

Assignment

1. Generate the datasets A and B in \mathbb{R}^2 with each of them consisting 2000 data points from normal distribution. The dataset A and B has been drawn from the $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$. Let us fix the $\mu_1 = [-1, 1]$ and $\mu_2 = [2, 2]$. Separate the 250 data points from each classes as testing set. Plot the optimal Bayesian decision boundary for following cases.

$$(i) \quad \Sigma_1 = \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$(ii) \quad \Sigma_1 = \Sigma_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

2. Write a function implementing the logistic regression model using the gradient descent method using the dataset generated for question 1. Plot the decision boundary obtained by the logistic regression. Compare it with the Bayesian decision boundary.
3. Consider the Iris dataset. The dataset contains three types of flower described by the four features. Consider only the data points with label 1 and 2. Divide the dataset into training and testing set in the ration 8:1:1. Use the training set to train the logistic regression model. Use the validation set to tune the parameter values of your model. Finally obtain the accuracy on the test set.
4. Learn about the different evaluation measure of a classification model namely Accuracy, Precision, Recall, Confusion matrix, F-Score and their significance. Write a brief note about them and submit it (Not less than 200 words). Compute all of them for problem 3.