Part D – Comparative Study

Find the average performance metrics of 10 random training and testing splits should be considered for this comparative study and Find out the best performing model and if possible explain the reasons for that model to outcast other models?

Models - PM1, PM2, PM3, PM4, FLDM1, FLDM2, LR1 and LR2

	PM1	PM2	PM3	PM4	FLDM1	FLDM2	LR1	LR2
Accuracies	91.489%	90.425%	96.276%	96.276%	95.744%	95.744%	91.255%	90.152%

Perceptron, Fisher Linear Discriminant Analysis (FLDA), and Logistic Regression are all popular linear classification algorithms. They all seek to separate a set of data points into different classes using a linear boundary.

Perceptron:

The Perceptron algorithm is an iterative algorithm that learns a linear discriminant function by updating the weights of the input features to minimize the misclassification rate. The algorithm is simple and easy to implement, but it can only handle linearly separable datasets. The perceptron algorithm works by iteratively adjusting the weights of the input features until the data points are correctly classified.

Fisher Linear Discriminant Analysis (FLDA):

FLDA is a supervised linear dimensionality reduction and classification technique that aims to find a linear combination of features that maximally separates the classes. The idea behind FLDA is to project the data onto a lower-dimensional subspace where the classes are well-separated. The goal is to find a linear projection that maximizes the ratio of the between-class variance to the within-class variance. FLDA can be used to classify data that are linearly separable, but it can also be used on datasets that are not linearly separable.

Logistic Regression:

Logistic Regression is a popular linear classification algorithm that models the probability of a data point belonging to a certain class. Logistic Regression uses a logistic function to model the relationship between the input features and the output class probabilities. Logistic Regression can handle both linearly separable and non-linearly separable datasets. The model is trained using the maximum likelihood estimation method to minimize the cross-entropy loss function.

In terms of performance, the best model depends on the characteristics of the dataset. If the dataset is linearly separable, all three models should perform well. However, if the data is not linearly separable, Logistic Regression and FLDA would outperform the Perceptron algorithm.

FLDA would perform better than Logistic Regression if the between-class variance is high and the within-class variance is low.

We observe that we get similar accuracy from three different models as the data set is linearly separable. Overall, Logistic Regression is a popular choice for binary classification tasks due to its flexibility, ease of implementation, and good performance on a wide range of datasets. However, it's important to note that each algorithm has its own strengths and weaknesses, and it's always a good idea to test multiple models to find the best one for your specific dataset.