

PRÀCTICA 2 CLASSIFICACIÓ

Oscar Moreno Ramos
Rubén Ramos Segarra
Sergi Diaz Lopez



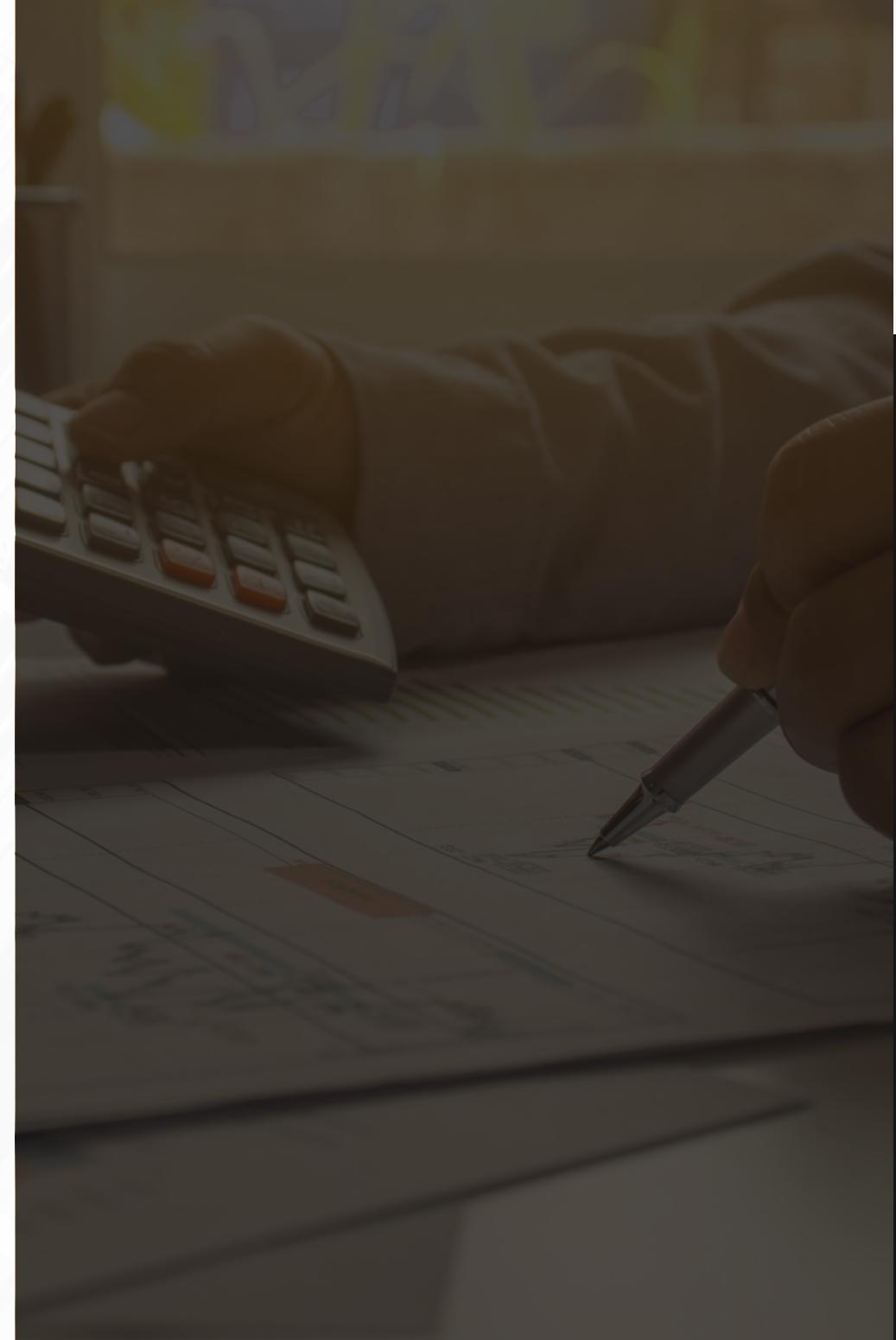


