Open your mind. LUT.
Lappeenranta University of Technology
LUT Machine Vision and Pattern Recognition 2015-09-22

BM40A0700 Pattern Recognition

Lasse Lensu

Exercise 3: Feature processing and Bayes rule

1. Assume that you are trying to classify samples belonging to two classes, and you have two normally distributed features available. The class means for Feature1 are 3 (Class1) and 7 (Class2), and the standard deviations (STDs) are 2 (Class1) and 1 (Class2). The class means for Feature2 are 5 (Class1) and 6 (Class2), and the STDs are 0.2 for both classes. Determine which feature is more useful in classification by computing the Fisher discriminant ratio (FDR) for both features.

2. Separability of features: Make two Matlab plots visualizing the probability density functions (PDFs) of the features in a two-class case, one plot for each feature. The case can be as follows: class means for Feature1 are 3 (Class1) and 7 (Class2), and the STDs are 2 (Class1) and 1 (Class2); the class means for Feature2 are 5 (Class1) and 6 (Class2), and the STDs are 0.2 for the both classes.

   Hints: You can calculate the PDFs with function `normpdf`. From the two plots, approximate the decision boundaries (threshold for `x`) that should be used in classification based on each feature.

3. Given the following data vectors:

   $$x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad x_4 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad x_5 = \begin{pmatrix} 1 \\ 10 \end{pmatrix},$$

   normalize them using the following:

   (a) Min-max normalization.
   (b) Mean-variance normalization (standardization).
   (c) Softmax-scaling.

   Hints:

   $$s_{ik} = \frac{x_{ik} - \bar{x}_k}{\sigma_k} \tag{1}$$

   $$t_{ik} = \frac{1}{1 + e^{-s_{ik}}} \tag{2}$$

   $$u_{ik} = \frac{x_{ik} - x_k^{min}}{x_k^{max} - x_k^{min}} \tag{3}$$

4. Bayes formula: John Doe is a medical doctor interested in statistics. He uses a specific test to determine if his patient A. N. Onymous has cancer. However, the test is not perfectly reliable: the test succeeds to reveal cancer only with probability 0.98. In addition, when the patient does not have cancer, the test has a probability of 0.03 to show positive. It is also known that the probability of cancer in the overall population is 0.008. What is the probability that Mr Onymous has cancer when the test is positive?

PO BOX 20, FI-53851 LAPPEENRANTA, FINLAND
E-mail: lasse.lensu@lut.fi

After the positive result, John Doe decides to perform the test once more. The second test also results positive. What is the probability of cancer after the second test? Assume that the results of the tests are independent.

Hints: Use the Bayes formula to calculate the a posteriori probabilities.