Guía docente
270217 - AA1 - Aprendizaje Automático 1

Última modificación: 30/01/2024

Unidad responsable: Facultad de Informática de Barcelona
Unidad que imparte: 723 - CS - Departamento de Ciencias de la Computación.

Titulación: GRADO EN CIENCIA E INGENIERÍA DE DATOS (Plan 2017). (Asignatura obligatoria).

Curso: 2023 **Créditos ECTS:** 6.0 **Idiomas:** Catalán

PROFESORADO

Profesorado responsable: MARTA ARIAS VICENTE

Otros: Segon quadrimestre:
MARTA ARIAS VICENTE - 11, 12, 13
LUIS ANTONIO BELANCHE MUÑOZ - 13
ALEXIS MOLINA MARTINEZ DE LOS REYES - 11, 12

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:
CE1. Utilizar con destreza los conceptos y métodos matemáticos que subyacen los problemas de la ciencia y la ingeniería de los datos.
CE3. Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos de estos tipos en situaciones concretas. Formular y resolver problemas de optimización matemática.
CE8. Capacidad de elegir y emplear técnicas de modelización estadística y análisis de datos, evaluando la calidad de los modelos, validándolos e interpretándolos.
CE9. Capacidad de elegir y emplear una variedad de técnicas de aprendizaje automático y construir sistemas que las utilicen para la toma de decisiones, incluso de forma autónoma.

Genéricas:
CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.
CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.

Transversales:
CT3. Comunicación eficaz oral y escrita. Comunicarse de forma oral y escrita con otras personas sobre los resultados del aprendizaje, de la elaboración del pensamiento y de la toma de decisiones; participar en debates sobre temas de la propia especialidad.
CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT7. Tercera lengua. Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

METODOLOGÍAS DOCENTES

Las clases de teoría introducen todo los conocimientos, las técnicas, conceptos y resultados necesarios para alcanzar un nivel bien fundamentado y comprensible. Estos conceptos se ponen en práctica en las clases de laboratorio. En estas se proporciona código python que permite resolver ciertos aspectos de un problema de análisis de datos con las técnicas correspondientes al tema en curso. Este laboratorio también sirve de guía para la parte correspondiente de la práctica, que desarrollan los alumnos a lo largo del curso. Algunas de las horas de laboratorio se podrán usar para resolver problemas (sin ordenador) en el aula de teoría.

Hay un trabajo práctico evaluable, que trabaja un problema real a elegir por el propio estudiante y que recoge e integra los conocimientos y las competencias de todo el curso. También se evalúa mediante el trabajo práctico la competencia genérica de comunicación eficaz escrita.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

1.Formular el problema del aprendizaje automático a partir de datos, y conocer los tipos de tareas que se pueden dar.
2.Organizar el flujo de resolución de un problema del aprendizaje automático, analizando las posibles opciones y eligiendo las más adecuadas al problema.
3.Decidir, defender y criticar una solución para un problema de aprendizaje automático, argumentando los puntos fuertes y débiles del acercamiento.
4.Conocer y saber aplicar técnicas lineales para la resolución de problemas de aprendizaje supervisado.
5.Conocer y saber aplicar técnicas de redes neuronales mono y multicapa para la resolución de problemas de aprendizaje supervisado.
6.Conocer y saber aplicar técnicas de máquinas de vectores soporte para la resolución de problemas de aprendizaje supervisado.
7.Conocer y saber aplicar las técnicas básicas para la resolución de problemas de aprendizaje no supervisado, con énfasis en herramientas de agrupamiento de datos.
8.Conocer y saber aplicar las técnicas básicas para la resolución de problemas de aprendizaje por refuerzo.
9.Conocer y saber aplicar técnicas de conjunto para la resolución de problemas de aprendizaje supervisado.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo pequeño	30,0	20.00
Horas aprendizaje autónomo	90,0	60.00
Horas grupo grande	30,0	20.00

Dedicación total: 150 h

CONTENIDOS

Introducción al Aprendizaje Automático

Descripción:
Información general y conceptos básicos. Descripción y planteamiento de los problemas atacados por el aprendizaje automático. Aprendizaje supervisado (regresión y clasificación), no supervisado (clustering) y semi-supervisado (por refuerzo y transductivo). Ejemplos modernos de aplicación.

Aprendizaje automático no supervisado: clustering

Descripción:
Definición y planteamiento del aprendizaje automático no supervisado. Introducción al clustering. Algoritmos probabilísticos: K-means y algoritmo Expectation-Maximization (E-M).

Aprendizaje automático supervisado (I): métodos lineales de regresión

Descripción:
Máxima verosimilitud para regresión. Funciones de error para regresión. Mínimos cuadrados: métodos analíticos (pseudo-inversa y SVD) e iterativos (gradient descent). Noción de regularización. Regresión regularizada L1 y L2: algoritmos ridge regression, LASSO y Elastic Net.

