



Guia docent

270217 - AA1 - Aprenentatge Automàtic 1

Última modificació: 30/01/2024

Unitat responsable: Facultat d'Informàtica de Barcelona
Unitat que imparteix: 723 - CS - Departament de Ciències de la Computació.

Titulació: GRAU EN CIÈNCIA I ENGINYERIA DE DADES (Pla 2017). (Assignatura obligatòria).

Curs: 2023 **Crèdits ECTS:** 6.0 **Idiomes:** Català

PROFESSORAT

Professorat responsable: MARTA ARIAS VICENTE

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

CAPACITATS PRÈVIES

Nocions mitjanes de probabilitat i estadística.
Nocions mitjanes d'àlgebra lineal, càlcul matricial i anàlisi real
Bon nivell de programació en llenguatges d'alt nivell

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

CE1. Utilitzar amb destresa els conceptes i mètodes matemàtics subjacents als problemes de la ciència i l'enginyeria de les dades.
CE3. Analitzar fenòmens complexos mitjançant la probabilitat i l'estadística, i plantejar models d'aquests tipus en situacions concretes. Formular i resoldre problemes d'optimització matemàtica.
CE8. Capacitat de triar i emprar tècniques de modelització estadística i anàlisi de dades, avaluant la qualitat dels models, validant-los i interpretant-los.
CE9. Capacitat de triar i emprar una varietat de tècniques d'aprenentatge automàtic i construir sistemes que les utilitzin per a la presa de decisions, fins i tot de forma autònoma.

Genèriques:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.
CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

Transversals:

CT3. Comunicació eficaç oral i escrita. Comunicar-se de forma oral i escrita amb altres persones sobre els resultats de l'aprenentatge, de l'elaboració del pensament i de la presa de decisions; participar en debats sobre temes de la pròpia especialitat.
CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.
CT7. Tercera llengua. Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i d'acord amb les necessitats que tindran els titulats i titulades.



METODOLOGIES DOCENTS

Les classes de teoria introdueixen tot els coneixements, les tècniques, conceptes i resultats necessaris per assolir un nivell ben fonamentat i entenedor. Aquests conceptes es posen en pràctica en les classes de laboratori. En aquestes es proporciona codi Python que permet resoldre certs aspectes d'un problema d'anàlisi de dades amb les tècniques corresponents al tema en curs. Aquest laboratori també serveix de guia per la part corresponent de la pràctica, que desenvolupen els alumnes al llarg del curs. Algunes de les hores de laboratori es podran usar per resoldre problemes (sense ordinador) a l'aula de teoria.

Hi ha un treball pràctic avaluable, que treballa un problema real a elegir pel propi estudiant i que recull i integra els coneixements i les competències de tot el curs. També s'avalua mitjançant el treball pràctic la competència genèrica de comunicació eficaç escrita.

OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

1. Formular el problema de l'aprenentatge automàtic a partir de dades, i conèixer els tipus de tasques que es poden donar.
2. Organitzar el flux de resolució d'un problema de l'aprenentatge automàtic, analitzant les possibles opcions i elegint les més adequades al problema.
3. Decidir, defensar i criticar una solució per a un problema d'aprenentatge automàtic, argumentant els punts forts i febles de l'apropament.
4. Conèixer i saber aplicar tècniques lineals per la resolució de problemes d'aprenentatge supervisat.
5. Conèixer i saber aplicar tècniques de xarxes neuronals mono i multicapa per la resolució de problemes d'aprenentatge supervisat.
6. Conèixer i saber aplicar tècniques de màquines de vectors suport per la resolució de problemes d'aprenentatge supervisat.
7. Conèixer i saber aplicar les tècniques bàsiques per la resolució de problemes d'aprenentatge no supervisat, amb èmfasi en eines d'agrupament de dades.
8. Conèixer i saber aplicar les tècniques bàsiques per la resolució de problemes d'aprenentatge per reforç.
9. Conèixer i saber aplicar tècniques de conjunt per la resolució de problemes d'aprenentatge supervisat.

HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANT

Tipus	Hores	Percentatge
Hores grup gran	30,0	20.00
Hores aprenentatge autònom	90,0	60.00
Hores grup petit	30,0	20.00

Dedicació total: 150 h

CONTINGUTS

Introducció a l'Aprenentatge Automàtic

Descripció:

Informació general i conceptes bàsics. Descripció i plantejament dels problemes atacats per l'aprenentatge automàtic. Aprenentatge supervisat (regressió i classificació), no supervisat (clustering) i semi-supervisat (per reforç i transductiu). Exemples moderns d'aplicació.

Aprenentatge automàtic no supervisat: clustering

Descripció:

Definició i plantejament de l'aprenentatge automàtic no supervisat. Introducció al clustering. Algorismes probabilístics: algorisme k-means i algorisme Expectation-Maximization (E-M).



Aprenentatge automàtic supervisat (I): mètodes lineals per regressió

Descripció:

Màxima versemblança per regressió. Funcions d'error per regressió. Mínims quadrats: mètodes analítics (pseudo-inversa i SVD) i iteratius (gradient descent). Noció de regularització. Regressió regularitzada L1 i L2: algorismes ridge regressió, LASSO i Elastic Net.

Aprenentatge automàtic supervisat (II): mètodes lineals per classificació

Descripció:

Màxima versemblança per classificació. Funcions d'error per classificació. Classificadors generatius Baiesians: LDA/QDA/RDA, Naïve Bayes i k-nearest neighbours.

Mètodes jeràrquics: arbres de decisió

Descripció:

Construcció general d'arbres de decisió. Criteris d'split: guany en entropia i criteri de Gini. Regularització en arbres de decisió. Arbres CART per regressió i classificació.

Mètodes de conjunt (ensemble)

Descripció:

Introducció als mètodes de conjunt (ensemble). Bagging i Random Forests. Boosting. Adaboost i variants.

Mètodes basats en kernels

Descripció:

Introducció a l'aprenentatge amb funcions de kernel. Regressió lineal regularitzada kernelitzada. Funcions de kernel bàsiques. Complexitat i generalització: dimensió de Vapnik-Chervonenkis. Màquines de Vectors Suport.

ACTIVITATS

Desenvolupament del tema 1 de l'assignatura

Objectius específics:

1

Competències relacionades:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CE1. Utilitzar amb destresa els conceptes i mètodes matemàtics subjacents als problemes de la ciència i l'enginyeria de les dades.

CE9. Capacitat de triar i emprar una varietat de tècniques d'aprenentatge automàtic i construir sistemes que les utilitzin per a la presa de decisions, fins i tot de forma autònoma.

Dedicació: 5h 18m

Grup gran/Teoria: 2h

Aprenentatge autònom: 3h 18m



Desenvolupament del tema 2 de l'assignatura

Objectius específics:

1, 3, 7

Competències relacionades:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CE1. Utilitzar amb destresa els conceptes i mètodes matemàtics subjacents als problemes de la ciència i l'enginyeria de les dades.

CE8. Capacitat de triar i emprar tècniques de modelització estadística i anàlisi de dades, avaluant la qualitat dels models, validant-los i interpretant-los.

CE9. Capacitat de triar i emprar una varietat de tècniques d'aprenentatge automàtic i construir sistemes que les utilitzin per a la presa de decisions, fins i tot de forma autònoma.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT3. Comunicació eficaç oral i escrita. Comunicar-se de forma oral i escrita amb altres persones sobre els resultats de l'aprenentatge, de l'elaboració del pensament i de la presa de decisions; participar en debats sobre temes de la pròpia especialitat.

Dedicació: 11h 36m

Grup gran/Teoria: 3h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 6h 36m

Desenvolupament del tema 3 de l'assignatura

Objectius específics:

1, 4

Competències relacionades:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CE1. Utilitzar amb destresa els conceptes i mètodes matemàtics subjacents als problemes de la ciència i l'enginyeria de les dades.

