K fold cross validation

MT21MCS013 Jay

Supervised learning classification techniques

- Support Vector Machine
  - Kernel - linear, polynomial, quadratic,RBF(Radial)
- Nearest Neighbour
  - City block distance(K=1), euclidean distance
- Decision Tree
  - Order in which features are selected is important
- Random Forest
  - Builds multiple decision trees and merges them together to get a more accurate and stable prediction.

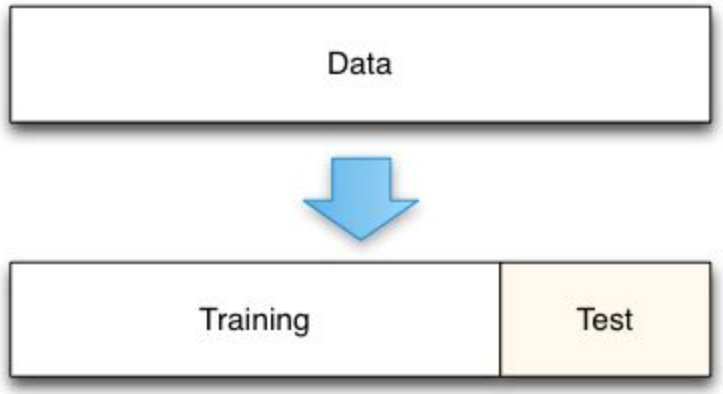
Train test split

Option 1 :

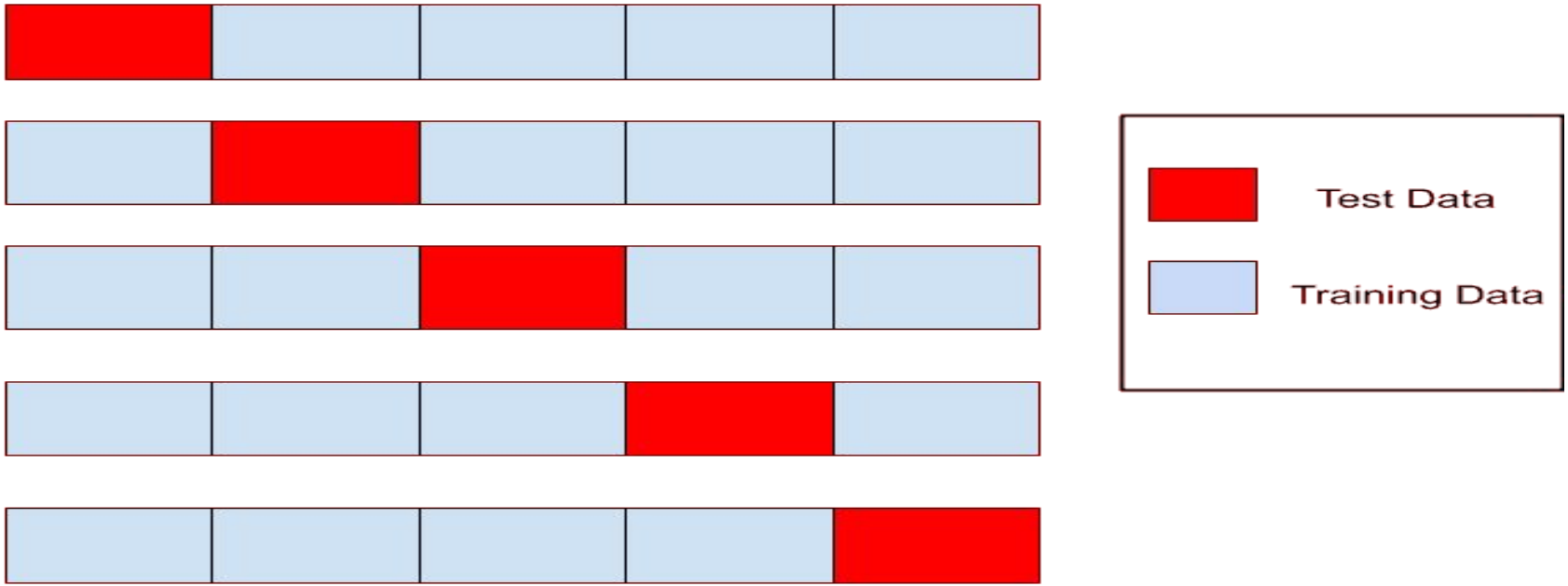
100% used as training and the same for testing