Aprendizaje automático supervisado (II): métodos lineales para clasificación

Descripción:
Máxima verosimilitud para clasificación. Funciones de error para clasificación. Clasificadores generativos Bayesianos: LDA/QDA/RDA, Naïve Bayes y k-nearest neighbours.

Métodos jerárquicos: árboles de decisión

Descripción:
Construcción general de árboles de decisión. Criterios de split: ganancia en entropía y criterio de Gini. Regularización en árboles de decisión. Árboles CART para regresión y clasificación.

Métodos de conjunto (ensemble)

Descripción:
Introducción a los métodos de conjunto (ensemble). Bagging y Random Forests. Boosting. Adaboost y variantes.

Métodos basados en kernels

Descripción:
Introducción al aprendizaje con funciones de kernel. Regresión lineal regularizada kernelizada. Funciones de kernel básicas. Complejidad y generalización: dimensión de Vapnik-Chervonenkis. Máquinas de Vectores Soporte.

ACTIVIDADES

Desarrollo del tema 1 de la asignatura

Objetivos específicos:
1

Competencias relacionadas:
CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.
CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.
CE1. Utilizar con destreza los conceptos y métodos matemáticos que subyacen los problemas de la ciencia y la ingeniería de los datos.
CE9. Capacidad de elegir y emplear una variedad de técnicas de aprendizaje automático y construir sistemas que las utilicen para la toma de decisiones, incluso de forma autónoma.

Dedicación: 5h 18m
Grupo grande/Teoría: 2h
Aprendizaje autónomo: 3h 18m

Desarrollo del tema 2 de la asignatura

Objetivos específicos:
1, 3, 7

Competencias relacionadas:
CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.
CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.
CE1. Utilizar con destreza los conceptos y métodos matemáticos que subyacen los problemas de la ciencia y la ingeniería de los datos.
CE8. Capacidad de elegir y emplear técnicas de modelización estadística y análisis de datos, evaluando la calidad de los modelos, validándolos e interpretándolos.
CE9. Capacidad de elegir y emplear una variedad de técnicas de aprendizaje automático y construir sistemas que las utilicen para la toma de decisiones, incluso de forma autónoma.
CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT3. Comunicación eficaz oral y escrita. Comunicarse de forma oral y escrita con otras personas sobre los resultados del aprendizaje, de la elaboración del pensamiento y de la toma de decisiones; participar en debates sobre temas de la propia especialidad.

Dedicación: 11h 36m
Grupo grande/Teoría: 3h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 6h 36m

Desarrollo del tema 3 de la asignatura

Objetivos específicos:
1, 4

Competencias relacionadas:
CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.
CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.
CE1. Utilizar con destreza los conceptos y métodos matemáticos que subyacen los problemas de la ciencia y la ingeniería de los datos.
CE3. Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos de estos tipos en situaciones concretas. Formular y resolver problemas de optimización matemática.
CE8. Capacidad de elegir y emplear técnicas de modelización estadística y análisis de datos, evaluando la calidad de los modelos, validándolos e interpretándolos.
CE9. Capacidad de elegir y emplear una variedad de técnicas de aprendizaje automático y construir sistemas que las utilicen para la toma de decisiones, incluso de forma autónoma.

Dedicación: 20h
Grupo grande/Teoría: 8h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 10h

Desarrollo del tema 4 de la asignatura

Objetivos específicos:
1, 2, 4

Competencias relacionadas:
CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.
CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.
CE1. Utilizar con destreza los conceptos y métodos matemáticos que subyacen los problemas de la ciencia y la ingeniería de los datos.
CE3. Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos de estos tipos en situaciones concretas. Formular y resolver problemas de optimización matemática.
CE8. Capacidad de elegir y emplear técnicas de modelización estadística y análisis de datos, evaluando la calidad de los modelos, validándolos e interpretándolos.
CE9. Capacidad de elegir y emplear una variedad de técnicas de aprendizaje automático y construir sistemas que las utilicen para la toma de decisiones, incluso de forma autónoma.
CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT7. Tercera lengua. Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Dedicación: 16h 18m
Grupo grande/Teoría: 6h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 8h 18m

Desarrollo del tema 5+6 de la asignatura

Objetivos específicos:
1, 2, 5

Competencias relacionadas:
CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.
CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.
CE1. Utilizar con destreza los conceptos y métodos matemáticos que subyacen los problemas de la ciencia y la ingeniería de los datos.
CE8. Capacidad de elegir y emplear técnicas de modelización estadística y análisis de datos, evaluando la calidad de los modelos, validándolos e interpretándolos.
CE9. Capacidad de elegir y emplear una variedad de técnicas de aprendizaje automático y construir sistemas que las utilicen para la toma de decisiones, incluso de forma autónoma.
CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT7. Tercera lengua. Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Dedicación: 18h 36m
Grupo grande/Teoría: 5h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 11h 36m

