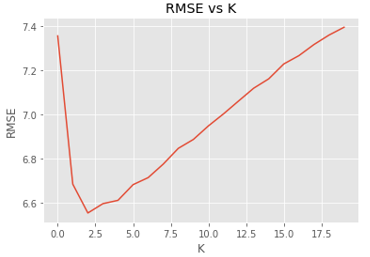
Feature importance

Hyperparameter tuning



Cross Validation

La validació creuada és un mètode estadístic per avaluar el rendiment generalitzat, és més estable i exhaustiu que utilitzar una divisió en un conjunt d’entrenament i un de test.

En cross-validation, les dades es divideixen repetidament i es formen diversos models. La versió més utilitzada de validació creuada és *k-fold cross-validation*, on k és un número especificat per l'usuari, normalment 5 o 10.

Quan es realitza la validació creuada de cinc vegades, les dades es divideixen en cinc parts de (aproximadament) la mateixa mida, anomenats *folds* o plecs. A continuació, s’entrena una seqüència de models.

El primer model s’entrena utilitzant el primer plec com a conjunt de prova, i els plecs restants (2-5) s’utilitzen com a conjunt d’entrenament. El model es construeix utilitzant les dades dels plecs 2-5, i després s'avalua la precisió en el plec 1. A continuació, es crea un altre model, aquesta vegada utilitzant el plec 2 com a conjunt de prova i les dades als plecs 1, 3, 4 i 5 com a conjunt d’entrenament.

Aquest procés es repeteix utilitzant els plecs 3, 4 i 5 com a conjunts de prova. Per a cadascun d’aquests cinc fraccionaments de les dades en conjunts d’entrenament i test es calcula la precisió. En la *il·lustració X* s’il·lustra el procés.

