

Parameter Tuning

Cross validation, KNN-regression

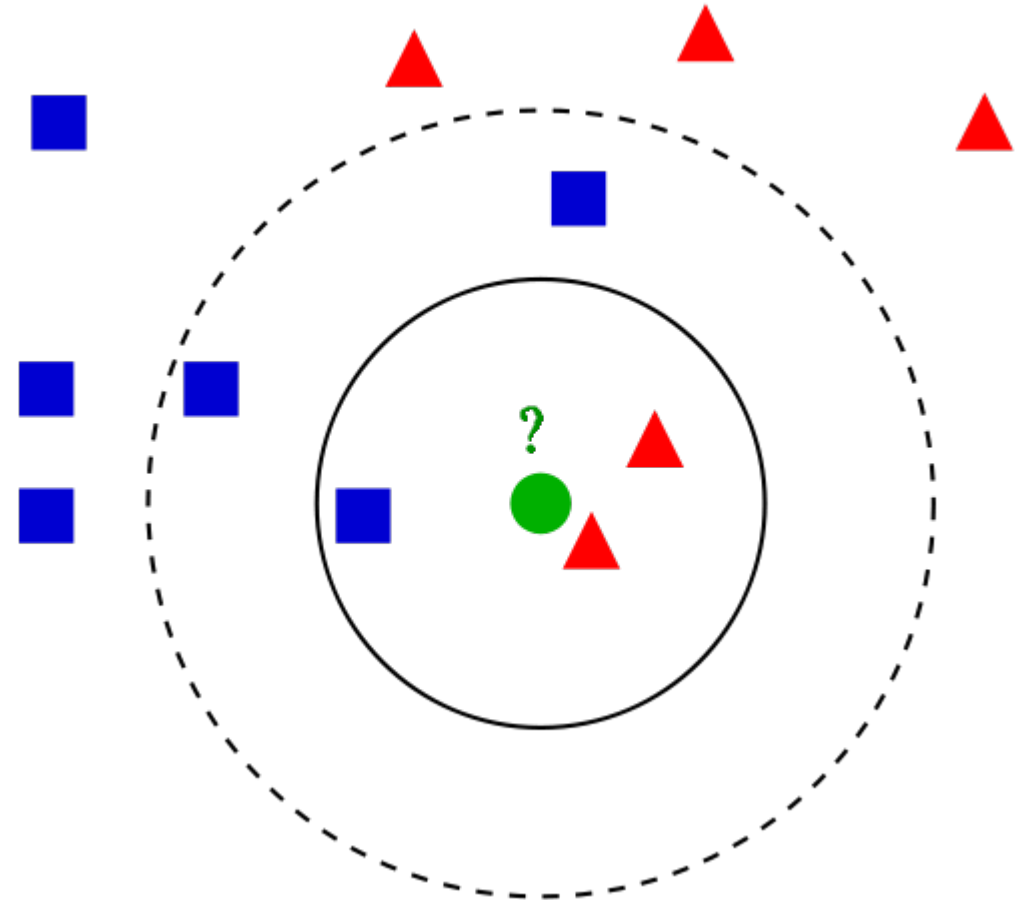
Akitaka Matsuo

Motivation

- In this lecture we will learn standard machine learning process using `sklearn` package
- We use KNN regression as an example
 - We need to tune the value of k in KNN regression

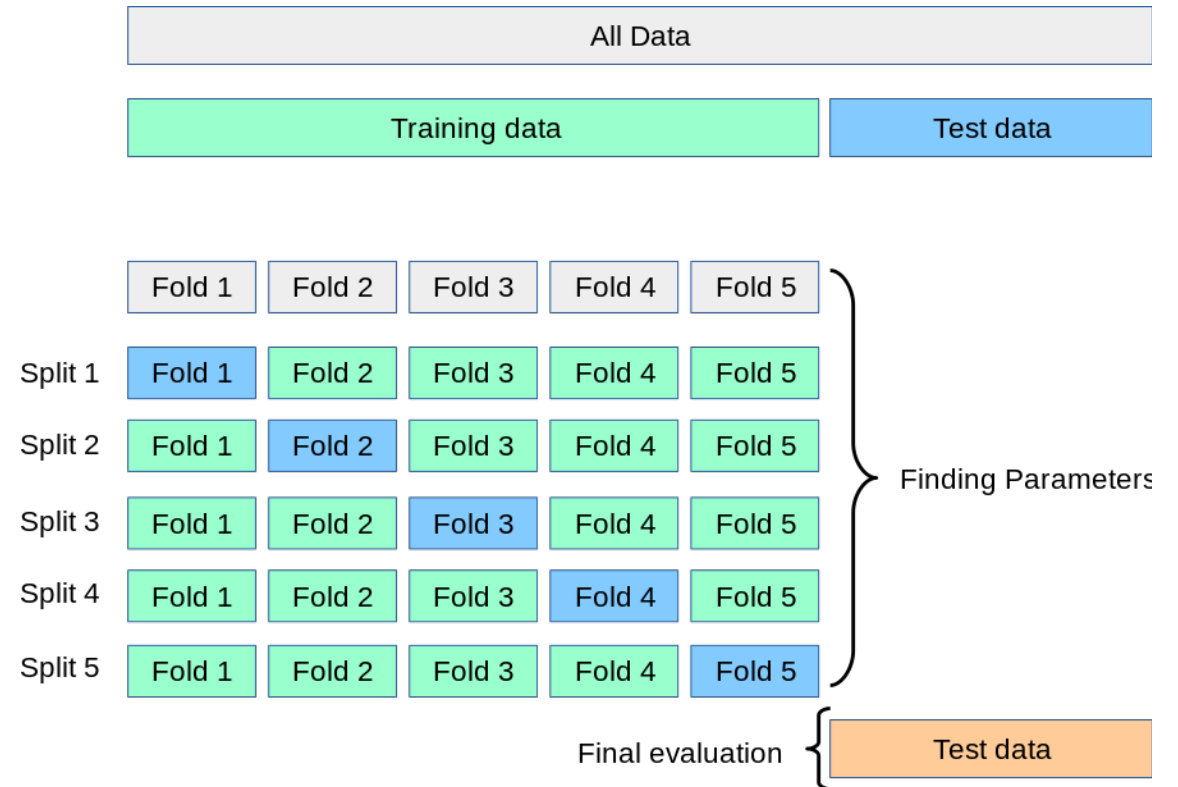
K-Nearest Neighbors (KNN)

- Very simple non-parametric method
- Classify based on the classes of neighbors
- Example:
 - Two classes (blue and red)
 - What is your prediction about the class of new point?
 - k is the number of points to consider
 - If $k = 3$: Red
 - If $k = 5$: Blue
- k is the tuning parameter



Cross Validation

- Cross-validation will be useful tool for parameter tuning
- The process looks like the figure
 - Split training set into small chunks
 - Hold one fold for validation, train temporary model using other folds
 - Calculate the performance indicator
 - repeat k times
 - (optionally) repeat n times



Tuning grid

- In `sklearn`, you can specify the experiment schedule using `tune grid`
- We can determine the specific value of parameters to experiment
- For KNN regression, there is only one tuning parameter, k