CE3. Analitzar fenòmens complexos mitjançant la probabilitat i l'estadística, i plantejar models d'aquests tipus en situacions concretes. Formular i resoldre problemes d'optimització matemàtica.

CE8. Capacitat de triar i emprar tècniques de modelització estadística i anàlisi de dades, avaluant la qualitat dels models, validant-los i interpretant-los.

CE9. Capacitat de triar i emprar una varietat de tècniques d'aprenentatge automàtic i construir sistemes que les utilitzin per a la presa de decisions, fins i tot de forma autònoma.

Dedicació: 20h

Grup gran/Teoria: 8h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 10h



Desenvolupament del tema 4 de l'assignatura

Objectius específics:

1, 2, 4

Competències relacionades:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CE1. Utilitzar amb destresa els conceptes i mètodes matemàtics subjacents als problemes de la ciència i l'enginyeria de les dades.

CE3. Analitzar fenòmens complexos mitjançant la probabilitat i l'estadística, i plantejar models d'aquests tipus en situacions concretes. Formular i resoldre problemes d'optimització matemàtica.

CE8. Capacitat de triar i emprar tècniques de modelització estadística i anàlisi de dades, avaluant la qualitat dels models, validant-los i interpretant-los.

CE9. Capacitat de triar i emprar una varietat de tècniques d'aprenentatge automàtic i construir sistemes que les utilitzin per a la presa de decisions, fins i tot de forma autònoma.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT7. Tercera llengua. Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i d'acord amb les necessitats que tindran els titulats i titulades.

Dedicació: 16h 18m

Grup gran/Teoria: 6h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 8h 18m

Desenvolupament del tema 5 i 6 de l'assignatura

Objectius específics:

1, 2, 5

Competències relacionades:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CE1. Utilitzar amb destresa els conceptes i mètodes matemàtics subjacents als problemes de la ciència i l'enginyeria de les dades.

CE8. Capacitat de triar i emprar tècniques de modelització estadística i anàlisi de dades, avaluant la qualitat dels models, validant-los i interpretant-los.

CE9. Capacitat de triar i emprar una varietat de tècniques d'aprenentatge automàtic i construir sistemes que les utilitzin per a la presa de decisions, fins i tot de forma autònoma.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT7. Tercera llengua. Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i d'acord amb les necessitats que tindran els titulats i titulades.

Dedicació: 18h 36m

Grup gran/Teoria: 5h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 11h 36m



Desenvolupament del tema 7 de l'assignatura

Objectius específics:

1, 2, 5

Competències relacionades:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CE1. Utilitzar amb destresa els conceptes i mètodes matemàtics subjacents els problemes de la ciència i l'enginyeria de les dades.

CE8. Capacitat de triar i emprar tècniques de modelització estadística i anàlisi de dades, avaluant la qualitat dels models, validant-los i interpretant-los.

CE9. Capacitat de triar i emprar una varietat de tècniques d'aprenentatge automàtic i construir sistemes que les utilitzin per a la presa de decisions, fins i tot de forma autònoma.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT7. Tercera llengua. Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i d'acord amb les necessitats que tindran els titulats i titulades.

Dedicació: 13h

Grup gran/Teoria: 6h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 5h

Control de la pràctica

Objectius específics:

1, 2, 3, 4, 5, 6, 7, 8, 9

Competències relacionades:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CE1. Utilitzar amb destresa els conceptes i mètodes matemàtics subjacents els problemes de la ciència i l'enginyeria de les dades.

CE3. Analitzar fenòmens complexos mitjançant la probabilitat i l'estadística, i plantejar models d'aquests tipus en situacions concretes. Formular i resoldre problemes d'optimització matemàtica.

CE8. Capacitat de triar i emprar tècniques de modelització estadística i anàlisi de dades, avaluant la qualitat dels models, validant-los i interpretant-los.

