25-26

GRADO EN INGENIERÍA INFORMÁTICA
CUARTO CURSO

GUÍA DE ESTUDIO PÚBLICA

MINERÍA DE DATOS (ING.TI)

CÓDIGO 71024062

UNED

25-26

MINERÍA DE DATOS (ING.TI)

CÓDIGO 71024062

ÍNDICE

NOMBRE DE LA ASIGNATURA	MINERÍA DE DATOS (ING.TI)
CÓDIGO	71024062
CURSO ACADÉMICO	2025/2026
DEPARTAMENTO	INTELIGENCIA ARTIFICIAL
TÍTULO EN QUE SE IMPARTE	GRADO EN INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN
CURSO - PERIODO - TIPO	- CUARTO - SEMESTRE 1 - OPTATIVAS
TÍTULO EN QUE SE IMPARTE	GRADO EN INGENIERÍA INFORMÁTICA
CURSO - PERIODO - TIPO	- CUARTO - SEMESTRE 1 - OPTATIVAS
TÍTULO EN QUE SE IMPARTE	PRUEBA DE APTITUD PARA HOMOLOGACIÓN DE GGRADO DE E.T.S. DE INGENIERÍA INFORMÁTICA (COMPLEMENTO)
Nº ETCS	6
HORAS	150.0
IDIOMAS EN QUE SE IMPARTE	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

Minería de Datos es sólo una de las denominaciones (la más popular, quizás, en el ámbito empresarial) de un área de investigación que podríamos llamar con más propiedad, Descubrimiento de Conocimiento a partir de datos. Corresponde con lo que desde antiguo se conoce como el principio de inducción en términos filosóficos. Hoy en día es una parte de lo que se conoce como Ciencia de Datos.

¿En qué consiste entonces la Minería de Datos? Se trata de conseguir reproducir con computadoras, tareas genuinamente humanas relacionadas con la extracción de conocimiento a partir de datos. Esas tareas pueden ser de varios tipos. Uno de ellos agrupa tareas en las que la computadora debe aprender a partir de un conjunto de ejemplos, generalizar las relaciones entre ellos, y aplicar el modelo resultante del aprendizaje a datos nuevos. La clasificación de casos en categorías responde bien a este patrón de tareas, pero también las actividades de control, en las que la máquina debe aprender a (generar un modelo para) controlar un sistema con unos objetivos explícitos, en problemas de planificación o de asignación de recursos, o las tareas de predicción, en las que el modelo aprendido a partir de los datos nos ayuda a inferir nuevos valores de unas variables desconocidas.

En todos estos casos, vemos la importancia que desempeñan los datos en este área. Se trata de producir modelos a partir de ejemplos que condensan el conocimiento que queremos aprehender, y para los que no disponemos de un modelo de conocimiento alternativo, expresado en lenguaje natural o estructurado.

Otro tipo de tareas encuadradas en la Minería de Datos, pero que no veremos en este curso, abordan la tarea de descubrir conceptos, relaciones o reglas en conjuntos de datos no etiquetados. Mientras que en el caso anterior (tareas de clasificación o regresión) disponemos de ejemplos que expresan los modelos que deseamos inferir, en las tareas de este tipo los datos están desnudos, y nuestra tarea consiste precisamente en descubrir esquemas clasificatorios, patrones repetidos, relaciones entre ellos, agrupamientos de datos o reglas que describan la distribución de los datos en un espacio de representación dado.

En este curso vamos a abordar los fundamentos del área. El objetivo del equipo docente ha sido, no abordar de manera extensiva pero superficial las diversas técnicas que se aplican en el área, sino proporcionar al estudiante los fundamentos que le permitan explorar en asignaturas sucesivas o por su cuenta, todas esas técnicas en las que aquí no podremos profundizar. Así pues, empezamos la casa por sus cimientos. Y los cimientos del edificio de la Minería de Datos son principalmente matemáticos y probabilísticos.

Esta asignatura, como podréis comprobar en la memoria de la titulación, se corresponde con la materia denominada Sistemas de Información, que comparte con las asignaturas de Bases de Datos y Gestión de Bases de Datos. Para su aprendizaje no es estrictamente necesario haber cursado las anteriores, pues lo que aquí se enseña se hace de manera independiente del sistema de almacenamiento de los datos. Sin embargo, sí es muy importante haber cursado las asignaturas de Fundamentos Matemáticos y Estadística.

Los conocimientos adquiridos a través de esta asignatura son los fundamentos de un área cuya exploración continúa en el Master de Inteligencia Artificial Avanzada o en el de Ingeniería y Ciencia de Datos, en las asignaturas relacionadas con la Minería de Datos. En ellas, se aplica todo lo aprendido aquí para entender las variadas técnicas avanzadas (como Máquinas de Vectores Soporte, Procesos Gaussianos, Redes Neuronales Artificiales, etc) y para adentrarnos en el mundo de la clasificación no supervisada o agrupamiento.

Existen multitud de vías en las que los conocimientos adquiridos aquí serán de utilidad en el futuro de los estudiantes. El aprendizaje estadístico (otra de las denominaciones de la Minería de Datos) abre un sinfín de perspectivas nuevas en una nueva era en la que los datos, en muchas ocasiones, desbordan la capacidad de los humanos de procesar información. Desde lo que se conoce como el cuarto paradigma de la Ciencia (o e-Ciencia, en una expresión poco afortunada) de aplicación en áreas como las bio-tecnologías o las grandes bases de datos científicas, a las aplicaciones empresariales en bancos o librerías virtuales, sin olvidar a los buscadores web. Lo que aquí aprenderemos es de aplicación general a todos esos campos, precisamente porque se trata de los fundamentos del área.

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR LA ASIGNATURA

Es necesario tener conocimientos bien asentados de Matemáticas (Análisis y Álgebra Matricial) y Estadística, adquiridos a través de las asignaturas de Fundamentos Matemáticos y Estadística.

EQUIPO DOCENTE

Nombre y Apellidos	LUIS MANUEL SARRO BARO (Coordinador de asignatura)
Correo Electrónico	lsb@dia.uned.es
Teléfono	91398-8715
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	INTELIGENCIA ARTIFICIAL

Nombre y Apellidos	JOSE MANUEL CASTILLO CARA
Correo Electrónico	manuelcastillo@dia.uned.es
Teléfono	
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	INTELIGENCIA ARTIFICIAL

HORARIO DE ATENCIÓN AL ESTUDIANTE

1. Equipo docente (en la sede central):

Dr. D Luis Manuel Sarro Baro

Horario de atención al estudiante:

Guardia: Lunes de 10 a 14 horas. Despacho 3.12**.** Tel.: 913988715. lsb@dia.uned.es

La dirección de contacto es:

ETSI Informática-UNED. Dpto. Inteligencia Artificial

c/Juan del Rosal, 16

28040 Madrid

La manera más rápida y sencilla de contactar con el equipo docente es a través del curso virtual. A través de los foros podemos compartir preguntas, respuestas y todo tipo de información relevante sobre la asignatura. Asimismo, en ellos se convocan videotutorías y se deciden los contenidos y fechas.

2. Profesores tutores (en el centro asociado correspondiente). Los horarios de atención del tutor serán suministrados por los propios centros asociados al inicio de curso.