Desarrollo del tema 7 de la asignatura

Objetivos específicos:
1, 2, 5

Competencias relacionadas:
CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.
CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.
CE1. Utilizar con destreza los conceptos y métodos matemáticos que subyacen los problemas de la ciencia y la ingeniería de los datos.
CE8. Capacidad de elegir y emplear técnicas de modelización estadística y análisis de datos, evaluando la calidad de los modelos, validándolos e interpretándolos.
CE9. Capacidad de elegir y emplear una variedad de técnicas de aprendizaje automático y construir sistemas que las utilicen para la toma de decisiones, incluso de forma autónoma.
CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT7. Tercera lengua. Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Dedicación: 13h
Grupo grande/Teoría: 6h
Grupo pequeño/Laboratorio: 2h
Aprendizaje autónomo: 5h

Control de la práctica

Objetivos específicos:
1, 2, 3, 4, 5, 6, 7, 8, 9

Competencias relacionadas:
CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.
CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.
CE1. Utilizar con destreza los conceptos y métodos matemáticos que subyacen los problemas de la ciencia y la ingeniería de los datos.
CE3. Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos de estos tipos en situaciones concretas. Formular y resolver problemas de optimización matemática.
CE8. Capacidad de elegir y emplear técnicas de modelización estadística y análisis de datos, evaluando la calidad de los modelos, validándolos e interpretándolos.
CE9. Capacidad de elegir y emplear una variedad de técnicas de aprendizaje automático y construir sistemas que las utilicen para la toma de decisiones, incluso de forma autónoma.
CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT7. Tercera lengua. Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.
CT3. Comunicación eficaz oral y escrita. Comunicarse de forma oral y escrita con otras personas sobre los resultados del aprendizaje, de la elaboración del pensamiento y de la toma de decisiones; participar en debates sobre temas de la propia especialidad.

Dedicación: 2h
Actividades dirigidas: 2h

Entrega de la práctica

Objetivos específicos:
1, 2, 3, 4, 5, 6, 7, 8, 9

Competencias relacionadas:
CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.
CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.
CE1. Utilizar con destreza los conceptos y métodos matemáticos que subyacen los problemas de la ciencia y la ingeniería de los datos.
CE3. Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos de estos tipos en situaciones concretas. Formular y resolver problemas de optimización matemática.
CE8. Capacidad de elegir y emplear técnicas de modelización estadística y análisis de datos, evaluando la calidad de los modelos, validándolos e interpretándolos.
CE9. Capacidad de elegir y emplear una variedad de técnicas de aprendizaje automático y construir sistemas que las utilicen para la toma de decisiones, incluso de forma autónoma.
CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT7. Tercera lengua. Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.
CT3. Comunicación eficaz oral y escrita. Comunicarse de forma oral y escrita con otras personas sobre los resultados del aprendizaje, de la elaboración del pensamiento y de la toma de decisiones; participar en debates sobre temas de la propia especialidad.

Dedicación: 3h
Actividades dirigidas: 3h

Seguiment i tutories de la pràctica

Descripción:
Seguiment i tutories de la pràctica

Objetivos específicos:
2, 3, 4, 6, 7, 9

Competencias relacionadas:
CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.
CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.
CE1. Utilizar con destreza los conceptos y métodos matemáticos que subyacen los problemas de la ciencia y la ingeniería de los datos.
CE3. Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos de estos tipos en situaciones concretas. Formular y resolver problemas de optimización matemática.
CE8. Capacidad de elegir y emplear técnicas de modelización estadística y análisis de datos, evaluando la calidad de los modelos, validándolos e interpretándolos.
CE9. Capacidad de elegir y emplear una variedad de técnicas de aprendizaje automático y construir sistemas que las utilicen para la toma de decisiones, incluso de forma autónoma.
CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT7. Tercera lengua. Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.
CT3. Comunicación eficaz oral y escrita. Comunicarse de forma oral y escrita con otras personas sobre los resultados del aprendizaje, de la elaboración del pensamiento y de la toma de decisiones; participar en debates sobre temas de la propia especialidad.

Dedicación: 26h
Grupo pequeño/Laboratorio: 6h
Aprendizaje autónomo: 20h

SISTEMA DE CALIFICACIÓN

La asignatura se evalúa mediante un examen parcial, un examen final y un trabajo práctico en el que se ataca un problema real, redactando el correspondiente informe.

La nota final se calcula como:

Nota Asignatura = 0,4 * Trabajo + 0,6 * max (Final, 1/3 * Parcial + 2/3 * Final)

Para los estudiantes que puedan y quieran concurrir a la reevaluación, la nota de examen de reevaluación sustituirá max (Final, 1/3 * Parcial + 2/3 * Final).

BIBLIOGRAFÍA

Básica:
- Bishop, C.M. Pattern recognition and machine learning. Springer, 2006. ISBN 0387310738.
- Murphy, K.P. Machine learning: a probabilistic perspective. Cambridge, Mass: MIT Press, 2012. ISBN 9780262018029.
- Cherkassky, V.S.; Mulier, F. Learning from data: concepts, theory, and methods. 2nd ed. John Wiley, 2007. ISBN 0471681822.
- Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Springer, 2009. ISBN 0387848576.