**Using Semi-Supervised Learning to Label the Fashion MNIST Dataset**

|  |  |  |
| --- | --- | --- |
| **Case** | **Logistic Model Accuracy** | **Time** |
| Normal Model (Full train data) | 0.83 | 3min 17s |
| Logistic regression model trained with only 25 random training points | 0.54 | 113 ms |
| Logistic regression model trained with 25 centroids of K-Means | 0.62 | 157 ms |
| Logistic regression model trained with full train data, points labelled based on centroids | 0.62 | 2min 36s |
| Logistic regression model trained with only 75 percentile train data, points labelled based on centroids | 0.61 | 2min 13s |

With 25 training points selected by using K-means centroid, instead of 25 random points, we increased accuracy from 54% to 62%. The next two cases of labelling the remaining training points based on the K-means centroid did not help much in increasing the accuracy, unlike in the MNIST dataset.