FEUP 2021/2022 (1ST SEMESTER)

Class Exercises. Set 3. Generative Classifiers

1. In a two-class, two dimensional classification task the feature vectors are generated by two normal distributions sharing the same covariance matrix: $\begin{bmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{bmatrix}$ and the mean vectors are $\mu_1 = [0\ 0]^t$ and $\mu_2 = [3\ 3]^t$, respectively. Classify the vector $[1.0\ 2.2]^t$ according to the Bayesian classifier, assuming equal priors.

2. Consider the data in 'heightWeightData.txt'. The first column is the class label (1=male, 2=female), the second column is height, the third weight.
   a) Write a Matlab/Python function to model each class data as a bi-dimensional Gaussian distribution with the mean and variance matrix learnt from the data using maximum likelihood estimation. The function should receive as input the training data and the test data, making prediction (male/female) for the test points, using the maximum a posteriori (MAP) probability estimators.
   b) What's the estimated $p([165\ 80]^t|(male))$?
   c) Use the previous function to make predictions (male / female) for the following test points: $[165\ 80]^t$, $[181\ 65]^t$, $[161\ 57]^t$ and $[181\ 77]^t$.
   d) Repeat b) and c) using as features $D = H - W$ and $S = (H + W)/2$, where H and W are the height and weight, respectively. (Note: use the same function from a) but changing the input data).

3. Several phenomena and concepts in real life applications are represented by angular data or, as is referred in the literature, directional data. Assume the directional variables are encoded as a periodic value in the range $[0, 2\pi]$. Assume a two-class (y_0 and y_1), one dimensional classification task over a directional variable x, with equal a priori class probabilities.
   a) If the class-conditional densities are defined as $p(x|y_0) = e^{2\cos(x-1)}/(2\pi 2.2796)$ and $p(x|y_1) = e^{3\cos(x+0.9)}/(2\pi 4.8808)$, what's the decision at $x = 0$?
   b) If the class-conditional densities are defined as $p(x|y_0) = e^{2\cos(x-1)}/(2\pi 2.2796)$ and $p(x|y_1) = e^{3\cos(x-1)}/(2\pi 4.8808)$, for what values of x is the prediction equal to y_0?
   c) Assume the more generic class-conditional densities defined as $p(x|y_0) = e^{k_0\cos(x-\mu_0)}/(2\pi I(k_0))$ and $p(x|y_1) = e^{k_1\cos(x-\mu_1)}/(2\pi I(k_1))$. In these expressions, k_i and μ_i are constants and $I(k_i)$ is a constant that depends on k_i. Show that the posterior probability $p(y_0|x)$ can be written as $p(y_0|x) = 1/(1 + e^{w_0+w_1\sin(x-\theta)})$, where w_0, w_1 and θ are parameters of the model (and depend on k_i , μ_i and $I(k_i)$).

4. Let the features $x = (x^{(1)}, x^{(2)}, \cdots, x^{(d)})$ be binary valued (1 or 0). Let p_{ij} denote the probability that the feature $x^{(i)}$ takes on the value 1 given class j. Assume that there are only two classes and that they are equally probable. Let the features be conditionally independent for both classes. Finally, assume that d is odd and the $p_{i1} = p > 1/2$ and $p_{i2} = 1 - p$, for all i. Show that the optimal Bayes decision rule becomes: decide class one if $x^{(1)} + x^{(2)} + \cdots + x^{(d)} > d/2$, and class two otherwise.

5. Fitting a naive bayes spam filter by hand. Consider a Naive Bayes model (multivariate Bernoulli version) for spam classification with the vocabulary V = "secret", "offer", "low", "price", "valued", "customer", "today", "dollar", "million", "sports", "is", "for", "play", "healthy", "pizza". We have the following example spam messages "million dollar offer", "secret offer today", "secret is secret" and normal messages, "low price for valued customer", "play secret sports today", "sports is healthy", "low price pizza". Give the MLEs for the following parameters: θ_{spam}, $\theta_{secret|spam}$, $\theta_{secret|non-spam}$, $\theta_{sports|non-spam}$, $\theta_{dollar|spam}$.

6. Consider the dataset D described in the Table. The set D is to be used as training data for a binary classifier to identify whether a point $[x1x2]^t$ falls inside some given target shape or not. Positive class labels (+)

X_1	-2.4	-2.1	-1.7	-1.6	-1.5	-1.2	-1.1	-0.5	0.0	0.0
X_2	0.4	-0.3	-1.6	-1.3	1.5	1.9	-2.0	0.1	0.4	2.0
Class	$-$	$-$	$-$	$-$	$-$	$-$	$-$	$+$	$+$	$-$
X_1	0.1	0.1	0.1	0.2	0.3	0.4	0.8	1.0	1.7	2.0
X_2	-0.7	-0.6	0.0	-0.5	-0.5	0.9	0.2	0.1	-1.0	0.4
Class	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$-$	$-$

Table: Training dataset containing 20 data-points pertaining to two different classes.

correspond to data-points falling inside the target shape, while negative class labels (-) correspond to data-points not falling inside the target shape.

Given a data-point $p = [x_1 \ x_2]^t$, consider the two binary attributes, A_1 and A_2, where attribute $A_i(p)$ is 1 if $|x_i| > 1$ and 0 otherwise, $i = 1, 2$. Using these attributes to represent each of the data-points in D, compute the parameters of a Naive Bayes classifier for the dataset D. Using this classifier, compute the class label for the point $(0.9; 0.9)$.

7. Given the dataset $\{4, 5, 5, 6, 12, 14, 15, 15, 16, 17\}$, use parzen window to estimate the density $p(x)$ at $x = 3, 10, 15$. Use a:
   a) symmetric rectangle kernel with base $h = 1$
   b) symmetric triangle kernel with base $h = 2$
   c) gaussian kernel with bandwidth of $h = 4$

8. Repeat the previous question but with the Nearest Neighbor Estimation, setting the number of neighbours equal to:

   a) 1
   b) 5