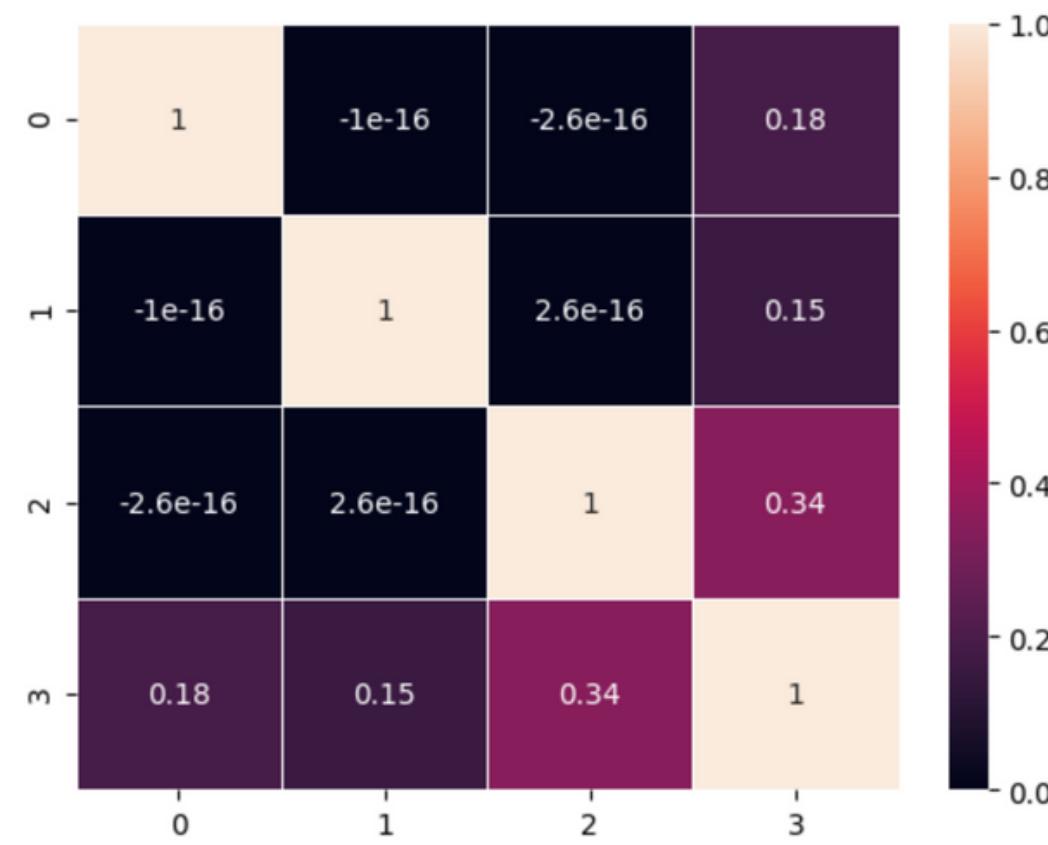
INTRODUCCIÓ

- Explorating data
- Preprocessing
- Model Selection
- Crossvalidation
- Metric Analysis