Option 2 :

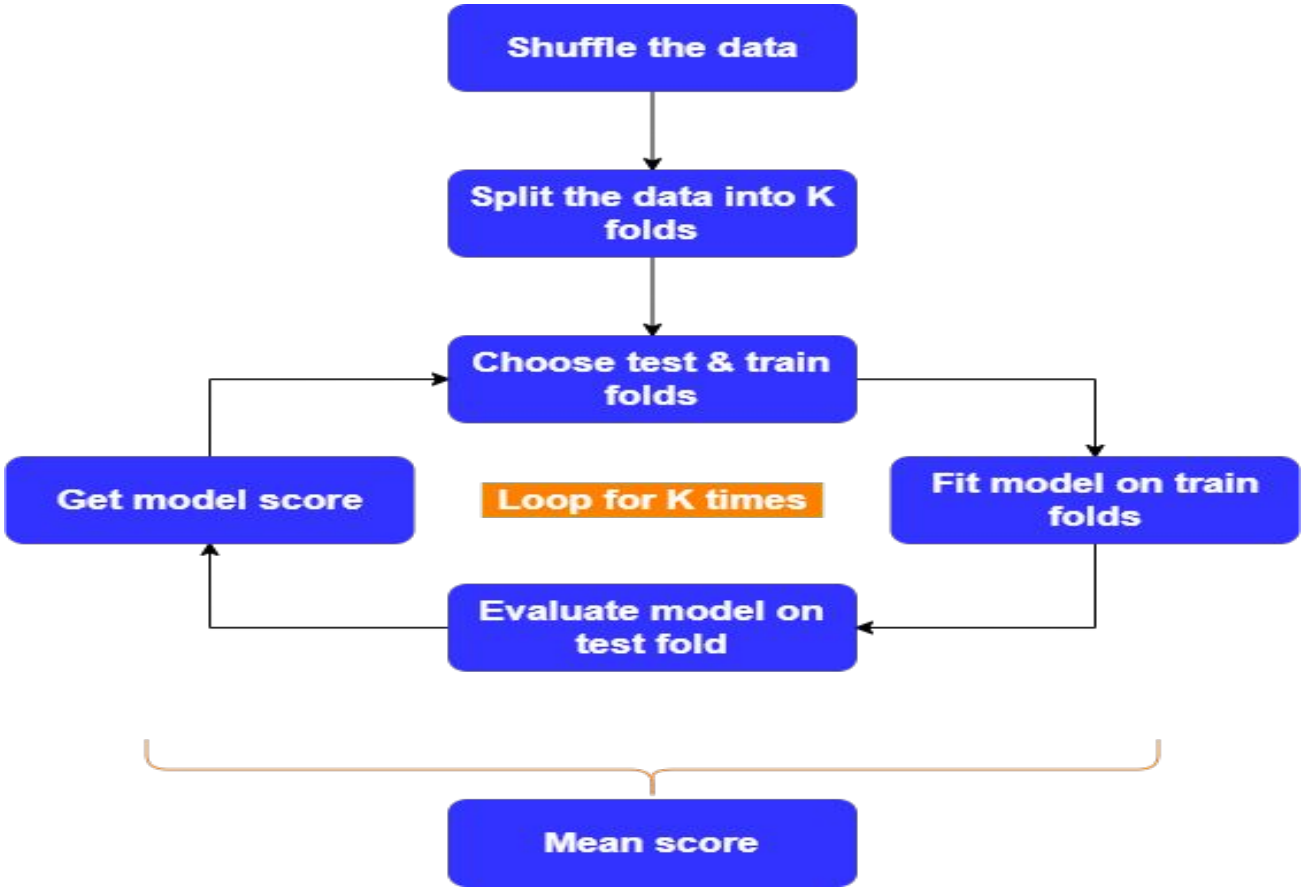
X:Y ratio, X% of dataset is used as training and Y% of it as test



K-Fold cross validation



K-Fold cross validation (Working)

