Implementing Minimum Error Rate Classifier

Rahat Bin Osman
160204083
dept. of CSE
Ahsanullah university of Science and Technology
Dhaka, Bangladesh

***Abstract*—The experiment is to implement the Minimum Error Rate Classifier to classify some sample points using the posterior probabilities. We used Multivariate Gaussian Density Function for calculating the likelihood probabilities and classifying the data points.**

***Index Terms*—Likelihood ,Prior, Posterior, Probability Density Function, Sigma, Mean, Variance, Co-variance, Quadratic function**

I. Introduction

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification. As we know, the minimum error rate classifier tries to minimize the error. In the experiment six sample data were given. The likelihood probabilities of a sample are determinate by the normal distribution.Any normal distribution can be express by two significant parameter μ and Σ.

II. Experimental Design / Methodology

1) Posterior , Prior and Likelihood: The decision result based on Bayes depends on the class conditional probability and the prior probability of the training samples given. Bayesian estimates the probability that a certain event may occur in the future through probabilistic knowledge and statistics of existing data, using the method of probability. The posterior probability is a multiplication of likelihood and prior divide by marginal probability. The posterior probability is elaborated as:

$$If\ p(w_1|x) > p(w_2)|x)\ then\ x\epsilon w_1$$

$$If\ p(w_1|x) < p(w_2)|x)\ then\ x\epsilon w_2$$

We can calculate posterior probabilities with the help of likelihood. So, it is as follows:

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

$$=> lnP(w_i|x) = lnP(x|w_i).P(w_i)$$

$$=> lnP(w_i|x) = lnP(x|w_i) + P(w_i)$$

Here $P(x|w_i)$ and $P(w_i)$ is likelihood and prior.
We use the following equation for normal distribution:

$$N_k(x_i|\mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^D|\Sigma_k|)}} e^{\frac{-1}{2}((x_i-\mu_k)^T)\Sigma_k^{-1}(x_i-\mu_k)^T))}$$

Taking ln we can get,

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

Here, N_k is normal disctribution,μ_k is mean, Σ is co-variance matrix,x_i data, and D is 2 for this experiment.

So,the decision boundary is :

$$g_1(x) = g_2(x)$$

$$=> p(w_1|x) = p(w_2|x)$$

$$=> p(w_1|x) - p(w_2|x) = 0$$

$$=> P(x|w_1).P(w_1) - P(x|w_2).P(w_2) = 0$$

Taking ln we can get,

$$=> lnP(x|w_1).P(w_1) - lnP(x|w_2).P(w_2) = 0$$

$$=> lnP(x|w_1) + lnP(w_1) - lnP(x|w_2) - lnP(w_2) = 0$$

$$=> lnP(x|w_1)/lnP(x|w_2) - lnP(w_2)/lnP(w_1) = 0$$

This is the equation of a decision boundary for minimum error rate classifier..

A. Classify and Plotting the data

We used the Normal Distribution formula to classify the sample input and plot them using different .

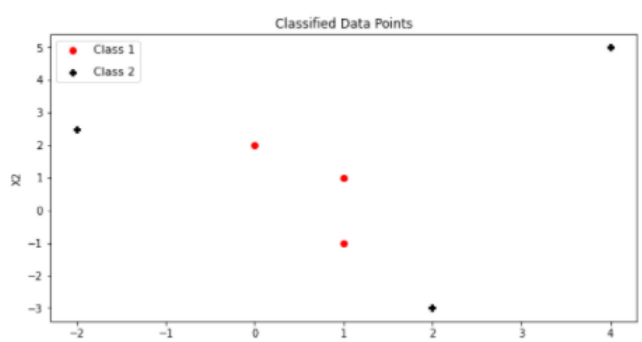

Fig. 1. Plotting the sample data

After plotting the points we can see that the two classes can not be separated with a linear boundary. We need to take them to a higher dimension so that we can possibly separate them linearly.

B. Decision Boundary in 2D

We cannot simplify the decision boundary equation, but we can rearrange to match a quadratic equation pattern. ($\Sigma_i = arbiraty$)

$g_i(x) = x^t W_i x + w_i^t x + w_{i0}$

Where, for Class, $i = 2$

$W_i = \Sigma_1^{-1} - \Sigma_2^{-1}$

$w_i = 2 * (\Sigma_1^{-1}\mu_1 - \Sigma_2^{-1}\mu_2)$

$w_{i0} = -ln(\frac{\Sigma_1}{\Sigma_2}) + 2 * ln(\frac{P(x_1)}{P(x_2)}) + \mu_1^t\Sigma_1^{-1}\mu_1 + \mu_2^t\Sigma_2^{-1}\mu_2$

now using the quadratic equation the plotting the decision boundary we get:

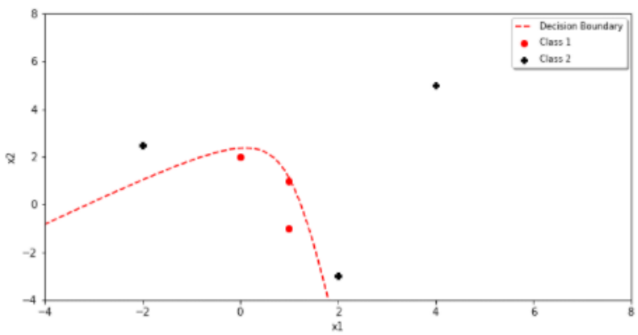

Fig. 2. Decision boundary in 2D

C. Drawing the Contour and 3D plot

Using Matplotlib 3d we added the Z axis and plotted all the new variables and contour of the graph.

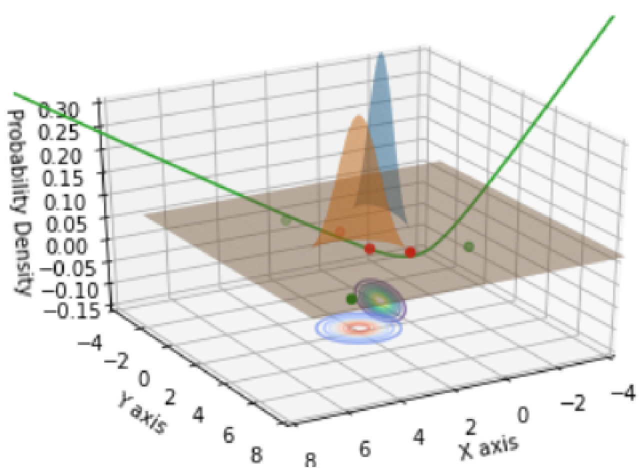

Fig. 3. Contour plotting in 3D

III. Result Analysis

Thus the discriminant function of Gaussian Density Function when general case Σ_i are arbitrary is Quadratic. Therefore the decision boundaries are quadratic (ellipses and paraboloids).

IV. Conclusion

Observing the results, we can come to a conclusion that, Minimum Error Rate Classifier can help classify multiple classes that are not linearly separable. As Gaussian Distribution is more natural, means real life scenarios is very much likely to be same and will help classify them with better accuracy than other linear classifiers.