Pattern Analysis and Machine Intelligence

Matteo Matteucci, Davide Eynard

23/02/2015

1 Statistical learning (8 points)

In statistical learning theory, Test and Training set Mean Squared Errors are related by the so called Bias-Variance trade-off:

1. Write and comment the formula representing the Bias-Variance trade off for the *Expected Prediction Error* in Regression

The previous formula does not hold for Classification, but a useful result exists for the *Classification Error Rate* as well

2. Write and comment what statistical learning theory states about the minimum achievable average test error rate.

Provided the previous result for Classification, answer the following questions

3. Describe in detail how the previous result for the optimal classifier is used to derive the Logistic Regression classifier; provide a detailed description of the underlining model for Logistic Regression and derive the shape of the decision boundary for Logistic Regression.

4. Describe in detail how the previous result for the optimal classifier is used to derive Linear Discriminant Analysis; provide a detailed description of the underlining model for Linear Discriminant Analysis and derive the shape of the decision boundary for Linear Discriminant Analysis.

2 Linear regression (8 points)

Provide detailed answers to the following

1. Let assume you have a dataset with n=1000 observations and you try to fit different models on the data:

   - A linear regression model, i.e. $Y = \beta_0 + \beta_1 X + \epsilon$

- The polynomial regression model $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4 + \epsilon$
- A smoothing spline (i.e. an even more flexible model than the previous two)

For each of the three models, you calculate both training and test RSS. How would you expect the values of RSS to be (both in the training and in the test case), supposing that the true relationship between X and Y is (a) linear or (b) cubic?

2. What is the *additive assumption* in a linear regression model? Show how you would detect and quantify a possible interaction between the variables in a regression model and how you would model it from a statistical perspective. Finally explain, with an example, how your model would take this interaction into account (e.g. explain how, given a change in the input, the dependent variable – and consequently the output – change).

3 Classification (8 points)

Consider the following dataset with three classes and the Linear Discriminant Analysis model (LDA hereafter)

X1	X2	class
1	1	A
2	2	A
2	3	A
3	1	A
4	1	A
1	4	B
2	5	B
3	4	B
4	5	B
4	3	C
6	1	C
6	2	C
6	3	C

1. Compute the parameters of a LDA classifier
2. Compute the discriminant functions for the LDA classifier
3. Compute the equations of the boundaries between the three classes according to the classifier
4. Draw the dataset and the boundaries between the classes according to the classifier
5. Compute the parameters in case of a Quadratic Discriminant Analysis model

4 Clustering (8 points)

Given the two datasets shown in figure (where blue diamonds are data points and red circles different centroids starting points), calculate and show the different steps of the K-Means algorithm for both the examples in the following way:

- At each step, specify the initial positions of the centroids
- Without actually calculating it (unless it is needed to verify distances you cannot tell apart at a glance), for each step specify which centroid the various dataset points belong to
- After you have assigned data points to the different centroids, calculate their new positions and proceed to next step

Tell how many iterations the algorithm needs to converge, compare its behavior in the two cases and write a comment about it: is the final situation the one you might expect/desire? If not, explain why.