

Pattern Analysis and Machine Intelligence

Matteo Matteucci, Davide Eynard

28/09/2016

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

$$Y_i = f(X_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2).$$

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. The previous formula does not hold for Classification, but a useful result exists for the Classification Error Rate as well. Write and comment what statistical learning theory states about the minimum achievable average test error rate.
4. Describe in detail how the previous result is used to derive the Linear Discriminant Analysis classifier and derive the shape of its decision boundary.

2 Linear regression (8 points)

Given the following observations

$$x = \{41, 45, 46, 47, 49, 50, 52, 54, 55, 56\}$$

$$y = \{98, 108, 105, 107, 112, 114, 124, 121, 124, 122\}$$

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. Is the trend identified by $\hat{\beta}_1$ significant or is it just due to spurious correlations? You have to provide supporting computations and justifications for your answer.

3 Generative vs. Discriminative Models (8 points)

A classical distinction between classification models is the generative vs discriminative one. Answer the following about this distinction.

- (a) What are discriminative and generative models? How do they differ? Which one should be preferred and why?
- (b) Which one, between *Logistic Regression*, *Linear Discriminant Analysis*, and *Support Vector Machines* is a discriminative model and which a generative one? Why?.
- (c) What is a Support Vector Machine? How is it defined and how is it trained? Why do Support Vector Machines have this name?
- (d) What is the kernel trick and how can it be applied to Support Vector Machines (i.e., what do you need to change with respect to the original algorithm)?

4 Clustering (8 points)

Suppose you want to evaluate some clustering algorithms using SSE and Accuracy.

1. Which of these measures is defined as “internal”, which is “external”, and what does this mean?
2. After running your function, you obtain the result SSE=15.3. How can you evaluate whether this is a good or a bad result? What would you compare this result with?
3. One of the clustering algorithms allows you to choose the number of clusters in advance. You calculate SSE after different executions of this algorithm, using $K=2,3,\dots,10$. SSE for $K=10$ provides the lowest value: what can you deduce from this?
4. Now suppose you have ground truth for your dataset. You run two different clustering algorithms on the same data and obtain the following results:

	SSE	Accuracy
Algorithm1	98.2	90%
Algorithm2	300	93%

What is the meaning of these results? Which algorithm is better?