

1 Descripció

Un problema fonamental en aprenentatge no supervisat és determinar K , el nombre de grups (clusters). La resposta no és fàcil perquè l'estructura dels grups pot ser jeràrquica i per tant el millor valor de K pot dependre del nivell de descomposició desitjat. És a dir, pot haver-hi més d'un bon valor de k .

En aquesta pràctica heu de trobat el millor valor de K sobre conjunts de dades diferents. Heu de fer servir,

- Un mètode per trobar el millor valor de K basat en dos passos.
 1. Trobar el millor agrupament per a valors de K fixats mitjançant un mètode de la família K -means.
 2. Triar, entre aquests, el millor valor de K fent servir algun criteri estadístic. Podeu fer servir algun dels índexs vistos a la sessió de laboratori corresponent.
- Mètodes de reducció de dimensionalitat.
- Validació visual del millor valor de K tal com s'ha explicat a la classe de teoria.

Heu de determinar el millor valor de k per als següents conjunts de dades amb un nombre variable de dimensions:

- Un sol atribut. Veure carpeta **dades/1D**.
- Molts atributs. Heu de fer servir el “Breast Cancer Wisconsin dataset”, format per 569 mostres de biòpsies, cadascuna amb 30 atributs sobre les característiques del nucli cel·lular. Tot i que les mostres estan classificades en malignes o benignes, suggerint que el millor valor de K és 2, cal que exploreu agrupaments significatius que no tenen perquè ser binaris. Podeu carregar el conjunt de dades amb els següent fragment de codi:

```
from sklearn.datasets import load_breast_cancer
import pandas as pd

cancer = load_breast_cancer()
cancer_df = pd.DataFrame(data=cancer.data, columns=cancer.feature_names)
cancer_df['target'] = cancer.target

print(cancer_df.head())
```

1.1 Tasques

Heu de preparar un informe seguint l'estructura IMRaD. La secció de resultats de l'informe ha d'incloure una taula resum que indiqui, per a cada conjunt de dades, quin és el millor valor de K (o els millors valors de K). El vostre informe ha d'incloure figures que validin o qüestionin els millors valors de K per a cada conjunt de dades.

2 Lliurament

Haureu de lliurar a través del Racó (secció “Pràctiques”) la vostra solució a la pràctica abans de la data indicada al mateix Racó.

La pràctica és per parelles. Només cal que un membre lliuri la solució pel Racó, això sí, indiqueu clarament els noms de les persones autores del projecte a l'informe.

1. El codi que hàgiu utilitzat. Dins una carpeta "codi".
2. Un informe en PDF. L'informe ha de ser breu (3-4 pàgines) i ha de detallar quina metodologia heu seguit, resultats, dificultats que us heu trobat i qualsevol cosa que penseu que hagi estat rellevant a l'hora de realitzar el projecte. Si us cal més espai, podeu fer servir apèndixs per a la informació complementària.

Algunes persones decideixen no formar una parella i treballar en solitari. Els motius són variats:

- No han trobat cap parella. Resposta: si els alumnes no han establert un canal de comunicació per formar parella és un problema seu pel qual seran avaluats. No és cap problema del professor. És habitual que els alumnes que no van a classe tinguin dificultats per trobar parella, per això la pràctica també avalua indirectament l'assistència a classe o saber suplir adequadament la no assistència a classe. **No és responsabilitat del professor ajudar els alumnes a buscar parella.** A la pràctica també s'avalua la capacitat per saber fer servir les xarxes socials. Els alumnes poden decidir pel seu compte i risc no estar connectats amb la resta d'estudiants de la classe mitjançant les xarxes socials.
- No saben o no els agrada treballar en grup. Resposta: a la pràctica s'avalua la capacitat per treballar en equip.
- Treballen millor sols. Resposta: demostren que han d'aprendre a aprendre a treballar en equip.
- No poden treballar en equip. Resposta: fals, fins i tot es pot col·laborar en línia. Si no tenen temps per treballar a la pràctica llavors no es poden contradir lliurant la pràctica. La realitat de moltes empreses avui en dia és el teletreball, i això implica fer servir eines de treball col·laboratiu. L'objectiu d'aquest màster és preparar per al món laboral. No poder treballar en equip és equivalent a no estar preparat per al món laboral.

Alguns alumnes d'aquesta mena escriuen al professor demanant-li que els deixi lliurar en solitari o unir-se a una parella per formar un trio amb l'excusa que no han aconseguit trobar parella o perquè ja és massa tard trobar-la. Aquestes excuses no serveixen (veure respostes més amunt). El truc és antic. El professor és vell. Per això, els criteris d'avaluació següents.

Es pot donar que hi hagi un nombre imparell d'alumnes que realment vulguin lliurar la pràctica. **Aleshores, la nota de la pràctica es calcula de la manera següent:**

- Si es tracta d'una parella, la nota és el resultat d'avaluar el contingut de la pràctica.
- **Si és un equip és més de dues persones la nota és un zero directe.**
- Si es tracta d'una estudianta en solitari, la nota és el resultat d'avaluar el contingut del lliurament si es tracta de l'únic lliurament en solitari. **Si hi ha més d'un lliurament en solitari, la nota és un zero directe per a cada estudiant que lliura en solitari.**