TUTORIZACIÓN EN CENTROS ASOCIADOS

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

Capacidad para conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.
Competencias Generales:

CG.1 - Competencias de gestión y planificación: Iniciativa y motivación. Planificación y organización (establecimiento de objetivos y prioridades, secuenciación y organización del tiempo de realización, etc.). Manejo adecuado del tiempo
CG.2 - Competencias cognitivas superiores: selección y manejo adecuado de conocimientos, recursos y estrategias cognitivas de nivel superior apropiados para el afrontamiento y resolución de d diversos tipos de tareas/problemas con distinto nivel l de com plejida
CG.5 - Competencias en el uso de las herramientas y recursos de la Sociedad del Conocimiento: Manejo de las TIC. Competencia en la búsqueda de información relevante. Competencia en la gestión y organización de la información. Competencia en la recolección de dat

Competencias Específicas:

BC.11 - Conocimiento y aplicación de las características, funcionalidades y estructura de los Sistemas Distribuidos, las Redes de Computadores e Internet y diseñar e implementar aplicaciones basadas en ellos
BC.12 - Conocimiento y aplicación de las características, funcionalidades y estructura de las bases de datos, que permitan su adecuado uso, y el diseño y análisis de aplicaciones basadas en ellos
BC.13 - Conocimiento y aplicación de las herramientas necesarias para el almacenamiento, procesamiento y acceso a los Sistemas de Información, incluidos los basados en web
BTEti.2 - Capacidad para seleccionar, diseñar, desplegar, integrar, evaluar, explotar y mantener las tecnologías de hardware, software y redes, dentro de los parámetros de coste y calidad adecuados
BTEti.5 - Capacidad para seleccionar, desplegar, integrar y gestionar sistemas de información que satisfagan las necesidades de la organización, con los criterios de coste y calidad identificados

BTEti.7 - Capacidad de comprender, aplicar y gestionar la garantía y seguridad de los sistemas informáticos
FB.3 - Capacidad para comprender y dominar los conceptos básicos de matemática discreta, lógica, algorítmica y complejidad computacional, y su aplicación para el tratamiento automático de la información por medio de sistemas computacionales y para la resolución
FB.4 - Conocimientos básicos sobre el uso y programación de los ordenadores, sistemas operativos, bases de datos y programas informáticos con aplicación en ingeniería.

RESULTADOS DE APRENDIZAJE

El estudiante, al concluir y aprobar la asignatura, dominará los conceptos básicos del área de la Minería de Datos: el análisis probabilístico del problema de reconocimiento de patrones en tareas de clasificación y regresión; su tratamiento desde la perspectiva bayesiana, con el manejo de las entidades que desempeñan papeles relevantes en ese ámbito (verosimilitud, probabilidades marginales, evidencias...); los problemas de la selección de modelos de complejidad creciente, y la maldición de la dimensionalidad. Habrá, asimismo, adquirido destreza en el manejo de probabilidades y de entidades algebraicas, principalmente matriciales, suficiente como para profundizar los contenidos de la materia contenidos en los ejercicios no resultos que propondrá el equipo docente.
En resumen, el estudiante habrá interiorizado los fundamentos del área a través de las aproximaciones más simples (los modelos lineales) y se hallará en disposición de abordar el estudio posterior de las técnicas más avanzadas (y no necesariamente a través de modelos lineales) como las Máquinas de Vectores Soporte, los Procesos Gaussianos o las Redes Neuronales (por citar sólo tres ejemplos)
En forma de lista:

- Conocimiento de las diversas herramientas y estructuras matemáticas que sirven de base a los principales lenguajes de manipulación de datos.
- Conocer los lenguajes estándar de definición y manejo de datos en un SGBD
- Utilizar de forma optimizada los lenguajes estándar de definición y manipulación de datos así como el uso de estos para el desarrollo de software avanzado.
- Conocer las principales técnicas de la minería de datos y saber elegir y aplicar la más adecuada en función del tipo de tarea a resolver.
- Conocer las principales técnicas de evaluación del conocimiento aprendido y aplicar la más adecuada así como la plataforma software de minería de datos a utilizar.

CONTENIDOS

Tema 1: Introducción al Aprendizaje Estadístico

Contenidos del tema 1:

1.1: Ajustar datos con un polinomio como ejemplo de partida
1.2: Teoría de la probabilidad
1.3: Selección de modelos
1.4: La maldición de la dimensionalidad
1.5: Teoría de la decisión
1.6: Teoría de la información
1.7: Distribuciones de probabilidad

--

Este tema se estudiará en los capítulos 1 y 2 del texto base. Los epígrafes del 1.1 al 1.6 se estudian en el capítulo 1 y el 1.7 en el capítulo 2.

Son contenidos fundamentales que introducen los conceptos clave de la asignatura y del área de la minería de datos. Tienen validez no sólo en el contexto de los modelos lineales sino en cualquier ténica de Minería de Datos.

Se ha dividido en siete sub-bloques siguiendo los epígrafes del texto base correspondientes a los temas 1 y 2. En este primer bloque se introducirán conceptos básicos de una manera descriptiva. Estos conceptos básicos aparecerán de forma recurrente durante el curso en contextos más específicos y, por lo tanto, más complejos.

Sub-bloque 1.1. Ajuste de datos mediante polinomios.

Corresponde al texto introductorio al tema 1 y al epígrafe 1.1 del libro.

Introducción

El texto introductorio presenta de manera informal los conceptos más generales relativos al área de aprendizaje automático por máquinas. El aprendizaje automático es una disciplina basada en la Estadística (en ocasiones se conoce como aprendizaje estadístico) y está en la base de las tareas de Minería de Datos.

En este primer sub-bloque se presenta en detalle un ejemplo de aprendizaje para un problema de regresión (es decir, se trata de un caso de aprendizaje supervisado en el que el vector objetivo o target vector no es una clase perteneciente a un esquema de clasificación, sino que se trata de una variable numérica continua). Mediante este ejemplo, se van a ilustrar esos conceptos clave que reaparecerán a lo largo del curso y que serán el lenguaje conceptual que emplearemos para entender las distintas técnicas disponibles en el área de Minería de Datos.

Además, este ejemplo servirá para introducir la notación matemática que se utilizará a lo largo del texto base y que se resume en la sección pre-índice 'Notación Matemática'.

Aunque la profundidad con la que se tratarán los conceptos sea sólo superficial (especialmente en lo que refiere a los métodos de máxima verosimilitud o Bayesianos), es importante que el alumno no sienta por ello frustración. Más adelante se explicarán con mayor detalle y rigurosidad los conceptos.

Resultados de aprendizaje

1. Comprensión por el alumno de los conceptos de conjunto de entrenamiento, fase de entrenamiento o aprendizaje, modelo, conjunto de test, generalización, preprocesado de características, aprendizaje supervisado (clasificación y regresión) y no supervisado.
2. Familiarización con la notación matemática.
3. Comprensión del concepto de modelo lineal **en los coeficientes**.
4. Comprensión del concepto de función de error.
5. Comprensión de la tarea de selección de modelos
6. Comprensión cualitativa del concepto de sobre-ajuste y su relación con la complejidad del modelo y el tamaño del conjunto de entrenamiento.
7. Comprensión del concepto de regularización y de sus efectos. En particular, del caso de la regresión ridge en la que los coeficientes son cuadráticos.
8. Comprensión del concepto de conjunto de validación.

Sub-bloque 1.2. Teoría de la probabilidad

Corresponde al epígrafe 1.2 del texto base.

Resultados de aprendizaje

1. Comprensión de los conceptos de probabilidad y probabilidad condicionada.
2. Comprensión de los conceptos de probabilidad conjunta.
3. Adquisición de familiaridad con las relaciones entre probabilidades conjuntas y condicionadas (regla del producto) y entre probabilidades conjunta y marginal (regla de la suma). Adquisición del concepto de marginalización.
4. Comprensión del Teorema de Bayes y de los conceptos de probabilidad a priori, a posteriori.
5. Comprensión del concepto de independencia de variables.
6. Comprensión del concepto de variables discretas y variables continuas.
7. Comprensión del concepto de densidad de probabilidad.
8. Familiarización con el factor de Jacobi o Jacobiano y con el hecho de que el máximo de una distribución de probabilidad depende de la elección de variable con la que parametrizamos.
9. Comprensión del concepto de función de distribución acumulativa.
10. Compresión de la generalización del concepto de densidad de probabilidad al caso de distribuciones multivariantes.
11. Comprensión de los conceptos de valor esperado y covarianza.

