CS4720/7720 Spring 2015
Introduction to Machine Learning and Pattern Recognition

# Homework Assignment 1

due Thursday 2/12/2015

**Problem 1**.
The joint probability distribution of X and Y is given below:

| | | X | | |
|---|---|---|---|---|
| | | -1 | 0 | 1 |
| Y | -1 | 1/2 | 1/8 | 0 |
| | 1 | 1/16 | 1/4 | 1/16 |

Determine:
(1) The marginal distribution of X.
(2) The conditional distribution of Y given X=0.
(3) The mean values of X and Y.
(4) The variances of X and Y.
(5) The cross correlation between X and Y.

**Problem 2.**
Given $x \sim N(\mu, \sigma^2)$, with $\mu = 3$ and $\sigma^2 = 4$, compute the followings:
(1) $\mathrm{P}_r(x \le 0)$.
(2) The value of $x_o$ that makes $\mathrm{P}_r(x \le x_o) = 0.758$.

**Problem 3.**
Assume that the bivariate Gaussian random vector $\underline{x} = [x_1 \quad x_2]^T$ ( $^T$ denotes transpose) has the following mean vector and covariance matrix:

$$\underline{\mu} = \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 9/5 & -2/5 \\ -2/5 & 6/5 \end{bmatrix}.$$

a) Compute the correlation coefficient $\rho$ between $x_1$ and $x_2$.
b) Determine the eigenvector matrix $U$ and the eigenvalue matrix $\Lambda$ of $\Sigma$.
c) Determine the probability density function for $\underline{z} = U^T \underline{x}$.
d) Given $\underline{q} \sim N(\underline{0}; \mathrm{I})$, with I the identity matrix, show how to obtain $\underline{x}$ from $\underline{q}$.
e) Compute the mean $\mu_{x_1|x_2}$ and the variance $\sigma^2_{x_1|x_2}$.
f) Compute the predicted value of $x_1$ given $x_2 = 3/\sqrt{5}$ by using the conditional mean, i.e., $\hat{x}_1 = E(x_1 \mid x_2 = 3/\sqrt{5})$.

**Problem 4.**
The lightness feature has been measured from some salmon and sea bass, and the measured data are stored in the two data files 'SalmonLightness.dat' and 'SeabassLightness.dat', respectively. Note: the lightness value varies from 0 to 10. In MATLAB, use the function load() to load the data files into vectors.

(1) Plot two histograms, one for salmon and one for sea bass, to show the number of fish samples with the measured lightness falling within the intervals of ((k-1)*0.5, k*0.5], with k = 1,2, …, 20.

Note: you can use the MATLAB function hist() to plot the histogram and use the vector given by the data file 'formathist.dat' as the second input parameter of the function hist() to implement the bin definition.

(2) Compute the prior probabilities for the two types of fish: P(salmon) and P(sea_bass), which define the chance of getting a salmon or sea bass when randomly picking up a fish.

(3) Compute the class conditional probabilities for the lightness values in the 20 bins that specify a quantization scheme of the lightness measurements, where the probability mass functions (pmfs) of the quantized lightness values for the two types of fish are defined as

$P(lightness \in ((k-1)*0.5, k*0.5] \mid salmon)$, and

$P(lightness \in ((k-1)*0.5, k*0.5] \mid sea\_bass)$, for $k = 1,2,...,20$.

Note: a vector Y of the data counts in the 20 bins is readily obtained from Y = hist(), and by making a proper normaliation on Y, a probability mass function can be obtained.

Plot the two probability mass functions in one figure, using two different colors to represent salmon and sea bass.

Note: the MATLAB function stem() can be used to plot the pmfs.

(4) Based on the two pmfs obtained in (3), compute the following four probabilities terms:

$P(lightness \le 5 \mid salmon)$ and $P(lightness \le 8 \mid salmon)$

$P(lightness \ge 5 \mid sea\_bass)$ and $P(lightness \ge 2 \mid sea\_bass)$

(5) Computer the evidence pmf $P(lightness \in ((k-1)*0.5, k*0.5])$, $k = 1,2,...,20$.

Plot the probability mass function by using the function stem().

(6) Compute the posterior probabilities for salmon and sea bass for each of the 20 lightness bins, i.e.,

$P(salmon \mid lightness \in ((k-1)*0.5, k*0.5])$,

$P(sea\_bass \mid lightness \in ((k-1)*0.5, k*0.5])$, for $k = 1,2,...,20$.

Plot the two posterior probabilities in one figure, using two different colors to represent salmon and sea bass.

The MATLAB function stem() can again be used to plot the posterior probabilities.

In your report, please include the five figure plots, the probabilities, and your MATLAB code.