EXPLORING DATA: WATER QUALITY

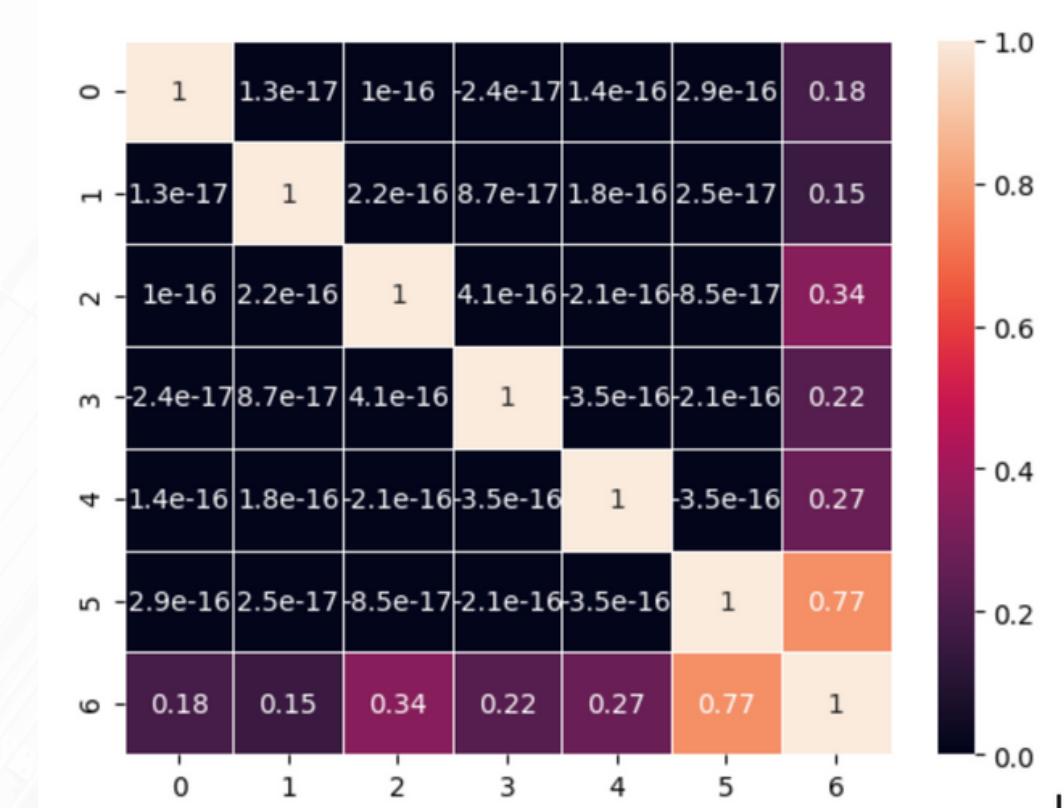
- Base de dades
- Tipus d'atributs
- Atribut objectiu: l'aigua és potable?





