Pattern Analysis and Machine Intelligence

Matteo Matteucci, Davide Eynard

06/07/2016

Abstract

Please write Ex1 and Ex3 on one sheet and Ex2 and Ex4 on a different one. Indicate clearly which exercise and question you are answering in you manuscript.

1 Statistical learning (8 points)

According to statistical learning theory, in regression, we assume a relationship exists between an observed variable and a dependent variable in the form

$$Y_i = f(X_i) + \epsilon_i, \quad \epsilon_i \sim N(o, \sigma^2).$$

1. What are the *two* sources of errors we have when estimating f from data and what are these errors due to?
2. According to statistical learning theory, Test and Training Mean Squared Errors are related by the Bias-Variance trade-off; write and comment the formula representing the Bias-Variance trade off for the Expected Prediction Error in Regression.
3. How is model complexity related to Bias and Variance? First provide a definition of model complexity, then according to that definition explain how bias and variance are influenced by an increased model complexity and why.
4. Provide two different examples for $f(X_i)$, describe their trade-off in terms of Bias and Variace, i.e., if one of them reduces either bias or variance with respect to the other, and explain when one should prefer one of the two models with respect to the other.

2 Linear regression (8 points)

Given the variables $x = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and $y = \{3.3, 3.6, 5.2, 5.6, 7.4, 8.3, 8.7, 9.7, 11.2\}$

1. Manually compute the parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ of a linear model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ which fits the given data

2. What is the value of MSE calculated between the values of y and the ones returned by the $\hat{y}$ function?

3. How can we compute if the trend identified by $\hat{\beta}_1$ is significant or it is just due to spurious correlations?

3 Classification (8 points)

Let's consider 4 "classical" classification algorithms: K-Nearest Neighbours (KNN), Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Support Vector Machines (SVM)

1. Provide a short description of each of the 4 algorithms highlighting the idea behind each of them, the basic assumptions, the complexity of the decision boundary, the learning process.

2. When should we chose each of them?

3. Describe what are the characteristics that could make each of the algorithms preferable to the others.

4. Which of the four algorithms has a non-linear decision boundary? In case of linear decision boundary, how it is possible to extend the algorithm so to have a non-linear one?

4 Clustering (8 points)

K-Means is a clustering algorithm that, despite some limitations, is still widely used for many applications.

1. Describe the K-Means algorithm in details, and elaborate about its initialization, i.e., what approach would you suggest to address the fact that the result of K-Means clustering depends on the initial positions of centroids?

2. Highlight the main advantages of using K-Means instead of another clustering algorithm (you can explicitly compare K-Means with other algorithms you choose) and suggest some applications for which you consider it better suited.

3. What clustering algorithm would you suggest to address K-Means limit of not being able to deal with non-globular clusters? Choose one (if there are many) and motivate your answer with respect to K-Means.

4. What approach would you suggest to address the need of knowing the number of clusters in advance in k-means? What alternative clustering algorithm would you choose not to have the issue of selecting an initial number of clusters (discuss your answer).