12. Adquisición de familiaridad con las dependencias funcionales de los valores esperados de densidades de probabilidad multivariantes o condicionales.
13. Adquisición de destreza en la obtención de varianzas y covarianzas en los casos de variables aleatorias y vectores de variables aleatorias. En particular, del álgebra matricial necesaria para el manejo de covarianzas de vectores de variables aleatorias.
14. Comprensión de la diferencia de entre la interpretación frecuentista y la Bayesiana al concepto de probabilidad e inferencia paramétrica.
15. Comprensión del concepto de verosimilitud y del estimador de máxima verosimilitud.
16. Ilustración de la interpretación frecuentista de la inferencia paramétrica mediante la estimación bootstrap.
17. Toma de contacto del alumno con la diatriba entre las aproximaciones frecuentista y Bayesiana.
18. Familiarización del alumno con las propiedades de la distribución gaussiana y de la estimación de sus parámetros a partir de una muestra.
19. Toma de contacto del alumno con el concepto de sesgo en relación con las estimaciones de máxima verosimilitud. Este concepto está relacionado con el de sobreajuste.
20. Adquisición de la capacidad de formular un problema de regresión desde la interpretación bayesiana de la probabilidad, generalizando el problema del ajuste de curvas.
21. Comprensión por parte del alumno, de la equivalencia entre la maximización de la verosimilitud y la minimización de la función del error basado en la suma de cuadrados en el caso de que los errores estén distribuidos según una densidad de probabilidad gaussiana.
22. Comprensión del concepto de hiper-parámetro.
23. Comprensión del concepto de estimador máximo a posteriori (MAP).
24. Comprensión por parte del alumno de la importancia –en el contexto de las técnicas bayesianas -- de utilizar las distribuciones de densidad de probabilidad a posteriori completas y no reducirlas a un único estimador.
25. El alumno debe haber adquirido destrezas algebraicas y de cálculo que le permitan operar con distribuciones de probabilidad: marginalizar, normalizar, descomponer probabilidades conjuntas, calcular valores esperados, varianzas y obtener estimadores de máxima verosimilitud.

Sub-bloque 1.3. Selección de modelos.

Corresponde al epígrafe del texto base 1.3.

Introducción

Resultados de aprendizaje

1. Toma de contacto con el problema de la selección de modelos en general y, en particular, con la selección de la complejidad óptima de un modelo.

2. Comprensión de la técnica de evaluación conocida como validación cruzada, de sus limitaciones y problemas asociados.
3. Primera toma de contacto con criterios basados en teoría de la información (AIC y BIC), y toma de conciencia de que existe una solución completamente Bayesiana para este problema.

Sub-bloque 1.4 La maldición de la dimensionalidad

Corresponde al epígrafe del texto base 1.4.

Introducción

Resultados de aprendizaje

1. Comprensión de la dificultad de obtener conjuntos de entrenamiento representativos en espacios de alta dimensionalidad, y las consecuencias que implica para las tareas de Minería de Datos-
2. Toma de contacto con la posibilidad de seleccionar dimensiones intrínsecas en las que el problema sea de dimensión mucho menor.

Sub-bloque 1.5. Teoría de la decisión

Corresponde al epígrafe del texto base 1.5.

Introducción

Resultados de aprendizaje

1. Toma de contacto con el problema de determinar el objetivo del aprendizaje: minimización de la tasa de error en la clasificación o del valor esperado del coste.
2. Comprensión de los conceptos de función de coste y de utilidad y de la matriz de costes.
3. La inclusión del rechazo de casos entre las posibles decisiones de un clasificador.
4. Comprensión de las diferencias entre los modelos generativos y los modelos discriminantes, y la relación con los modelos no probabilísticos. Muy importante para aprendizajes posteriores.
5. Primera toma de contacto con el concepto de dato anómalo y su detección como problema de Minería de Datos.
6. Comprensión de las ventajas de los modelos generativos.
7. Comprensión de la generalización de los conceptos de la teoría de la decisión, ejemplificados inicialmente para problemas de clasificación, al caso de regresión.
8. Adquisición de destrezas correspondientes al cálculo variacional (explicados en el Apéndice D) para la minimización de la función de coste.
9. Comprensión conceptual de las implicaciones de elegir una función de coste cuadrática y su minimización: la predicción óptima resulta ser la media condicional.
10. Adquisición de la capacidad de interpretar el coste esperado en términos del sesgo y la varianza. Comprensión del significado de ambos términos.
11. Extensión de los conceptos aprehendidos relativos a los modelos generativos, discriminantes y no probabilísticos al caso de la regresión.

12. Ampliación de los recursos del estudiante a la hora de definir funciones de coste diferentes de la función cuadrática: funciones basadas en la métrica de Minkowski.

Sub-bloque 1.6. Teoría de la Información.

Corresponde al epígrafe del texto base 1.6.

Introducción

Resultados de aprendizaje

1. Aprehensión del concepto de información.
2. Comprensión de la definición de entropía de una variable aleatoria y de sus propiedades.
3. El alumno debe saber relacionar la entropía de una variable con la longitud mínima promedio de un mensaje que comunica el resultado de una medición de esa variable aleatoria.
4. El alumno debe saber generalizar el concepto de entropía al caso de variables continuas.
5. El alumno debe conocer las distribuciones de probabilidad de máxima entropía para variables discretas y continuas y sus propiedades, y debe ser capaz de llegar a demostrarlo matemáticamente.
6. El alumno debe haber adquirido la destreza suficiente para realizar cálculos de los valores de entropía para una distribución de probabilidad dada.
7. El alumno debe haber comprendido la diferencia entre la entropía y la entropía condicional y cómo se relacionan.
8. El alumno debe haber comprendido en qué sentido son importantes las definiciones de contenido de información y entropía para la Minería de Datos.
9. Comprensión del concepto de distancia de Kullback-Leibler entre dos distribuciones de probabilidad y de su definición matemática. El alumno debe conocer las propiedades de la distancia K-L.
10. El alumno conocerá el significado de la información mutua, su definición matemática y la relación con la entropía condicional.
11. El alumno sabrá las reglas para operar con estas entidades matemáticas tomadas del ámbito de la teoría de la información.

Sub-bloque 1.7. Distribuciones de probabilidad.

Corresponde a los epígrafes 2.3 a 2.3.6 (ambos incluidos) y al epigrafe 2.5.2 (Método de vecinos mas cercanos).

Introducción

Hemos visto anteriormente que una aproximación rigurosa a los problemas clásicos de la Minería de Datos requiere la utilización de un formalismo basado en la teoría de la probabilidad. Para ello, en muchas ocasiones será conveniente trabajar con distribuciones de densidad de probabilidad que están parametrizadas por unas pocas variables. En este sub-bloque, estudiaremos la distribución más importante en el área de la Minería de Datos:

la distribución normal o Gaussiana. Aunque sería deseable estudiar otras distribuciones incluidas en el capítulo 2 del texto base (beta, multinomiales, la distribución t...) la limitada extensión de la asignatura nos obliga a restringirnos a una única distribución.

Resultados de aprendizaje

1. El alumno debe comprender el concepto de probabilidades conjugadas en el análisis bayesiano y comprender su utilidad en este contexto.
2. El alumno debe entender la necesidad de recurrir a estos métodos bayesiano como resultado del sobreajuste que puede aparecer al emplear métodos frecuentistas en determinadas situaciones.
3. El alumno debe conocer la importancia de los métodos secuenciales para conjuntos de datos extensos.
4. El alumno debe conocer la propiedad asintótica que relaciona las estimaciones bayesiana y de máxima verosimilitud de la distribución predictiva de una variable.
5. El alumno debe conocer la importancia que la distribución normal o Gaussiana tiene en el contexto de la Minería de Datos.
6. El alumno debe conocer y memorizar la forma funcional de la distribución Gaussiana.
7. El alumno debe conocer el concepto de distancia de Mahalanobis y memorizar su definición. Asimismo debe conocer la relación entre la distancia euclídea y la distancia de Mahalanobis.
8. El alumno debe conocer la forma geométrica de la distribución Gaussiana. Debe saber que la distribución toma el mismo valor en hiper-superficies elipsoidales.
9. El alumno debe ser consciente de que, para que una distribución Gaussiana esté bien definida la matriz de covarianza debe ser definida positiva (y conocer lo que esto significa). El alumno debe conocer el concepto de matriz semi-definida positiva y lo que implica.
10. El alumno debe ser capaz de derivar y entender que en el espacio de coordenadas dado por los autovectores de la matriz de covarianza, ésta está alineada con los ejes y por lo tanto se puede representar por un producto de distribuciones gaussiana univariantes.
11. El alumno debe conocer las implicaciones de asumir las diversas simplificaciones de la distribución Gaussiana (matriz de covarianza diagonal o proporcional a la matriz identidad) en el análisis estadístico.
12. El alumno debe entender la definición de las distribuciones condicionales de probabilidad y, en particular, la forma de las distribuciones condicionales de una distribución Gaussiana.
13. El alumno debe conocer la definición de la matriz de precisión
14. El alumno debe entender (pero no memorizar) el procedimiento de obtención de la media y covarianza de una distribución condicional de probabilidad Gaussiana en el

caso general multivariante. No es importante que el alumno recuerde ni los pasos intermedios ni el resultado final, pero sí que entienda el procedimiento conocido como 'completar el cuadrado'.

