SGN-13006 Introduction to Pattern Recognition and Machine Learning
Department of Signal Processing, Tampere University of Technology
Joni Kamarainen and Katariina Mahkonen
Exercise 4

Complete / start preparing your solutions before the exercise session. During the exercise session you can consult the teacher, and once finished your work, show your solutions to the teacher to get the exercise points.

1. **Bayesian decision making** (3 points)

   (a) A mail delivery service is using four automatical distributors. The incoming mail is sorted by the distributors in proportions 20 %, 25 %, 35 % and 20 %. The error probabilities of the distributors are observed to be 2 %, 3 %, 2 % and 1 % respectively. Calculate the error probability of the whole mail delivery service.

   The following problems can be solved using the equation of Bayesian posterior probability:

$$P(C=\omega_i|X=x) = \frac{P(X=x|C=\omega_i)P(\omega_i)}{\sum_{c=1}^{N_C} P(X=x|C=\omega_c)P(\omega_c)}$$

   (b) We have a classification task between three classes. Generally 40% of the cases that we have to classify come from the class number 3 and the rest of the cases come equally often from class nr 1 and class nr 2. For the classification of a case we have binary (attributes,features) answers (yes/no) to three questions. The probabilities of answers to be positive to each of the questions for a case from each class are given in the table below. Classify a case where the answers are [no, no, yes]

Probability of positive answer $P(A_q = \text{yes}\|C_c)$		Class 1	Class 2	Class 3
	1	0.9	0.3	0.5
Question	2	0.1	0.7	0.5
	3	0.5	0.9	0.1

   according to Bayesian minimum error criterion.

   (c) Again we have a classification problem with three classes. For the classification of an item we take three measurements, whose values are always between 0 and 2. The probability density functions of each measurement for each class are given as follows:

$$p(\vec{x}|\omega_1) = \begin{cases} 1 & \text{when} 0 \leq x_1, x_2, x_3 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$p(\vec{x}|\omega_2) = \begin{cases} 8 & \text{when} 0.75 \leq x_1, x_2, x_3 \leq 1.25 \\ 0 & \text{otherwise} \end{cases}$$

$$p(\vec{x}|\omega_3) = \begin{cases} 1/8 & \text{when} 0 \leq x_1, x_2, x_3 \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

   The prior probabilities of classes are $P(\omega_1) = 0.25$, $P(\omega_2) = 0.25$ and $P(\omega_3) = 0.5$. Find the decision boundaries of a Bayes-classifier. **Tips:** Data is *t*hree dimensional, thus you can draw (fairly easily) the borders of volymes with nonzero probability density for each class. Then calculate the Bayesian posterior probability for each sub-volyme. According to those, you can define the sub-volymes where the classifier selects class 1, class 2 and class 3.

2. **Naive-Bayes Classifier** (3 points)

   Data: We are using the tiny subset of Imagenet dataset. Use the Matlab template `imagenet_tiny5_process.m` to handle the data.

   Feature extraction: For Bayesian classifier we use feature vectors of length 3 from each image. The feature vector contains an average value from each color channel of an image.

Classifier training: To perform Naive-Bayes classification with Gaussian (Normal) class density distribution models, calculate the mean μ_c and variance σ_c^2 of each feature within the training data feature vectors from each class $c = 1 \ldots 5$. Calculate also the prior probabilities $P(C = c)$.

Classification: For each evaluation image, calculate the posterior probabilities of each class $c = 1 \ldots 5$ as:

$$P(C = c|X = x) = \frac{P(X = x|C = c)P(C = c)}{\Sigma_{k=1}^{5} P(X = x|C = k)P(C = k)} \tag{1}$$

There is a Matlab function `mvnpdf`, which can be used to get a point value of the multi variate Gaussian (Normal) distribution function:

$$P(X = x|\theta = [\mu, \Sigma]) = \frac{1}{(2\pi)^{-k/2}|\Sigma|^{-1/2}} e^{(x-\mu)^T \Sigma^{-1} (x-\mu)} \tag{2}$$

Note that, as we are doing 'Naive' Bayesian classification, which means that correlations between different features are defined zeros, i.e. $\Sigma_{i,j} = 0$ for $i \neq j$, the covariance matrix Σ is diagonal. $|\Sigma|$ denotes the determinant of the covariance matrix.