Correlació amb PCA 3 elements

Millorem la regressió logística fins a 0.708



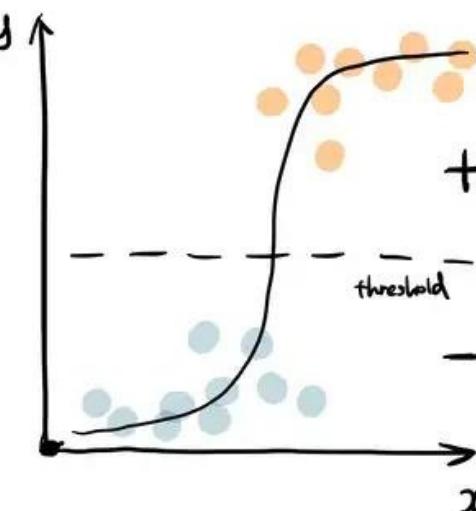
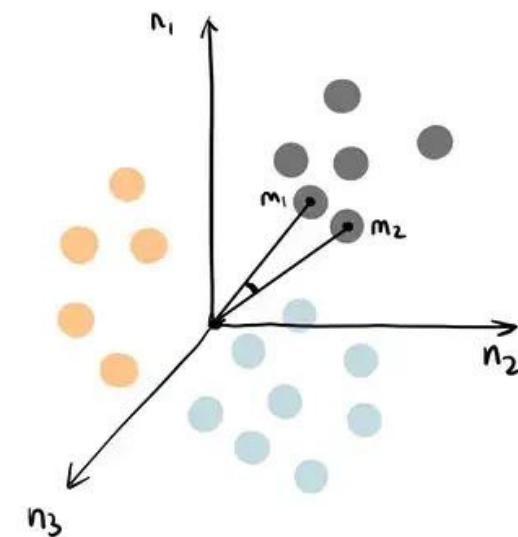
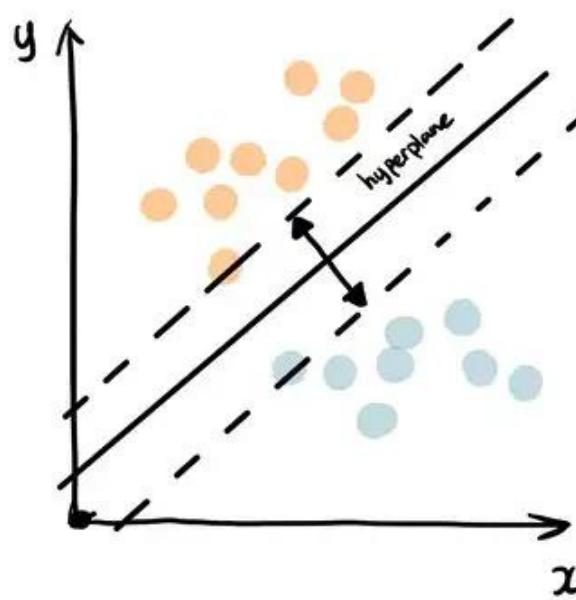
Correlació amb PCA 6 elements

Millorem la regressió logística fins a 1

PREPROCESSING

**CORRELACIÓ
NORMALITZACIÓ
PCA**

MODEL SELECTION



REGRESSIÓ LOGÍSTICA

Aconseguim els millors resultats possibles, ideal per a separar dades binàries. Sense tenir en compte preprocessing arribem al 0.63.