15. El alumno debe entender el concepto y definición de probabilidad marginal y entender (pero no memorizar el desarrollo) la derivación de los valores de la media y de la varianza.
16. El alumno debe ser capaz de seguir (entender los pasos pero no memorizar) la aplicación de los métodos de compleción del cuadrado para derivar todos los factores del Teorema de Bayes para variables Gaussianas multivariantes.
17. El alumno debe entender la derivación de las estimaciones de máxima verosimilitud de la media de una distribución Gaussiana multivariante. Debe ser capaz de reproducir dicha derivación y recordar sus resultados y el valor de la estimación de máxima verosimilitud de la covarianza (aunque no su derivación que no aparece en el texto base).
18. El alumno debe entender el concepto de estimación sesgada y cómo corregirlo en el caso de la estimación de máxima verosimilitud de la matriz de covarianza.
19. El alumno debe ser capaz de derivar la probabilidad a posteriori del parámetro media de una distribución univariante gaussiana suponiendo conocida su varianza.
20. El alumno debe ser capaz de interpretar los resultados obtenidos al aplicar una estimación bayesiana del valor de la media de una distribución Gaussiana dada su varianza. En particular, debe ser capaz de analizar los valores asintóticos de dicha estimación cuando el número de casos tiende a cero y a infinito. También, en el caso asintótico en que la varianza de la probabilidad a priori de dicho parámetro (la media) tiende a infinito.
21. El alumno debe ser capaz de interpretar la inferencia bayesiana con probabilidades a priori conjugadas como un proceso de actualización secuencial de las estimaciones.
22. El alumno debe conocer la distribución Gamma y su importancia para el problema de estimar la precisión de una distribución Gaussiana de media conocida (con probabilidades a priori conjugadas).
23. El alumno debe conocer la solución al problema de inferencia bayesiana de los (dos) parámetros de una distribución Gaussiana (con probabilidades a priori conjugadas), y debe conocer las propiedades de la distribución Gaussiana-Gamma.
24. El alumno debe conocer la extensión de las técnicas de estimación bayesiana de los parámetros de una Gaussiana con probabilidades a priori conjugadas al caso multivariante: las distribuciones de Wishart y Gaussiana-Wishart. No debe conocer las fórmulas ni las constantes de normalización de memoria, pero sí sus dependencias funcionales.

25. El alumno debe conocer la definición de clasificador óptimo.

Tema 2: Modelos Lineales de Regresión

Contenidos del tema 2:
2.1: Modelos basados en funciones de base lineales
2.2: La descomposición sesgo-varianza
2.3: Regresión lineal Bayesiana
2.4:Comparación Bayesiana de Modelos
2.5:La aproximación de la evidencia
--
Los problemas de regresión y clasificación son conceptualmente idénticos, pero los espacios matemáticos sobre los que se definen (los tipos de variable aleatoria que definen el espacio de llegada) son diferentes. Vamos a utilizar el caso más general de la regresión para introducir los conceptos fundamentales sobre los que pivota gran parte del área del aprendizaje supervisado. En el tema siguiente profundizaremos en las particularidades de la clasificación supervisada a partir de lo aprendido aquí.

Es fundamental entender el concepto de modelo lineal. Se produce frecuentemente la confusión de definir un modelo lineal en términos de las variables aleatorias independientes o de entrada. Lo que define un modelo lineal es su dependencia respecto de los parámetros que lo definen, de manera que un modelo puede ser lineal porque lo es su dependencia respecto de éstos, aunque dependa de manera no lineal de las variables independientes o de entrada.
En general, un problema de regresión se define como el problema de predecir el valor de una variable dependiente (que puede ser multidimensional) a partir de una variable independiente (que también puede ser multidimensional) empleando para ello un modelo construido a partir de un conjunto de observaciones (es decir, pares constituidos por las variables independiente y dependiente correspondientes).
Desde una perspectiva probabilista, la información completa se halla en la distribución de probabilidad de la variable dependiente dada la variable independiente.

Sub-bloque 2.1. Modelos basados en funciones de base lineales

Corresponde a los epígrafes 3.1 y 3.1.1 a 3.1.5.

Introducción

Resultados de aprendizaje

1. El alumno debe conocer la expresión matemática que define una regresión lineal en los parámetros y en la/las variable(s) de entrada y en los parámetros, y el caso más general de modelos que son una combinación lineal de funciones no lineales de las variables de

entrada.

2. El alumno debe conocer los casos particulares en los que las funciones de la base son i) potencias de las variables de entrada, ii) funciones normales, iii) funciones sigmoideas logísticas, iv) términos de una serie de Fourier.
3. El alumno debe conocer el problema de utilizar una base de funciones (lineales o no lineales) que son globales para todo el espacio de variables de entrada, y la solución mediante funciones locales conocida como funciones de spline.
4. El alumno debe ser capaz de seguir y entender hasta los últimos detalles la derivación de la expresión correspondiente a la estimación de máxima verosimilitud de los parámetros de un modelo lineal en el caso de que las mediciones de la variable dependiente estén afectadas por ruido Gaussiano.
5. El alumno debe entender la relación entre el método de estimación por máxima verosimilitud y la estimación de mínimos cuadrados y conocer en qué circunstancias ambas estimaciones son equivalentes.
6. El alumno debe conocer y memorizar la expresión de la estimación de máxima verosimilitud de los parámetros de un modelo lineal en el caso de que las mediciones de la variable dependiente estén afectadas por ruido Gaussiano en función de la matriz de diseño.
7. Debe conocer las ecuaciones normales para la estimación por mínimos cuadrados.
8. Debe conocer la extensión del concepto de matriz inversa de una matriz no cuadrada (la pseudo-inversa de Moore-Penrose).
9. El alumno debe ser capaz de interpretar la estimación de máxima verosimilitud de la precisión de la distribución Gaussiana del ruido en términos de la varianza de las medidas respecto al modelo de regresión de máxima verosimilitud.
10. El alumno debe ser capaz de interpretar problemas de minimización de una función cuadrática del error en términos geométricos como la distancia mínima de un punto a un hiper-plano (obtenida mediante una proyección ortogonal).
11. El alumno debe conocer la adaptación del método de mínimos cuadrados a un aprendizaje secuencial: el método de descenso del gradiente secuencial o estocástico y el algoritmo *least-mean-squares (LMS)*.
12. El estudiante debe conocer el concepto de regularización y el caso particular conocido como decaimiento de pesos o contracción de parámetros. Asimismo, debe conocer la ventaja de cálculo que proporciona la forma matemática del decaimiento de pesos: la preservación de la forma cuadrática de la función de error.
13. El alumno debe ser capaz de derivar la solución de la ecuación de minimización del error con regularización cuadrática de manera algebraica.
14. El alumno debe conocer la generalización de este tipo de regularización a exponentes diferentes de 2 (regularización cuadrática) y el nombre y propiedades del caso de

exponente 1 (regularización lasso).

15. El estudiante debe entender la equivalencia entre minimizar la función de error sin regularizar sujeta a restricciones y la minimización de la función de error regularizada.
16. El estudiante debe entender la utilidad de la regularización en el caso de conjuntos de datos de tamaño pequeño, para evitar sobreajuste de modelos mediante la limitación efectiva de la complejidad del modelo.
17. El alumno debe ser capaz de extender todo lo aprendido en este sub-bloque al caso de múltiples variables dependientes, y de deducir el desacoplamiento de los diferentes problemas de regresión que representan cada una de las variables.

