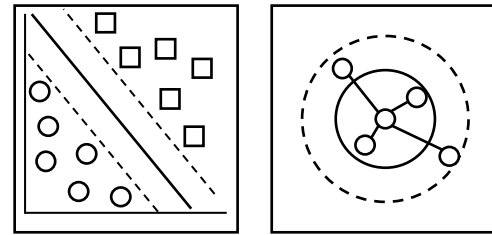


Tutorial 3.3a: Classification with Early Integration



(i) Import libraries

Import pandas, matplotlib and sklearn

(ii) Load flux and gene expression values and combine them

*Read fluxes and gene_expression_data
Concatenate [metabolic_data,
gene_expression_data]*

(iii) Split data into training and test sets

Specify proportion of test data and randomize training:test split for reproducibility

(iv) Normalize training and test sets

Perform feature scaling on X training samples and use same parameters to normalize X test samples

(v) Select hyperparameters

Select hyperparameters for fine tuning SVC and kNN models

(vi) Cross-validate hyperparameters

Use grid-search to compare hyperparametric combinations for robustness

(vii) Train model with optimal parameters

Fit best combination of parameters to the model with training samples

(viii) Generate predictions for test set

Obtain final predictions using test set samples

(ix) Compare robustness between SVC and kNN

Compute balanced accuracy scores and plot ROC curves to evaluate SVC and kNN performance