



TIPOLOGIA I CICLE DE VIDA DE LES DADES

PAC1: Què son les dades i quin el seu cicle de vida?

Marc Cervera Rosell

Semestre: febrer 2025 - juny 2025

Màster en ciència de dades

Índex

Exercici 1	1
Pregunta 1	1
Pregunta 2	1
Pregunta 3	1
Pregunta 4	1
Pregunta 5	1
Pregunta 6	1
Pregunta 7	1

Exercici 1

Pregunta 1

Per aplicar la tècnica de normalització caldria, en primer lloc, seleccionar aquelles variables numèriques a les quals volem aplicar la tècnica i després ajustar-les a una escala més reduïda, per tal de, fer-les més comprensibles. Típicament, les escales són entre -1.0 i 1.0 o entre 0.0 i 1.0.

Per aplicar la tècnica de discretització cal, com en el cas anterior, seleccionar aquelles variables numèriques a les quals volem aplicar la tècnica. Seguidament, cal substituir els valors numèrics per una sèrie d'etiquetes les quals poden ser conceptuais o intervals. Aquesta tècnica permet consolidar un possible criteri d'agrupació de dades. Per exemple, un *dataset* amb els resultats dels balanços d'empreses. Es podria substituir els resultats dels balanços per 'positiu', 'negatiu' i 'neutre' i, posteriorment, agrupar pel resultat del balanç.

Aquestes tècniques són molt importants per als models de predicció, ja que permeten que les dades siguin més senzilles d'interpretar i això comporta que el rendiment del model sigui millor i que aquest tingui la capacitat de trobar patrons més fàcilment, és a dir, el model es fa més senzill.

Pregunta 2

Pregunta 3

Pregunta 4

Pregunta 5

Pregunta 6

Pregunta 7