CE9. Capacitat de triar i emprar una varietat de tècniques d'aprenentatge automàtic i construir sistemes que les utilitzin per a la presa de decisions, fins i tot de forma autònoma.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT7. Tercera llengua. Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i d'acord amb les necessitats que tindran els titulats i titulades.

CT3. Comunicació eficaç oral i escrita. Comunicar-se de forma oral i escrita amb altres persones sobre els resultats de l'aprenentatge, de l'elaboració del pensament i de la presa de decisions; participar en debats sobre temes de la pròpia especialitat.

Dedicació: 2h

Activitats dirigides: 2h



Lliurament de la pràctica

Objectius específics:

1, 2, 3, 4, 5, 6, 7, 8, 9

Competències relacionades:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CE1. Utilitzar amb destresa els conceptes i mètodes matemàtics subjacents als problemes de la ciència i l'enginyeria de les dades.

CE3. Analitzar fenòmens complexos mitjançant la probabilitat i l'estadística, i plantejar models d'aquests tipus en situacions concretes. Formular i resoldre problemes d'optimització matemàtica.

CE8. Capacitat de triar i emprar tècniques de modelització estadística i anàlisi de dades, avaluant la qualitat dels models, validant-los i interpretant-los.

CE9. Capacitat de triar i emprar una varietat de tècniques d'aprenentatge automàtic i construir sistemes que les utilitzin per a la presa de decisions, fins i tot de forma autònoma.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT7. Tercera llengua. Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i d'acord amb les necessitats que tindran els titulats i titulades.

CT3. Comunicació eficaç oral i escrita. Comunicar-se de forma oral i escrita amb altres persones sobre els resultats de l'aprenentatge, de l'elaboració del pensament i de la presa de decisions; participar en debats sobre temes de la pròpia especialitat.

Dedicació: 3h

Activitats dirigides: 3h



Seguiment i tutories de la pràctica

Descripció:

Seguiment i tutories de la pràctica

Objectius específics:

2, 3, 4, 6, 7, 9

Competències relacionades:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CE1. Utilitzar amb destresa els conceptes i mètodes matemàtics subjacents als problemes de la ciència i l'enginyeria de les dades.

CE3. Analitzar fenòmens complexos mitjançant la probabilitat i l'estadística, i plantejar models d'aquests tipus en situacions concretes. Formular i resoldre problemes d'optimització matemàtica.

CE8. Capacitat de triar i emprar tècniques de modelització estadística i anàlisi de dades, avaluant la qualitat dels models, validant-los i interpretant-los.

CE9. Capacitat de triar i emprar una varietat de tècniques d'aprenentatge automàtic i construir sistemes que les utilitzin per a la presa de decisions, fins i tot de forma autònoma.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT7. Tercera llengua. Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i d'acord amb les necessitats que tindran els titulats i titulades.

CT3. Comunicació eficaç oral i escrita. Comunicar-se de forma oral i escrita amb altres persones sobre els resultats de l'aprenentatge, de l'elaboració del pensament i de la presa de decisions; participar en debats sobre temes de la pròpia especialitat.

Dedicació: 26h

Grup petit/Laboratori: 6h

Aprenentatge autònom: 20h

SISTEMA DE QUALIFICACIÓ

L'assignatura s'avalua mitjançant un examen parcial, un examen final i un treball pràctic en què s'ataca un problema real, redactant el corresponent informe.

La nota final es calcula com:

$$\text{Nota Assignatura} = 0,4 \cdot \text{Treball} + 0,6 \cdot \max(\text{Final}, 1/3 \cdot \text{Parcial} + 2/3 \cdot \text{Final})$$

Pel estudiants que puguin i vulguin concórrer a la re-avaluació, la nota de examen de reavaluació substituirà $\max(\text{Final}, 1/3 \cdot \text{Parcial} + 2/3 \cdot \text{Final})$.

BIBLIOGRAFIA

Bàsica:

- 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.