Sub-bloque 2.2. La descomposición sesgo-varianza

Corresponde al epígrafe 3.2

Introducción

Las técnicas de ajuste paramétrico vistas con anterioridad y basadas en máxima verosimilitud o, de manera equivalente, minimización del error cuadrático, presentan una tendencia al sobreajuste cuando los conjuntos de entrenamiento (en presencia de ruido) son de tamaño insuficiente. Hemos visto que la regularización es una alternativa útil, pero que en exceso puede limitar excesivamente la complejidad de los modelos. En este subbloque vamos a utilizar una perspectiva frecuentista para profundizar en la comprensión de las fuentes de error en que incurrimos al ajustar modelos de regresión y su dependencia respecto de la complejidad efectiva de los modelos empleados.

Resultados de aprendizaje

1. El alumno debe ser capaz de escribir el valor esperado de la función de pérdida cuadrática (square loss function) en términos de la función objetivo y de la función resultado de una estimación (por el método que sea).
2. El estudiante debe entender el experimento mental que conduce a la descomposición del valor esperado de la pérdida cuadrática en términos de sesgo, varianza y ruido.
3. El alumno debe ser capaz de seguir la derivación de la fórmula que descompone el valor esperado de la función pérdida (promediado sobre potenciales conjuntos de entrenamiento) en términos del sesgo (bias), de la varianza y del error debido al ruido.
4. El alumno de saber la expresión que define a cada uno de estos términos (sesgo, varianza y ruido).

Sub-bloque 2.3. Regresión lineal Bayesiana

Corresponde a los epígrafes 3.3 y 3.3.1 a 3.3.3.

Introducción

Una solución alternativa a la regularización en el problema de sobreajuste que aparece en las estimaciones de máxima verosimilitud es la utilización de métodos de estimación bayesianos que proporcionan estimaciones óptimas de la complejidad requerida a partir del conjunto de entrenamiento. En este sub-bloque introducimos los conceptos básicos de

regresión lineal bayesiana.

Resultados de aprendizaje

1. El estudiante debe ser capaz de formular el problema de estimación de parámetros de un modelo lineal en términos de inferencia bayesiana (es decir, de escribir las expresiones de la probabilidad a priori, de la verosimilitud y de la probabilidad a posteriori) para el caso de ruido gaussiano, con probabilidad a priori conjugada.
2. El estudiante debe ser capaz de deducir las expresiones para la media y matriz de covarianza de la distribución a posteriori de los parámetros del modelo lineal.
3. El estudiante debe entender la equivalencia entre la estimación máximo a posteriori de este modelo y la de una estimación de máxima verosimilitud regularizada con un término cuadrático.
4. El estudiante debe entender y conocer la definición de distribución predictiva.
5. El estudiante debe ser capaz de derivar los estadísticos resumen de la distribución predictiva (para el caso de ruido Gaussiano y distribuciones a priori centradas en cero) a partir de los resultados conocidos de la convolución de dos Gaussianas.
6. El alumno debe ser capaz de interpretar las dos componentes de la varianza de la distribución predictiva y distinguir su dependencia con la variable independiente. También debe conocer las propiedades asintóticas de dicha varianza en el caso de un conjunto de datos infinito.
7. El estudiante debe conocer el valor asintótico de la varianza de la distribución predictiva cuando se evalúa lejos de los centros de funciones de base localizadas (extrapolación).
8. El alumno debe conocer la forma de la distribución predictiva, solución a este problema (ver punto anterior) en el caso de que queramos determinar simultáneamente los parámetros del modelo y la varianza del ruido.
9. El alumno debe ser capaz de seguir (entender cada paso de) la derivación del valor de la media de la distribución predictiva.
10. El alumno debe ser capaz de reescribir la expresión de la media de la distribución predictiva en términos de una función kernel equivalente o matriz de suavizado.
11. El alumno debe conocer el concepto de función de suavizado lineal.
12. El alumno debe saber y entender que la media de una distribución predictiva basada en funciones localizadas como las Gaussianas se puede expresar como una combinación lineal de contribuciones de los puntos del conjunto de entrenamiento en la que cada contribución está ponderada de manera que el peso en la combinación disminuye con la distancia entre el punto de entrenamiento y el punto donde se evalúa la predicción.
13. El alumno debe ser consciente de que, como consecuencia de lo anterior, las predicciones en puntos próximos están tanto más correlacionadas cuanto más próximos estén los puntos en los que se evalúa la predicción.

14. El alumno debe saber que la suma de la función kernel evaluada en todos los puntos del conjunto de entrenamiento es la unidad.
15. El alumno debe conocer la propiedad de las funciones kernel de poder ser expresadas como producto escalar o interno de funciones no lineales de las variables de entrada.

Sub-bloque 2.4. Comparación Bayesiana de Modelos.

Corresponde al epígrafe 3.4

Introducción

Como hemos visto con anterioridad, las estimaciones de máxima verosimilitud son propensas al sobre-ajuste. Hemos visto una técnica para evitar ese sobre-ajuste denominada regularización, y el problema que plantea la determinación de los valores de sus parámetros.

En este sub-bloque vamos a introducir una alternativa de vigencia global en todas las áreas de la Minería de Datos. Por ello, es difícil exagerar su importancia.

Resultados de aprendizaje

1. El estudiante debe conocer los elementos básicos de los que se parte para realizar la comparación bayesiana de modelos: el concepto de modelo, la probabilidad a posteriori de un modelo dado un conjunto de datos, la probabilidad a priori de un modelo, y la evidencia de un modelo o su verosimilitud marginal.
2. El alumno debe saber interpretar el concepto de verosimilitud marginal de un modelo. ¿Marginal respecto de qué?
3. El estudiante debe conocer el concepto y definición de factor de Bayes.
4. El alumno debe saber utilizar las probabilidades a posteriori de los modelos para realizar inferencias utilizando la distribución predictiva. ¿Qué alternativas hay para dicha distribución predictiva?
5. El estudiante debe entender el concepto de selección de modelo desde esta perspectiva.
6. El estudiante debe ser capaz de computar la evidencia de un modelo como una marginalización sobre los parámetros.
7. El estudiante debe ser capaz de reproducir la aproximación a la evidencia basada en la simplificación de asumir que tanto la distribución a posteriori de los parámetros como el prior son distribuciones constantes de una determinada anchura. Asimismo debe ser capaz de interpretar dicha aproximación como una descomposición de la evidencia en dos términos: la bondad del ajuste para los parámetros más probables a posteriori y un término que mide la anchura relativa de las distribuciones a posteriori y a priori. El estudiante debe entender las implicaciones de esta descomposición en una y varias dimensiones.

8. El estudiante debe entender en qué consiste el concepto de complejidad óptima de un modelo en el sentido bayesiano y qué relación tiene con la capacidad expresiva de un modelo medida por la evidencia.
9. El estudiante debe ser capaz de demostrar que la comparación bayesiana de modelos favorecerá en promedio (sobre conjuntos de datos) el modelo correcto si se cumple que el modelo empleado para generar los datos está entre las alternativas. Para ello, hará uso del concepto de distancia de Kullback-Leibler.
10. El alumno debe ser consciente de que la comparación bayesiana de modelos no requiere la división del conjunto de datos en dos subconjuntos (entrenamiento y validación o test) para evitar el sobre-ajuste y que, por tanto, hace uso de todos los datos disponibles para entrenar.

Sub-bloque 2.5. La aproximación de la evidencia.

Corresponde a los epígrafes 3.5, 3.5.1 a 3.5.3 y 3.6.

Introducción

Calcular la evidencia (marginalizar sobre todos los parámetros e hiperparámetros que hemos visto en sub-bloques anteriores) puede ser intratable desde un punto de vista analítico. Por lo tanto, la técnica que hemos visto en el sub-bloque anterior precisa, para su aplicación práctica, de alguna aproximación o simplificación que la haga factible en casos generales.

