1. Show 2 Gaussian blobs

Show nearest neighbors for 3 test points

Make train-test split, fit 1-NN and show train and test scores

Show 3 nearest neighbors for 3 test points

Fit 3-NN and show train and test scores

Write a function which computes and shows the predictions on a large grid. It displays the decision boundary

--- find X\_grid by hand for 2\times 2 grid

Apply this function to 1, 5, 10, 30 neighbors

1. For odd number of neighbors from 1 to 29 compute and show train and test scores

Idealized representation of the train, test (validation) and true (generalization) accuracy

Train/validation/test split

For odd number of neighbors from 1 to 13 find the classifier with the best validation score, output its test score

Cross-validation definition and illustration

For the previous example find the classifier with the best cross-validation score, output its test score

Model selection and evaluation scheme

Use GridSearchCV for the previous problem

Show the graphs of mean train score and mean test score with one standard deviation

1. Load load-digits, show one digit, use apply k-NN GridSearchCV for k=1,3, … 29, output the test score of the best classifier

Stratified k-fold cross-validation: illustration

Example: x\_i=1, i=0,..,19; x\_i=1, i=20,..,99. Use 5-fold cross-validation for Dummy classifier(‘most\_frequent’) with and without stratification. Output cross\_val\_scores

--- do it by hand

Use shuffling

Load iris dataset, show the labels. Use 3-fold cross-validation for 5-NN without stratification (fails) and with stratification (success)

LOO, the same example

1. k-NN for regression

sin + noise, 1, 3, 9 neighbors, illustration

Define 𝑅^2 coefficient of determination

For k=1,..,29 for apply GridSearchCV to find the best classifier

1. Wrong and right ways to perform cross-validation