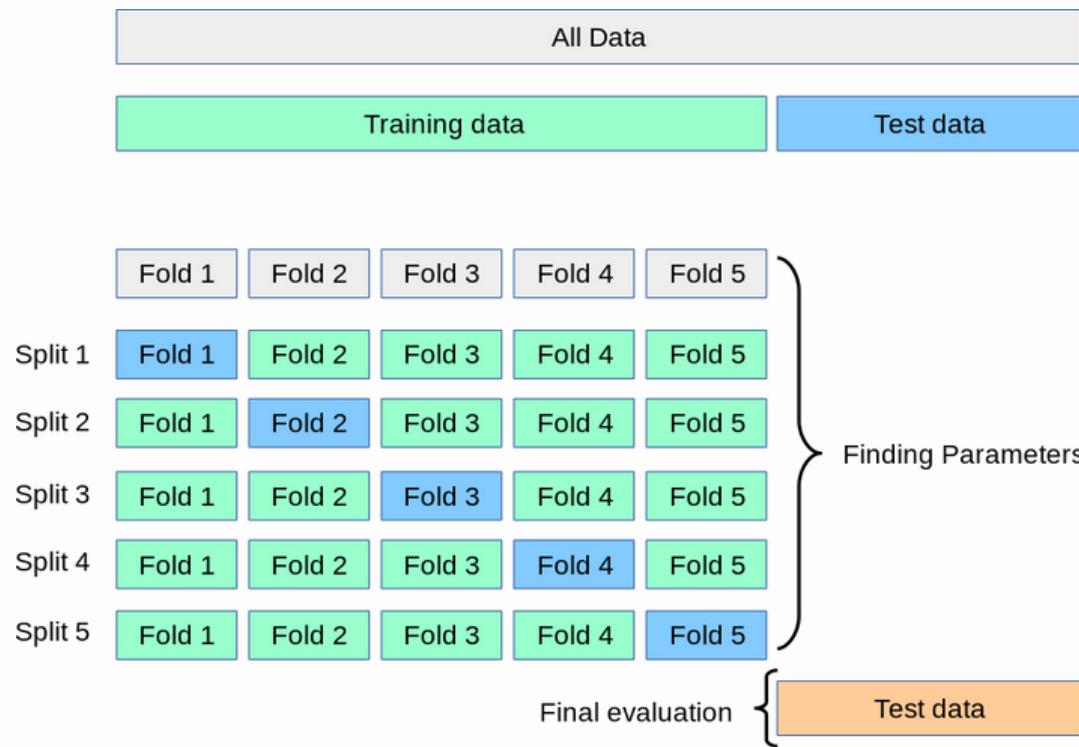
SVM

Aconseguim molt bon resultats, ideal per a separar dades de dues classes similars. Dades semblants a la regressió logística, de fet amb el kernel poli trobem les mateixes dades.

KNN

Busca trobar la classificació mitjançant els veïns més propers, millor cas amb 2 neighbors i 0.598