Resultados de aprendizaje

1. El estudiante debe conocer el fundamento de la técnica conocida como aproximación a la evidencia, Bayes empírico, máxima verosimilitud de nivel 2, o máxima verosimilitud generalizada.
2. El alumno debe entender pero no memorizar la derivación de la aproximación a la evidencia. Sí debe recordar las dependencias funcionales de dicha aproximación.
3. El alumno debe recordar que la maximización de la evidencia implica soluciones implícitas.
4. El estudiante debe ser capaz de interpretar el valor de como un número efectivo de parámetros. Es muy importante que el alumno reflexione sobre el concepto de número efectivo de parámetros.
5. El estudiante debe entender la relación entre el parámetro , , los valores de los parámetros w y el número efectivo de parámetros de un modelo.
6. El alumno debe conocer la relación entre el tamaño del conjunto de datos, el número efectivo de parámetros y las estimaciones de los hiperparámetros y . ¿Qué ocurre si se dispone de un número elevado de observaciones pero concentradas en un entorno reducido del espacio de variables independientes?
7. El alumno debe ser consciente de las limitaciones de los modelos lineales y conocer su origen (el de las limitaciones).

Tema 3: Modelos lineales de clasificación

Contenidos del tema 3:

3.1: Funciones discriminantes

3.2: Modelos Generativos Probabilísticos

3.3: Modelos discriminantes probabilísticos

3.4: La aproximación de Laplace y su utilidad para comparar modelos

3.5: Regresión Logística Bayesiana

--

En este bloque vamos a extender lo aprendido en el anterior al caso en que la variable dependiente que queremos predecir es una variable de clase. Este problema es equivalente a dividir el espacio de entrada (de las variables independientes) en regiones cada una de las cuales corresponde a una clase. Las fronteras que separan unas regiones de otras se conocen como fronteras de decisión. En el caso particular que nos ocupa en este bloque, veremos que las fronteras de decisión inducidas por los modelos lineales de clasificación son líneas rectas, planos o hiperplanos en más de dos dimensiones.

A lo largo del bloque veremos que existen al menos tres formas de abordar el problema. En la primera, el modelo no es probabilista y simplemente asigna a cada vector del espacio de entrada una clase. El siguiente nivel de complejidad consiste en emplear modelos probabilistas que estimen para un vector de entrada dado, una distribución de probabilidad de pertenencia a cada una de las clases del esquema empleado. Esta segunda posibilidad (la que denominamos probabilista) se puede llevar a cabo desde una metodología discriminante que modele directamente las probabilidades de pertenencia que deseamos encontrar, o desde una metodología generativa que llegue al mismo objetivo pero modelando otras distribuciones de probabilidad previas cuya combinación con el teorema de Bayes nos conduzca al objetivo.

Es importante ser conscientes de que el espacio de variables dependientes es esencialmente diferente al caso general de la regresión y que ello implica i) la utilización de una función de activación y ii) el hecho de que los modelos ya no sean lineales en los parámetros. Es extraordinariamente importante que el alumno sea consciente del abuso del lenguaje que representa seguir hablando de modelos lineales, y de las diferencias entre los modelos lineales de regresión y clasificación.

Sub-bloque 3.1. Funciones discriminantes

Corresponde a los epígrafes 4.1 y 4.1.1 a 4.1.7.

Introducción

Resultados de aprendizaje

1. El alumno debe entender el concepto de función discriminante y de función discriminante lineal.
2. El alumno debe ser capaz de distinguir el concepto de sesgo o *bias* en sentido estadístico y sesgo como constante en una función lineal.
3. El alumno debe ser capaz de interpretar el sesgo de una función lineal como un umbral de decisión.
4. El alumno debe adquirir la destreza de interpretar geométricamente un modelo lineal como un hiperplano frontera que separa dos regiones del espacio de entrada.
5. El alumno debe entender por qué en el espacio de coordenadas extendidas el hiperplano de separación pasa por el origen.
6. El alumno debe conocer la problemática asociada a las aproximaciones “clase frente al resto” y “clase frente a clase” para la generalización de los clasificadores al caso de más de dos clases.
7. El alumno debe conocer la solución de mínimos cuadrados para clasificación, y por qué ésta no puede ser interpretada como una probabilidad.
8. El estudiante debe conocer los problemas asociados con las soluciones de mínimos cuadrados para clasificación, en particular, los ligados a los datos atípicos.
9. El estudiante debe ser capaz de interpretar la solución de mínimos cuadrados como equivalente a una solución de máxima verosimilitud en la que los errores asociados a las medidas están distribuidos según una función gaussiana. Como consecuencia, debe entender las limitaciones de su aplicación a problemas de clasificación.
10. El estudiante debe conocer la interpretación de los modelos lineales de clasificación como problemas de reducción de la dimensionalidad, y ser capaz de explicarla.
11. Debe ser capaz de enunciar los ojetivos deseables de una reducción de la dimensionalidad (proyección) óptima para un problema de clasificación.
12. El estudiante debe ser capaz de llegar a la solución para la proyección que reduce a 1 la dimensionalidad de un problema de clasificación conocida como Discriminante Lineal de Fisher, y enunciar las definiciones que la caracterizan.
13. El estudiante debe ser capaz de enunciar al menos una forma de definir el umbral de decisión en la dimensión proyectada.
14. El estudiante debe ser capaz de explicar la relación entre el discriminante lineal de Fisher y la solución de mínimos cuadrados.
15. El estudiante debe ser capaz de enunciar la definición de un perceptrón: una transformación **fija** no lineal, seguida de un modelo lineal generalizado. Debe ser capaz de definir la función de activación y cómo se integra en la definición del perceptrón y de explicar en qué consiste la componente de sesgo (*bias*).
16. El estudiante debe entender en qué consiste el proceso de aprendizaje del perceptrón y qué alternativas sencillas existen para la definición del error que aquél debe minimizar.

17. El estudiante debe saber contestar por qué no es sencillo utilizar el número total de errores en la clasificación como medida del error que debemos minimizar.
18. El estudiante debe conocer la función de error inherente a la definición del perceptrón. ¿Por qué soluciona el problema de la elección el punto anterior?
19. El alumno debe ser capaz de enunciar la fórmula de cálculo iterativo del vector de pesos de un perceptrón en el marco del descenso estocástico del gradiente (ver sección 5.2.4 Stochastic Gradient Descent, del texto base, edición de 2006), y de dar una explicación intuitiva de su significado.
20. El estudiante debe ser capaz de enunciar las propiedades del algoritmo de aprendizaje del perceptrón en lo referente a la reducción del error **en cada paso**. Además, debe ser capaz de enunciar el teorema de convergencia del perceptrón.
21. El estudiante debe entender que las soluciones obtenidas para los perceptrones no pueden ser interpretadas probabilísticamente.

Sub-bloque 3.2. Modelos Generativos Probabilísticos

Corresponde a los epígrafes 4.2 y 4.2.1 a 4.2.4.

Introducción

En este sub-bloque vamos a pasar de los modelos discriminantes vistos en el sub-bloque anterior, a los modelos generativos, y a los modelos generativos probabilísticos en particular. Para ello, tendremos que modelar explícitamente las probabilidades condicionadas a la clase (verosimilitudes) y las probabilidades a priori.

Resultados de aprendizaje

1. El estudiante debe ser capaz de expresar la probabilidad a posteriori de pertenencia a una clase para una observación, en función de las verosimilitudes y de las probabilidades a priori (a partir del teorema de Bayes).
2. El estudiante debe conocer la definición y las propiedades analíticas de la función sigmoide.
3. El estudiante debe ser capaz expresar dicha probabilidad de pertenencia a posteriori como una función logística sigmoidea. Debe asimismo ser consciente de que no ha necesitado proporcionar expresiones funcionales de las verosimilitudes ni de las probabilidades a priori para llegar a dicha expresión.
4. El estudiante debe conocer la propiedad de simetría de la función logística sigmoidea y la forma funcional de su inversa: la función *logit* o '*log odds*'. Esta función *log odds* representa el logaritmo del cociente entre las probabilidades a posteriori de pertenencia a dos clases, en un problema de clasificación dicotómica.
5. El estudiante debe conocer la extensión (sencilla y directa) de la función logística sigmoidea al caso de problemas multi-clase: la función exponencial normalizada o softmax.

