Module-Lab 7: Probabilistic Machine Learning

Exercise 1 This is the continuation of Module-Lab6 exercise 2 i.e., the dataset to be used is the Ionosphere dataset – to facilitate the work, you can load the Ionosphere directly in Matlab by simply >> load ionosphere

a) Implement a multivariate Normal Bayes classifier i.e., assuming the class conditional density is modelled by Gaussians.
Hint: see the '*Lec6_BayesianInference.pdf*' slides.

b) Performance measures: assuming identical priors for both classes (ie, priors = 0.5) and using the maximum a-posteriori decision, calculate the TPrate, FPrate, the Balanced Acc, F1-score, Precision, and the Recall on the **test set**.

c) Implement the Naive Bayes classifier (NBC) using Gaussian likelihoods and compare the results obtained in (b).
Hint: see page 12 of the *Lec6_BayesianInference.pdf* slides

Hint: to avoid the underflow problem (due to the product of 32 likelihood terms), it is preferable to use the sum of the logarithms. Thus, the NBC expression becomes

$$\hat{y} = \underset{k\in\{1,\ldots,|C|\}}{\operatorname{argmax}} \left(\log(p(C_k)) + \sum_{i=1}^{n} \log(p(x_i|C_k)) \right)$$

Some results using Matlab implementation:
>> ********** Classifier Normal Bayes
Acc = 0.70
BAcc = 0.79
Pre = 0.98
Rec = 0.65
F1 = 0.78

>> ********** NBC
Acc = 0.49
BAcc = 0.69
Pre = 1.00
Rec = 0.38
F1 = 0.55

Exercise 2 Now, we will use the Fisher Iris dataset (https://archive.ics.uci.edu/ml/datasets/iris) which can be loaded within Matlab by simply >> load fisheriris

a) This is a "single-use" 3-classes dataset with no separated test set. So, the learned models will be tested on the same set that has been used for training. Create your own code, in Matlab or Python, to open/load the dataset, and then to separate the dataset in 3 subsets as function of the class/label. You can use some bits of the code below

```
load fisheriris %meas species
N = size(meas,1); %150
nf = size(meas,2); %4
unique(species); % 'setosa', 'versicolor', 'virginica'
t = zeros(N,1);
for i=1:N
    if species{i}(4) == 's', t(i) = 1; end
    if species{i}(4) == 'g', t(i) = 2; end
end
X = meas;
```

Hint1: see page14 of the *Lec6_BayesianInference.pdf* – slides.
Hint2: there are exactly 50 examples per class and 4 features/attributes in total.

b) Train 3 variants of the NBC (Naive Bayes Classifier) using Normal/**Gaussian** class-conditional densities (likelihoods) for increasing number of features:
(i) features 1 and 2
(ii) features 1, 2 and 3
(iii) all the 4 features

Hint: some numbers regarding the Gaussian parameters of the *likelihoods*:
Mean = 5.0060 3.4280 1.4620 0.2460
Variance = 0.1242 0.1437 0.0302 0.0111

C) Test the 3 variants on the Iris dataset in terms of Precision and Recall.

Hint: some numbers,
// features 1-2
Pre1 = 0.98 Rec1 = 0.98 F1 = 0.98
Pre2 = 0.74 Rec2 = 0.74 F2 = 0.74
Pre3 = 0.62 Rec3 = 0.62 F3 = 0.62

// features 1-3
Pre1 = 1.00 Rec1 = 1.00 F1 = 1.00
Pre2 = 0.88 Rec2 = 0.88 F2 = 0.88
Pre3 = 0.76 Rec3 = 0.76 F3 = 0.76

// features 1-4
Pre1 = 1.00 Rec1 = 1.00 F1 = 1.00
Pre2 = 0.94 Rec2 = 0.94 F2 = 0.94
Pre3 = 0.94 Rec3 = 0.94 F3 = 0.94

d) Compute the **confusion matrix** for the model in **b (ii)** ie, using features 1,2,3.