

Assignment 7

Generate the datasets A and B in \mathbb{R}^2 , each consisting of 2000 data points from a 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 $\mu_2 = [2, 1]$. Separate the 500 data points from each class as a testing set. Plot the optimal Bayesian decision boundary.

1. Write a function implementing the logistic regression model using the gradient descent method. Obtain the best accuracy on the test set by tuning the value of the parameter λ . Plot the decision boundary obtained by the logistic regression.
2. Consider the Iris dataset. The dataset contains three types of flowers described by the four features. Consider only the data points with labels 1 and 2. Divide the dataset into training, testing, and validation in the ratio 8:1:1. Use the training set to train the logistic regression model. Use the validation set to tune the parameter value λ . Finally, obtain the accuracy of the test set.