6. El estudiante debe ser capaz de aplicar el formalismo anterior al caso en que las verosimilitudes de clase sean funciones Gaussianas multi-variantes, todas ellas con la misma matriz de covarianza. Debe saber que las hiper-superficies de separación inducidas en ese caso son lineales (hiper-planos) y que el efecto de las probabilidades a priori se limita a desplazar de forma paralela dichos hiper-planos.
7. Es muy importante que el estudiante reconozca un modelo lineal generalizado en las probabilidades a posteriori de pertenecia a las clases cuando las verosimilitudes son gaussianas multivariantes de igual covarianza.
8. El estudiante debe conocer que las propiedades de linealidad de las hiper-superficies de separación se mantienen en el caso de problemas multi-clase, pero no cuando las matrices de covarianza de las clases difieren, en cuyo caso se trata de cuádricas o discriminantes cuadráticos.
9. El estudiante debe saber por qué los hiper-planos de separación mencionados en el punto anterior son las fronteras de decision que producen un error de clasificación mínimo.
10. El estudiante debe entender que no basta con definir modelos paramétricos para las verosimilitudes. Una vez definidos, necesitamos encontrar los valores de los parámetros que minimizan el error.
11. El estudiante debe conocer la definición de solución de máxima verosimilitud, es decir, el procedimiento para llegar a ella, al menos en el caso de dos clases y matrices de covarianza iguales. No es necesario que memorice la derivación, pero sí que conozca la solución: los centros de las gaussianas son las medias de los casos de entrenamiento, y la matriz de covarianza, la suma ponderada de las covarianzas de los casos de cada clase respecto a la media de clase.
12. El estudiante debe saber que en el caso de variables de entrada discretas (no continuas) el modelo resultante de aplicar un modelo naïve Bayes es, de nuevo, una probabilidad a posteriori que es lineal en los parámetros del modelo.
13. El estudiante debe conocer las propiedades que caracterizan a las probabilidades a posteriori cuando las verosimilitudes pertenecen a la familia exponencial, en particular la linealidad de las fronteras de decisión.

Sub-bloque 3.3. Modelos discriminantes probabilísticos

Corresponde a los epígrafes 4.3 y 4.3.1 a 4.3.6.

Introducción

Hasta ahora se han tratado modelos discriminantes no probabilísticos, y modelos generativos probabilísticos. ¿No se pueden realizar modelos discriminantes probabilísticos? La respuesta es sí. En este bloque nos ocupamos de ello.

Resultados de aprendizaje

1. El estudiante debe entender que los modelos discriminantes probabilísticos se fundamentan en la determinación directa de los parámetros del modelo lineal generalizado sin pasar previamente por una definición de las verosimilitudes de clase.
2. El estudiante debe recordar el concepto de funciones base: transformaciones no lineales del espacio de características. Estas funciones han aparecido en apartados anteriores como el perceptrón.
3. El estudiante debe ser consciente de que si un modelo lineal de clasificación induce fronteras de decisión lineales en el espacio transformado, las fronteras de decisión en el espacio de características original serán, en general, no lineales.
4. El estudiante debe ser consciente de que un problema de clases no separables por hiper-planos puede ser linealmente separable en el espacio transformado por las funciones de base no lineales.
5. El estudiante debe saber que el solapamiento entre clases en el espacio de características original no se reduce en el espacio transformado por las funciones de base, y puede aumentar.
6. El estudiante debe ser capaz de definir qué se entiende por modelo de regresión logística, y entender que se refiere a un problema de clasificación aunque el nombre se refiera a la regresión.
7. El estudiante debe reconocer en el modelo de regresión logística un modelo lineal generalizado.
8. El estudiante debe ser capaz de explicar cómo podemos determinar los parámetros del modelo de regresión logística mediante la máxima verosimilitud, y reconocer que el número de parámetros que necesitamos determinar con la aproximación discriminante es mucho menor que en la aproximación generativa. Deber entender que la aproximación de máxima verosimilitud a la determinación de parámetros en este caso implica la definición de la verosimilitud de los parámetros del modelo discriminante.
9. Como es habitual, en lugar de maximizar la verosimilitud, minimizamos el logaritmo de la verosimilitud cambiado de signo: el *cross-entropy error.*
10. El estudiante debe saber (y es muy importante) que las soluciones de máxima verosimilitud al problema de la regresión logística pueden sufrir un severo sobreajuste en problemas separables linealmente, y debe ser capaz de explicar por qué.
11. El estudiante debe ser capaz de enunciar dos soluciones al problema del sobreajuste de los modelos de regresión logística en casos linealmente separables.
12. El estudiante debe ser capaz de explicar intuitivamente en qué consiste el método de mínimos cuadrados iterativo y con ponderación adaptativa. Debe entender que se basa en el hecho de que, a pesar de que la función sigmoidea imposibilita la obtención de soluciones analíticas al problema de regresión logística, la función de error todavía es cóncava y por lo tanto, se puede diseñar un algoritmo iterativo para inferir los parámetros

que hacen mínimo ese error.

Sub-bloque 3.4. La aproximación de Laplace y su utilidad para comparar modelos.

Corresponde a los epígrafes 4.4 y 4.4.1.

Introducción

Este sub-bloque es un interludio necesario para abordar el problema de la clasificación desde una perspectiva totalmente Bayesiana. Esta perspectiva no se puede seguir de forma analítica hasta sus últimas consecuencias, y sólo podemos acercarnos de maneras aproximadas. Aquí veremos una de ellas. De manera adicional, veremos una aplicación en un contexto ajeno al tratado en este bloque, pero de una importancia fundamental: la comparación de modelos (ya vista con anterioridad) y el criterio de información bayesiano (BIC).

Resultados de aprendizaje

1. El estudiante debe conocer el procedimiento para aproximar una función multivariante en un máximo, mediante una función gaussiana, a través del desarrollo de Taylor de orden 2 y de la integración de las funciones Gaussianas.
2. El estudiante debe conocer las limitaciones de la aproximación de Laplace.
3. El estudiante debe conocer por qué la aproximación de Laplace es útil para hacer una comparación Bayesiana de modelos exhaustivos.
4. El estudiante debe ser capaz de explicar con sus propias palabras los cuatro términos que se obtienen al calcular la evidencia de un modelo con la aproximación de Laplace para la probabilidad conjunta de datos y parámetros.
5. El estudiante debe conocer cómo se simplifican esos cuatro términos en el caso de que la probabilidad a priori sea muy ancha y que la matriz de covarianza de la aproximación de Laplace sea no singular. Es muy importante que el estudiante recuerde el resultado: el criterio BIC también conocido como de Schwartz.

Sub-bloque 3.5. Regresión Logística Bayesiana

Corresponde a los epígrafes 4.5 y 4.5.1 a 4.5.2.

Introducción

En este sub-bloque vamos a ampliar lo estudiado en 3.2 y a aplicar una metodología totalmente Bayesiana. Desgraciadamente, no es posible hacerlo de manera analítica por varios motivos, por lo que tendremos que emplear aproximaciones como la vista en el bloque anterior.

Resultados de aprendizaje

1. El estudiante debe entender el procedimiento general que nos lleva a la expresión de la probabilidad a posteriori de los parámetros dado el conjunto de entrenamiento, y qué problemas plantea su uso inferencial.
2. El estudiante debe ser capaz de explicar con sus propias palabras cómo podemos obtener una aproximación Gaussiana a dicha probabilidad a posteriori.

3. El estudiante debe ser capaz de explicar en palabras cómo se define la probabilidad predictiva de clase, bajo la aproximación Laplaciana de la probabilidad a posteriori de los parámetros del modelo.
4. Sabiendo que la convolución del producto de una Gaussiana y una función sigmoidea se puede aproximar por la misma convolución pero con una función probit, el estudiante debe saber cuál es la forma funcional de la probabilidad predictiva de la clase.
5. En todo lo anterior, no es necesario que el estudiante memorice fórmulas, pero sí que entienda el problema general que se trata en este sub-bloque y cómo se aborda.

