Examen APC

# Bloc 1: Introducció

## Defineix Aprenentatge Computacional i dona exemples d’aplicacions

L’Aprenentatge Computacional és la ciència que fa que els computadors aprenguin i actuïn com ho fan els humans. A més, milloren el seu aprenentatge amb el temps i a mida que van fent proves. La forma de aprendre d’aquests computadors és exposar-los a dades i informació en forma d’observacions i de interacció amb el mon real. Dintre de l’aprenentatge computacional hi ha 4 grans tipus:

* **Aprenentatge Supervisat:** En aquesta classe, els data scientists apliquen els algorismes amb les dades etiquetades i definint les variables les quals es vol que l’algorisme accedeixi per fer les correlacions. L’input i l’output de l’algorisme són especificats.

**Exemples:** Classificadors i models regressors.

* **Aprenentatge No Supervisat:** Aquest tipus d’aprenentatge computacional implica algorismes que s’entrenen amb dades sense etiquetar. L’algorisme busca entre les dades alguna connexió útil. Les dades amb les que l’algorisme s’entrena i les prediccions o recomanacions que allibera son predeterminades.

**Exemples:** Detecció d’anomalies, Reducció de la dimensionalitat

* **Aprenentatge Semi-Supervisat:** Aquest tipus d’aprenentatge involucra un mix dels dos tipus previs. El data scientist pot ficar la majoria de dades d’entrenament etiquetades però el model és totalment lliure d’explorar les dades pel seu compte i desenvolupar el seu propi enteniment sobre el data set.

**Exemples:** Traducció computacional, Etiquetar dades.

* **Aprenentatge Reforçat:** És típicament utilitzat per ensenyar a una maquina com completar un procés amb molts nivells pels quals hi ha unes regles ben definides. Els data scientists programen un algorisme per a completar la tasca i donar senyals positives i negatives sobre com completar la tasca. En gran mesura però, l’algorisme decideix pel seu compte quins nivells va escollint.

**Exemples:** Robòtica, Gameplay en videojocs.

Per tal d’escollir el millor model d’aprenentatge computacional, hi ha un procés per facilitar-nos el treball:

1. Ajuntar el problema amb les dades d’entrada que s’haurien de considerar per la solució. En aquest procés es necessita un gran entendriment del problema.
2. Recollir dades, transformar-les i etiquetar-les si es necessari.
3. Escollir quin algorisme s’ha d’utilitzar i testejar-lo per veure com rendeix
4. Canviar petits detalls i anar comprovant l’output per tal d’aconseguir un nivell acceptable d’exactitud.

## Descriu els diferents tipus d’aprenentatge computacional

Dintre de l’aprenentatge computacional hi ha 4 grans tipus:

* **Aprenentatge Supervisat:** En aquesta clase, els data scientists apliquen els algorismes amb les dades etiquetades i definint les variables les quals es vol que l’algorisme accedeixi per fer les correlacions. L’input i l’output de l’algorisme són especificats. Aquest algorisme va canviant com opera amb les dades i quins pesos atribueix a les dades fins que fa prediccions acurades al donar-li dades noves.

**Exemples:** Classificadors i models regressors.

* **Aprenentatge No Supervisat:** Aquest tipus d’aprenentatge computacional implica algorismes que s’entrenen amb dades sense etiquetar. L’algorisme busca entre les dades alguna connexió útil. Les dades amb les que l’algorisme s’entrena i les prediccions o recomanacions que allibera son predeterminades.

**Exemples:** Detecció d’anomalies, Reducció de la dimensionalitat

* **Aprenentatge Semi-Supervisat:** Aquest tipus d’aprenentatge involucra un mix dels dos tipus previs. El data scientist pot ficar la majoria de dades d’entrenament etiquetades però el model és totalment lliure d’explorar les dades pel seu compte i desenvolupar el seu propi enteniment sobre el data set.

**Exemples:** Traducció computacional, Etiquetar dades.

* **Aprenentatge Reforçat:** És típicament utilitzat per ensenyar a una maquina com completar un procés amb molts nivells pels quals hi ha unes regles ben definides. Els data scientists programen un algorisme per a completar la tasca i donar senyals positives i negatives sobre com completar la tasca. En gran mesura però, l’algorisme decideix pel seu compte quins nivells va escollint.

**Exemples:** Robòtica, Gameplay en videojocs.

## Defineix i diferencia entre Classificació i Regressió

* **Classificació:** S’encarrega de predir una etiqueta discreta Y. Classificació es el problema d’assignar noves observacions a la classe que s’hi assembli més, basada en el model de classificació construït a partir del model amb les dades d’entrenament.

L’exactitud de les classificacions dependrà de l’efectivitat de l’algorisme que s’escull, de com s’aplica i de quant valuoses son les dades que s’han utilitzat per entrenar el model.

Dintre dels classificadors existeixen 2 tipus. El primer seria el **regressor logístic.** Aquest model té com a sortida la probabilitat d’una variable categòrica objectiu d’una classe especifica. Normalment és de 0 o 1 (binaria), però es pot utilitzar per classificar entre moltes categories.

