HW5

CS4720 Machine Learning

By HG King

Question 1 (Practice). This question is to demonstrate that: 1) when the selected model is poor, the maximum-likelihood classifier does not produce satisfactory results; and 2) proper transformation of the data can compensate for poor models. The dataset1 used for this question is divided into training2 and test3 data, with each one consisting of 3 classes in a 2D-feature space.

a) Assuming Gaussian distribution for all three class conditional densities, and with unknown means and covariances, compute the maximum likelihood estimates for each class using the training data.

This one was pretty straight forward, estimate the mu and sigma for the data based on the maximum likelihood estimation method. From the notes, in the multivariate case:

**μ** = 1/n \* Σ **x**k

and

**Σ** = 1/n \* Σ( **x**k - **μ**)( **x**k - **μ**)T

b) Ignore the priors, i.e. assume 1/3 for all three classes, and redo parts a) and b) of Homework 2, Question 1, and use those same Matlab functions in part c) below. Use the means and variances from part a) above.

Again, this one was easy since the functions for mahalanobis and the discriminants were already written out. All I had to do was transfer over the code and run it on the new data.

c) Classify the test data and compute the test error using confusion matrix.

The confusion matrix is the flattened, real values of the classes and the estimated values of said classes. The estimated classes make up the rows and the columns are the real values. To find the percentage of correct classification, just divide the matrix value over the number of data values.

d) Bayesian estimates. The data has a simpler description when seen in polar coordinates. Use cart2pol() to transform all the data points to polar coordinates. Use scatter() to plot the transformed points. What you should find is that the transformed data looks Gaussian on the radius r and uniform on the angle θ. So, ignore the angle θ and classify the test data only on r as follows. The problem is now 1-D and again, if you inspect the data, Gaussian distribution is a more suitable pdf to describe all three classes. Assume then that each class has p(r|ωi) = N µi ,σ 2 with µi unknown and variance σ 2 = 0.25 for all three classes. Let the only prior knowledge about µi be p(µi) = N µ0 = 0, σ 2 0 = 100 and compute the Bayes estimates for µi and the posterior distribution p(µi |Di) of all three classes. Next compute p(r|ωi ,Di) = ´ p(r|µi) p(µi |Di)dµi and use this density estimate to classify the test data and compute the test error using confusion matrix. Do not forget to comment on the results from each step above