K-FOLD



Trobem el millor Score amb K=8

Score: 0.567

LEAVE ONE OUT

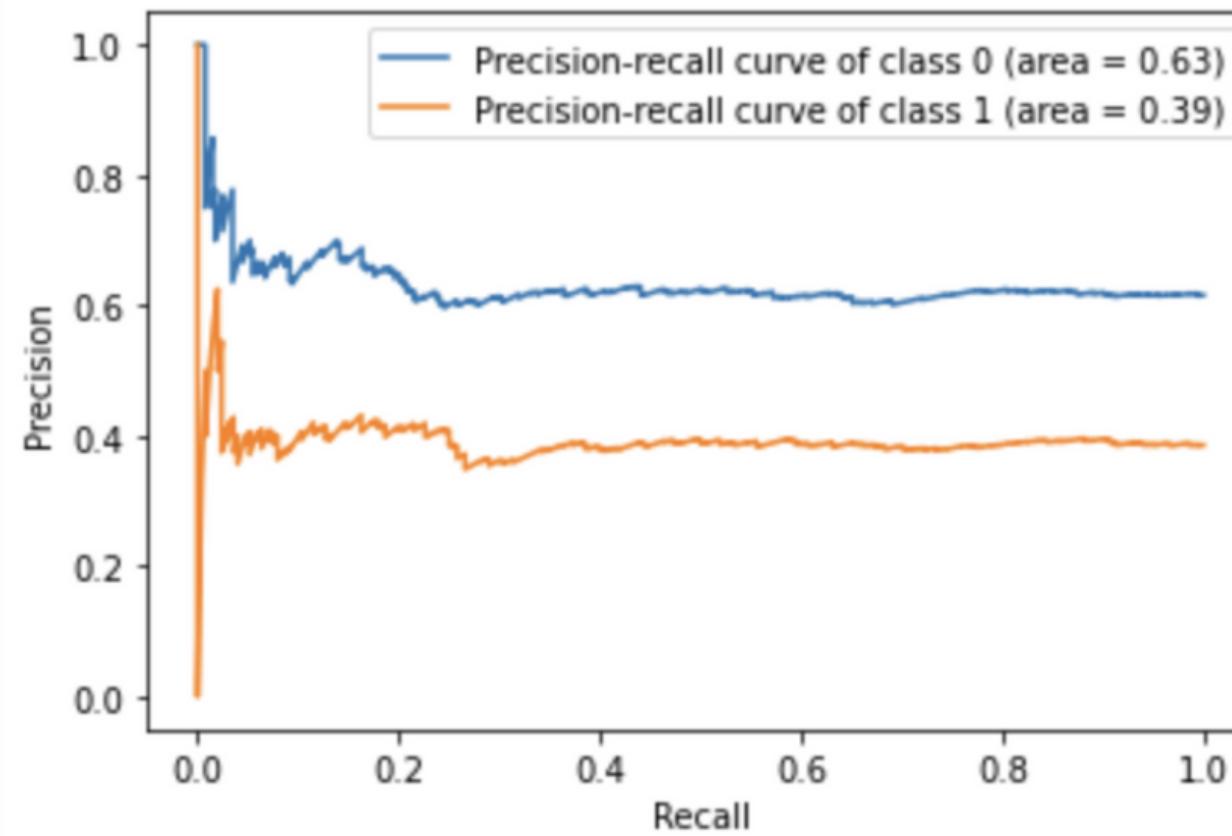
Leave-One-Out Cross Validation



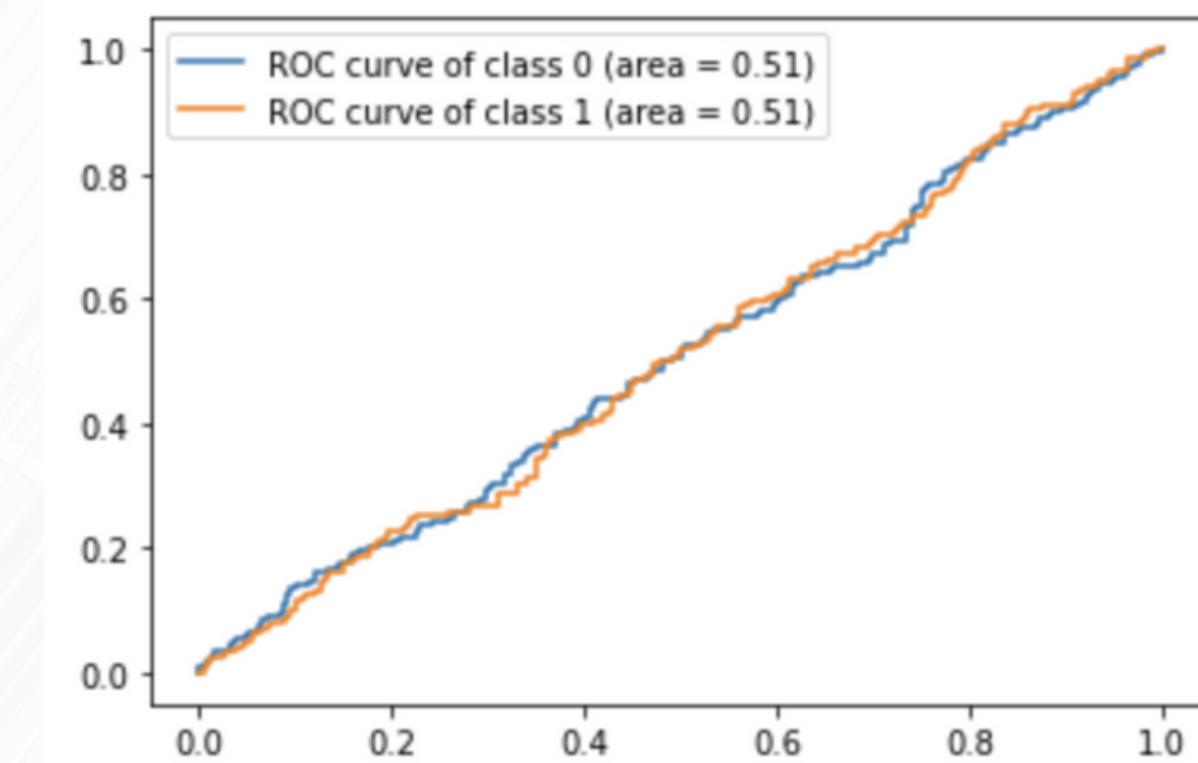
Score: 0.590

CROSSVALIDATION

PRECISION-RECALL



ROC-CURVE



METRIC ANALYSIS



**TENIU ALGUNA
PREGUNTA?**



**GRÀCIES PER LA
VOSTRA ATENCIÓ**