Després vindrien els **Vectors de Suport (SVM).** Soluciona el mateix problema que la regressió logística (classificació amb dos classes), però l’algorisme es presenta en forma geomètrica. Els SVM es basen en definir línies geomètriques que defineixen la frontera entre una classe o una altra. Una mida que es fa servir en els SVM és el marge, que és la distancia entre la línia i els punts més pròxims de les dos classes. Els SVM s’encarreguen de maximitzar aquest marge.

* **Regressió:** S’encarrega de predir una variable objectiu continua Y. Permet estimar un valor basat en una dada d’entrada X. En aquest model la variable objectiu és la variable desconeguda que es vol predir, sent continua (no hi ha forats al valor on Y es pot trobar). Variables discretes serien variables que nomes poden prendre valors exactes. La data d’entrada X inclou tota la informació rellevant sobre el data set que es pot utilitzar per predir l’objectiu. Aquests atributs s’anomenen característiques (features), que poden ser numèrics o categòrics. A la teva regressió voldràs les màximes observacions d’entrenament com sigui possible dels teus features per així que el teu model estigui ben entrenat per saber la relació entre X i Y. Les dades es separen en train i test, on el train té etiquetes per saber com entrenar-se el model. El test no té aquestes etiquetes, ja que serveixen per predir la variable Y i comparar-lo amb el resultat real d’aquest valor Y. Aquest pas és important ja que el nostre model ha de saber desenvolupar-se en tots els terrenys, sobretot en aquells en els que encara no ha estat. Dintre de la regressió es poden trobar 3 fases: La **regressió linial** s’encarrega de dibuixar la línia predictora on depenent de X estarà la nostra variable de sortida Y. Despres vindria el **Descens de Gradient**, que s’encarrega de buscar la pèrdua mínima en la nostra funció del model fent iteracions, obtenint cada vegada una millor aproximació al resultat. L’**Overfitting** és el resultat de fer massa aproximació al problema conegut, sense pensar en una possible situació nova o imprevista. El nostre model ha entrenat tant unes dades en concret que la línia de regressió que no es podrà adaptar a nous paràmetres. L’**Underfitting** és tot el contrari, hem entrenat massa poc el nostre model i no tindrà uns valors que s’adeqüin als resultats reals.
* **Diferències:**
  + Classificació = Predir una classe discreta / Regressió = Predir una quantitat continua
  + Un classificador pot predir un valor continu però es un forma de probabilitat per l’etiqueta de classe / Un regressor pot predir un valor discret però en forma d’un enter (quantitat)
  + Les prediccions dels classificadors es poden evaluar utilitzant l’exactitud, mentre que els regressors no / Els regressors es poden evaluar amb el RootMSE mentre que els classificadors no

## Defineix i diferencia entre L’IA simbòlica i la subsimbòlica

Una IA és simbòlica quan es desenvolupa una intel·ligència que es basa en regles i coneixement interpretades. Aquesta interpretació és òbviament humana, i tots ho podem entendre. Aquestes IAs simbòliques tenen un grau d’enteniment molt gran, ja que som nosaltres mateixos qui ho dissenyem i qui ho interpretem, per tant la maquina esta treballant amb dades “humanes”. Aquest simbolisme es pot expressar en les pròpies etiquetes que es donen a les variables, a la representació gràfica que es dona a un resultat o a un conjunt d’entrada de dades o a una interfície gràfica amb la que ens ajudem per controlar el programa.

En canvi, una IA subsimbòlica no manipulen representacions simbòliques per tal de trobar una solució als problemes, sinó que fan càlculs segons principis que la IA mateixa ha pogut demostrar per tal de arribar a una solució. Els conceptes estan representats com a un set de nombres, vectors, matrius o tensors. S’adrecen a un problema d’una forma local, tenint en compte un concepte únic i particular del problema a solucionar.

Punts a tenir en compte a l’hora de comparar una IA simbòlica d’una subsimbolica:

* Una IA simbòlica implica que tots els seus nivells estan basats en representacions llegibles simbòliques humanes on s’utilitza la lògica i la cerca per solucionar el problema.
* Un punt fort d’una IA simbòlica es que el procés d’enteniment pot ser fàcilment entenedor. En una representació simbòlica es pot explicar fàcilment com s’ha arribat a una certa conclusió i quines passes raonables s’han pres.
* Un punt molt fluix d’una representació subsimbòlica és que és molt difícil trobar una lògica o una causa de perquè s’ha arribat a un resultat en concret. És un tema delicat ja que redueix de gran forma el seu manteniment o la seva modificació.
* Un punt molt feble d’una representació simbòlica és que al procés d’aprenentatge les regles i el coneixement ha de ser codificat pel programador, cosa que pot ser un problema.
* Alguns sistemes subsimbòlics no poden prendre decisions d’alt risc.
* Les representacions simbòliques han estat limitades al mon acadèmic i als laboratoris universitaris en recerques petites amb els gegants industrials.

# Bloc 2.1: Regressió de Dades Numèriques

## Defineix el Descens del Gradient en un problema de Regressió Lineal

El descens del gradient és l’algorisme més utilitzat en aprenentatge computacional. Dins del descens de gradient s’inclou una funció de cost que es la que volem minimitzar. Aquesta funció de cost s’utilitza per monitoritzar els errors en les prediccions del nostre model.