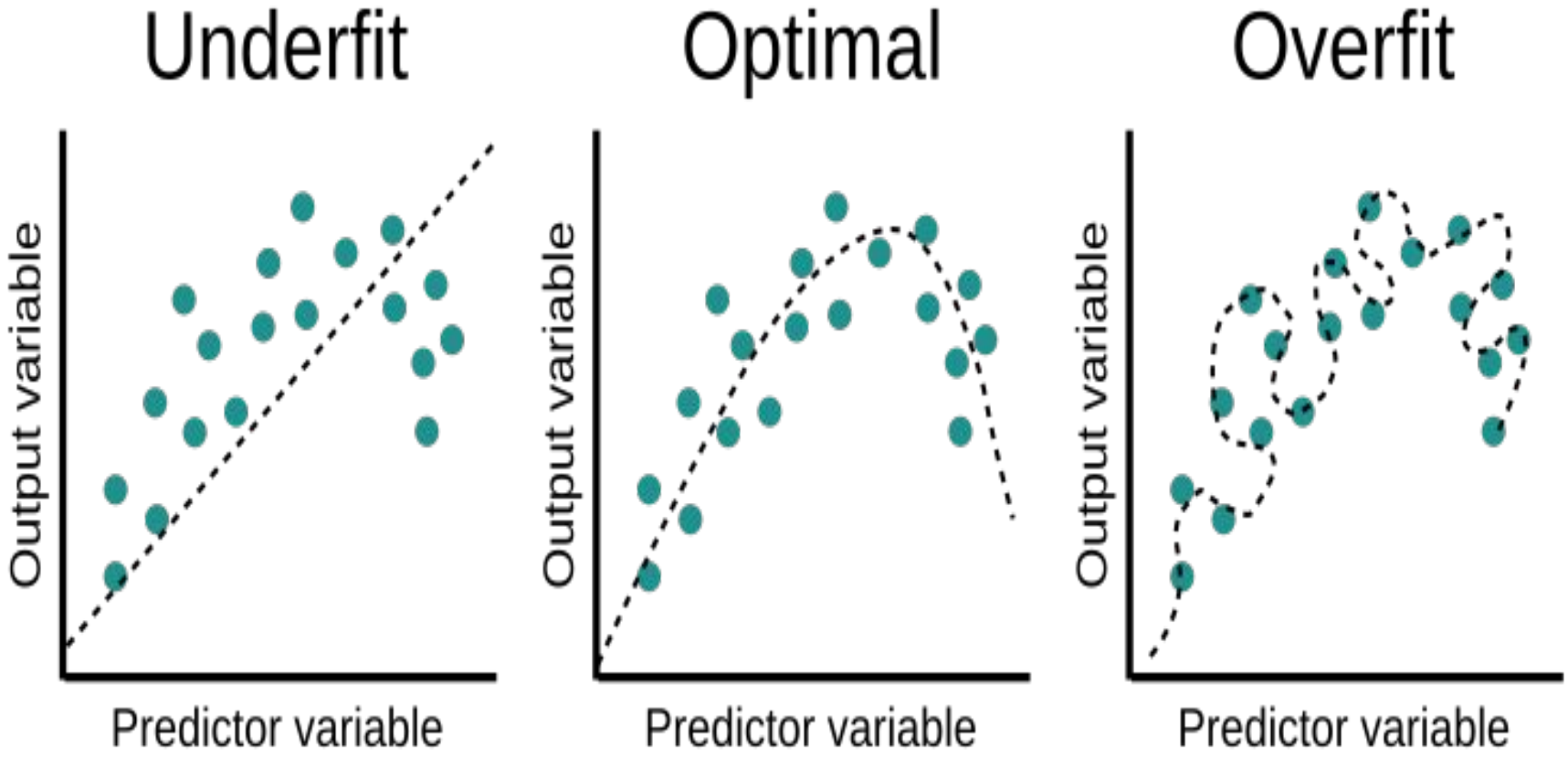


Overfitting

- KNN
- Parzen Window
- SVM
  - RBF(Radial Basis Function)
  - Polynomial with high degree

Underfitting

- Linear classifier
- SVM
  - Linear kernel


Underfit
Optimal
Overfit
Output variable
Predictor variable
Output variable
Predictor variable
Output variable
Predictor variable

Reference

[1] Image - fitting https://www.olexsys.org/images/overfitting.png

[2] Image - K fold https://www.mltut.com/wp-content/uploads/2020/05/cross-validation.png

[3] Image - K fold - working
https://raw.githubusercontent.com/satishgunjal/images/master/Inner_Working_KFold.png

[4] Image - Train test split -
https://www.researchgate.net/figure/Train-Test-Data-Split_fig6_325870973