1   Team Details

Group 19 - Pavan Ravishankar(**CS17S026**), Gouthaman KV(**CS16D018**)

2   Perceptron

Data is normalized per feature per class for quick convergence as weight vector is uniformly changed across dimensions. Batch gradient descent is used to get 100% accuracy. Stochastic gradient descent with shuffling gives 64% accuracy. Since time difference observed is not significant, we are using Batch gradient descent as to not comprimise on accuracy.

One vs One classifier(3 classifiers) is used to train perceptron. We observed that this converged [8] for all 3 classifiers in less than 200 epochs contrary to one vs rest approach, which did not converge even after 500 epochs for tall building vs rest classifier. This gives us insight into how the data might be present. Coast, Mountain and Tall Building data points are pairwise linearly separable but Tall Building points are not so far to be linearly separable as given by one vs rest approach. [8]

3   Linear Discriminant Analysis

Data is scale normalized per feature per class and projected to C-1 dimension space as rank of S_b is C-1. Programatically one can confirm that project space is C-1 as top 2 eigen values are 0.29, 0.11 and rest are lesser than 10^7. One can observe that data overlaps [4] and calculating accuracy by putting a unimodal bayesian on projected trained data gives 64.41% accuracy. Performing whitening transformation per class, mean shifting after whitening and performing LDA gives linearly separable data given accuracy of 100%. [3]

4   Parzen window

Image data is unfolded into features and normalized by standard deviation per feature per class. Gaussian kernel is placed on feature point and number of points of a particular class is calculated by summing over gaussian distances. Feature point is assigned a score for every class and classified to class with maximum score. Image is assigned a score for every class by summing over scores of features and classified to class with maximum score (likelihood approach).

h-value was varied from 0.005 to 100 and found to give 100% accuracy for values less than 0.5. So we chose h-value of 0.5 which does not give too much smoothening(large h) and does not contain few points(small h). Calculating exponentiation for small h gives underflow error which is addressed by using logsumexp.

5 Nature of Image Data

When GMM was used to classify image data, we found that with just one mixture per class we are able to classify with 100% accuracy leading to a hypothesis that data is unimodal. Classifiers used in this class further confirm this hypothesis that data is unimodal as Parzen window gives 100% accuracy and Perceptron converges for all one-vs one classifiers. Further, perceptron fails to converge Tall building vs rest classifier which makes Tall Building class close to other 2 classes. Our view of how the data might be is illustrated here. [8]
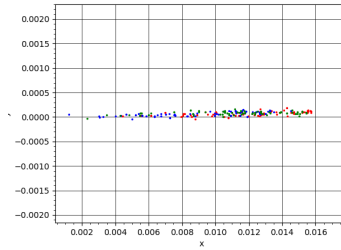
Figure 1: LDA Before normalization
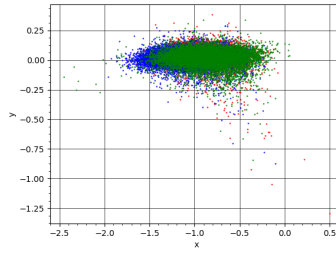
Figure 2: Without Whitening
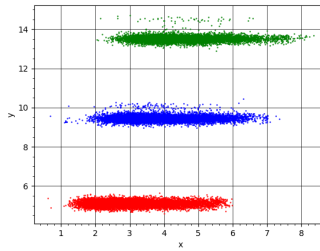
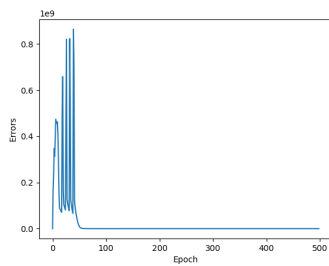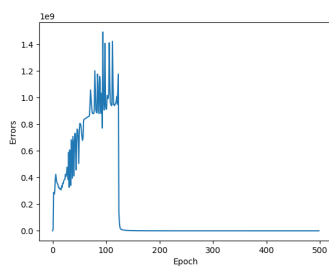Figure 3: Whitening Linearly separable
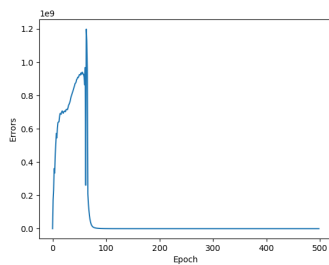
Figure 4: One vs Two

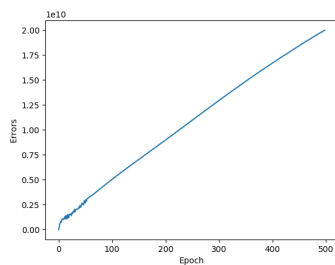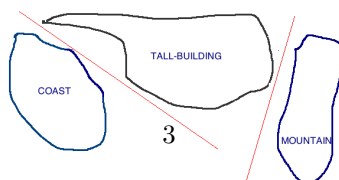Figure 5: One vs Third

Figure 6: Two vs Third

Figure 7: One vs Rest Non Converging

Figure 8: Data Visualization