METODOLOGÍA

La asignatura se cursa de una manera clásica en la educación a distancia. Al tratarse de una asignatura de fundamentos, se concede una importancia especial a los aspectos teóricos, y las actividades prácticas están supeditadas a la consolidación de los conceptos teóricos.
La interacción con el equipo docente se realizará principalmente a través de la plataforma de aprendizaje virtual de la asignatura, donde se pretende que sean los propios alumnos los que resuelvan sus dudas (con la ayuda y supervisión en todo momento del profesor) de manera colaborativa. Todo ello, desde la convicción de que lo que se descubre se aprende mucho mejor que lo que se asimila de forma pasiva. El equipo docente valorará muy positivamente la participación en los foros con mensajes que colaboren en la resolución de problemas y dudas.
El 5% de los créditos asignados se destina a la preparación para el estudio del contenido teórico, lo que incluye la lectura de las orientaciones y una primera lectura del índice del texto base.

El segundo bloque, el más importante en cuanto a fracción del total (80%) lo constituye el estudio de los contenidos teóricos (60%) y el desarrollo de ejercicios de consolidación de lo aprendido (20%) mediante la resolución de problemas propuestos cuya respuesta estará a disposición de los alumnos (ejercicios de auto-evaluación).

Finalmente, el tercer bloque se asigna a una práctica entregable a la que corresponde el 10% de la nota (y el 15% de los créditos de la asignatura). Consistirá en la resolución de tres de los ejercicios propuestos en el texto base pero cuya solución no está disponible a través de Internet (ejercicios de descubrimiento).

SISTEMA DE EVALUACIÓN

TIPO DE PRUEBA PRESENCIAL

Tipo de examen: Examen de desarrollo

Preguntas desarrollo

Duración del examen: 120 (minutos)

Material permitido en el examen

El texto base de la asignatura

Criterios de evaluación

El 90% de la nota final corresponde a la nota obtenida en la prueba presencial. Sea NPP la nota de la prueba presencial calificada de 0 a 10. Los enunciados del examen se corresponderán con algunos de los ejercicios de descubrimiento propuestos a comienzo de curso por el equipo docente. Por lo tanto, no habrá preguntas que exijan la memorización de demostraciones. Si es necesario, el equipo docente proporcionará definiciones complejas que sean necesarias para la resolución de los enunciados de examen. Consideramos que lo más importante no es la memorización de fórmulas sino la comprensión de los conceptos clave del área y la adquisición de las destrezas algebraicas necesarias para manipular fórmulas.

% del examen sobre la nota final: 0

Nota del examen para aprobar sin PEC: 0

Nota máxima que aporta el examen a la calificación final sin PEC: 0

Nota mínima en el examen para sumar la PEC: 0

Comentarios y observaciones

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC? Si

Descripción

El equipo docente propondrá cada año dos subconjuntos de ejercicios tomados del texto base. El primero estará compuesto de ejercicios cuya respuesta se encuentra disponible en el sitio web del libro (auto-evaluación); el segundo contendrá ejercicios cuya respuesta no está disponible a través de Internet y de entre los que se seleccionarán los enunciados del examen (descubrimiento).

El 10% de la nota final de la asignatura corresponderá a la evaluación por parte de los tutores o equipo docente de un conjunto de 3 ejercicios de descubrimiento elegidos por el estudiante de entre los propuestos por el equipo docente. Sea NED la nota de 0 a 10 asignada a estos ejercicios de descubrimiento. Entonces, la nota combinada NC será igual a 0.9*NPP+0.1*NED.

La PEC es voluntaria pero no entregarla equivale a renunciar a ese 10% de la nota final.

Criterios de evaluación

Ponderación de la PEC en la nota final	0
Fecha aproximada de entrega	Final del periodo de pruebas presenciales.
Comentarios y observaciones	

Si se realizan dos entregas de la PEC (convocatorias de junio y septiembre), ambas serán evaluadas y se utilizará en cada convocatoria la entrega correspondiente. En caso de entregarse la PEC en la convocatoria de junio y no en la de septiembre, se mantendrá la evaluación de junio.

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s?	Si

Descripción

Un último tipo de tarea incluye la realización de una práctica de experimentación numérica evaluable. Dicha práctica será evaluada por los tutores, y podrá suponer hasta 2 puntos sobre 10 en la nota final. La nota final se calculará sumando a la nota combinada NC la puntuación de la práctica (NPEN) siempre y cuando ésta última supere los 4 puntos sobre 10. Si la suma de ambas notas supera los 10 puntos, la nota evidentemente será de 10. Su objetivo (el de la práctica evaluable) es facilitar que el alumno adquiera familiaridad con los casos prácticos de experimentación numérica a los que se les aplica todo el bagaje conceptual adquirido durante el curso. El enunciado de la práctica se hará público cada año a comienzo de curso.

Esta práctica es voluntaria y se puede alcanzar la máxima calificación (10) sólo con la prueba presencial y la PEC.

Criterios de evaluación

Ponderación en la nota final	0
Fecha aproximada de entrega	Final del periodo de pruebas presenciales.
Comentarios y observaciones	

¿CÓMO SE OBTIENE LA NOTA FINAL?

La nota final se obtendrá siempre como la suma 0.9*NPP+0.1*NED+NPEN o 10 en el caso de que la suma supere la nota máxima (10). Se tendrán en cuenta las consideraciones referentes a la posibilidad de que haya varias entregas de la PEC como indicado más arriba.

BIBLIOGRAFÍA BÁSICA

ISBN(13):9780387310732
Título:PATTERN RECOGNITION AND MACHINE LEARNINGnull
Autor/es:Christopher M. Bishop ;
Editorial:Springer

BIBLIOGRAFÍA COMPLEMENTARIA

ISBN(13):9780387848587
Título:THE ELEMENTS OF STATISTICAL LEARNINGnull
Autor/es:Hastie, Trevor ; Tibshirani, Robert J. ; Friedman, Jerome ;
Editorial:Springer

RECURSOS DE APOYO Y WEBGRAFÍA

El curso se desarrolla fundamentalmente a través de la plataforma aLF de la UNED. La información y el material complementario se encuentra en dicho curso, y la interacción con el equipo docente se desarrollará principalmente a través de los foros de la asignatura en dicha plataforma. Por supuesto, el equipo docente también atenderá a los alumnos a través del teléfono (913988715) o de manera presencial en el horario de guardia (Lunes de 10:00 a 14:00). Es recomendable acordar una cita previamente. Finalmente, el equipo docente estará también disponible a través de software de videoconferencia, preferiblemente skype. De nuevo, será necesario concertar una cita con anterioridad.

El equipo docente será el responsable de responder a las dudas que surjan sobre el funcionamiento de la asignatura y sobre los contenidos teórico de ésta (siempre que sea posible, a través de los foros, pues de esta manera las respuestas quedan a disposición de otros alumnos que puedan compartirlas). En principio, y salvo circunstancias excepcionales que lo impidan, el tiempo máximo de espera para las respuestas a las preguntas del foro es de 7 días, (el tiempo entre guardia y guardia). Por regla general, nunca se alcanza ese periodo y en la medida de lo posible el equipo docente intenta responder a las cuestiones con la máxima celeridad que permiten las otras obligaciones del profesorado, entre las que cabe destacar la docencia en otras asignaturas y las tareas de investigación y administración. Como orientación, se puede decir que en periodo lectivo, fuera de épocas de examen o viajes al extranjero para reuniones o congresos, el tiempo de espera no debe rebasar las 48 horas.
Es muy importante que el alumno que solicite una respuesta directa del equipo docente lo haga constar en su mensaje al foro.

IGUALDAD DE GÉNERO

En coherencia con el valor asumido de la igualdad de género, todas las denominaciones que en esta Guía hacen referencia a órganos de gobierno unipersonales, de representación, o miembros de la comunidad universitaria y se efectúan en género masculino, cuando no se hayan sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.