COURSE 02457

Non-Linear Signal Processing: Exercise 1

This exercise is based on C.M. Bishop: *Neural Networks for Pattern Recognition*, Chapter 1. The objective of this exercise is to become familiar with the relations between probability densities and histograms, Bayes' theorem, conditional distributions and decision rules.

Print and comment on the figures produced by the software main1.m, norm1d.m, probconfus.m as outlined below at the three **Checkpoints**.

Densities and histograms

A probability density function $p(x)$ specifies that the probability of the variable x lying in the interval between any two points a, b is

$$P(x \in [a,b]) = \int_a^b p(x)dx \tag{1}$$

If $\{x_1, x_2, ..., x_N\}$ is a set of points, the histogram of this set, evaluated on the ordered set of points $[z_1, z_2, ..., z_M]$ is defined

$$H_j = \sum_{x_k \in [z_j, z_{j+1}]} 1 \quad , j = 1, ..., M-1 \tag{2}$$

and the normalized histogram is given by

$$\tilde{H}_j = \frac{H_j}{\sum_{j'=1}^{M-1} H_{j'}}. \tag{3}$$

The normalized histogram can be compared with the histogram approximation to the density

$$P_j = \int_{z_j}^{z_{j+1}} p(x)dx \quad j = 1, ..., (M-1). \tag{4}$$

We here focus on the density of the normal distribution:

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{5}$$

Checkpoint 1.1: Use the program main1.m to illustrate the relation between densities, histogram approximations to densities. Create samples from the univariate normal density using randn.m and compare the sample histrograms with the density of the normal distribution and with its histogram.

Bayes' theorem

For real variables x and labels $C_k = 1, ..., c$, Bayes' theorem reads,

$$P(C_k|x) = \frac{p(x|C_k)P(C_k)}{p(x)} \tag{6}$$

If the class-conditional distributions are univariate normals with individual parameter sets we can use `main1.m` to illustrate Bayes theorem.

Checkpoint 1.2: Define three univariate normals and plot the resulting densities. Set the prior probabilities $P(C_k)$ and plot the resulting $p(x)$ and the posterior probabilities $P(C_k|x)$. Do this for different setting of the prior probabilities and comment on the densities you get. Can you create a situation where for one of the classes there are no points x where it is the most likely class?

Decision boundaries

A decision rule is a division of the space of x so that each point is uniquely associated with a class C_k. We can set up a simple 1D linear discriminant by dividing the real axis into three intervals $I_1 =]-\infty, d_1], I_2 =]d_1, d_2], I_3 =]d_2, \infty[$.

One way to summarize the errors of a decision rule is the error confusion matrix $R_{j,k}$ defined, e.g., as

$$R_{j,k} = \int_{I_j} p(x|C_k)dx \tag{7}$$

Checkpoint 1.3: Define the decision boundaries as above and use the histogram approximation to estimate the error confusion matrix. Do this for different decision boundaries, plot the posteriors and the decision regions. Explain and comment on the confusion matrix.

Challenge

Implement an error-reject mechanism as discussed by in Bishop section 1.10.1. Plot the relation between the error rate and the reject rate. Comment on the initial slope of the relation for small reject rates.

DTU, August 2005,

Lars Kai Hansen