ASU

Project 1 Report

Statistical Learning and Pattern Recognition

Option 1

Prerna Satija (Student)

10/4/2013

Abstract

Calculating the mean and covariance maximum likelihood estimates for the 3 classes as per the data given in Training.txt.

Using these estimates, getting the posterior probability of each sample for each row for minimum error rate classification condition.

Finding the maximum probability and the class corresponding to the max posterior probability.

Finding the error count as the variations in the class labels between original class labels and the ones computed.

Project Requirements

This project requires the use of Matlab software for computing values.

Question 1

Given the feature vector $X = [X_1, X_2]$ for the classes labeled 1, 2, 3.

μ = [sum of all rows of X_1, sum of all rows of X_2]/n

$\sum = ((X_1 - \mu_1)((X_1 - \mu_1)^T + (X_2 - \mu_2)((X_2 - \mu_2)^T)/n$

Thus, the maximum likelihood estimate for the mean is the sample mean and the maximum likelihood estimate for the covariance matrix is the covariance as defined above.

For the data given in Training.txt, separating the number of rows for each class,

Class 1

Mean_1 = [3.0766 3.1020]

Covariance_1 = [2.0819 0.8684

0.8684 1.8769]

Class 2

Mean_2 = [1.9521 -1.0157]

Covariance_2 = [1.9526 0.1324

0.1324 1.8510]

Class 3

Mean_3 = [-0.0631 1.9365]

Covariance_3 = [1.0310 0.0303

0.0303 1.9915]

Question 2

Applying minimum error rate classification for 2 D multivariate densities of each of the 3 classes will give:

$g_i(x) = P(w_i|x)$

Where $g_i(x)$ is the discriminant function

$p(w_1)=p(w_2)=p(w_3)=1/3$

$P(w_i|x) = p(x|w_i)P(w_i)/p(x)$

Ignoring the evidence p(x).

Using the equation for Multivariate density, (when d = 2)

$P(w_1|x)$ = (1/3)*(exp(-0.5*(X-Mean_1)*inv(Covariance_1)*transpose(XMean_1)))/(((2*pi)^(d/2))*(det(Covariance_1))^(1/2))

$P(w_2|x)$ = (1/3)*(exp(-0.5*()X-Mean_2)*inv(Covariance_2)*transpose(XMean_2)))/(((2*pi)^(d/2))*(det(Covariance_2))^(1/2))

$P(w_3|x)$ = (1/3)*(exp(-0.5*(X -Mean_3)*inv(Covariance_3)*transpose(XMean_3)))/(((2*pi)^(d/2))*(det(Covariance_3))^(1/2))

Did this for all 1000 rows of the data set.

Found the maximum posterior probability for each row and compared the maximum value with each probability calculated. Find the class label for which the posterior probability is maximum. Number of discrepancies between class labels already given in the set and classes found using this method will give the number of errors. Error rate is the number of such errors divided by the total number of rows.

Training error count = 133

Training Error rate is 0.133

Similarly did for testing data set using mean and covariances of training data set.

Testing error count = 143

Testing Error rate is 0.143

Question 3

Prior Probability = num of samples of a class/total num of samples

Thus, $P(w_i) = C_i/N$

$P(w_1) = 202/1000 = 0.202$

$P(w_2) = 298/1000 = 0.298$

$P(w_3) = 500/1000 = 0.500$

Computed Posterior Probabilities for each sample in a similar way as in Question 2.

Training Error Rate = 0.128

Testing Error Rate = 0.135

Conclusion

Using the Maximum Likelihood Estimates for means and covariances computed for training data set, found the error count and rate for testing data.

References

Pattern Classification, 2nd edition, R. O. Duda, P. E. Hart, and D. G.

Stork, Wiley-Interscience.