Open your mind. LUT.
Lappeenranta University of Technology
LUT Machine Vision and Pattern Recognition 2018-09-13

BM40A0701 Pattern Recognition

Lasse Lensu

Exercise 3: Feature processing

1. Receiver operating characteristic (ROC) curve (1 point):
   - What does a ROC curve represent?
   - How can the ROC curve be produced for a binary classification case?
   - How is it possible to use the ROC curve to determine the optimal threshold for a binary classifier?

2. Class separability (1 point): 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.

   Hints:

$$FDR_1 = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{1}$$

$$FDR_n = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\intercal (\mathbf{C}_1 + \mathbf{C}_2)^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \tag{2}$$

3. Separability of features (1 point): 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. From the two plots, approximate the decision boundaries (threshold for the feature value) that should be used in classification based on each single feature.

   Hints: You can compute the PDFs with function `normpdf`.

4. Feature normalisation (1 point): Given the following data vectors:

$$\boldsymbol{x}_1 = \begin{pmatrix} 8 \\ 9 \end{pmatrix}, \quad \boldsymbol{x}_2 = \begin{pmatrix} 1 \\ 9 \end{pmatrix}, \quad \boldsymbol{x}_3 = \begin{pmatrix} 6 \\ 1 \end{pmatrix}, \quad \boldsymbol{x}_4 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \quad \boldsymbol{x}_5 = \begin{pmatrix} 10 \\ 10 \end{pmatrix},$$

   normalize the data using the following methods:

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

   Hints:

$$s_{ik} = \frac{x_{ik} - \bar{x}_k}{\sigma_k} \tag{3}$$

$$t_{ik} = \frac{1}{1 + e^{-s_{ik}}} \tag{4}$$

$$u_{ik} = \frac{x_{ik} - x_k^{min}}{x_k^{max} - x_k^{min}} \tag{5}$$

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