Project Part – I

-Janaki Venkata Ramachandra Sai Nayani

(jnayani@asu.edu)

Input: Data Set Containing Images

Classes: Images of 3 and 7

Goal: Classify the Data Set and find bayes Error

Task-I Feature Extraction

We extract features from the given image, we take features as skewness and bright to dark ratio for a given threshold.

Before calculation we flatten the image and convert the 28x28 Matrix into a linear array and then we calculate the using the formula g1=∑Ni=1(Yi−Y¯)3/Ns3.

We calculate the ratio of bright pixels to dark pixels using a Threshold T, if pixel value is greater than the threshold then it is bright, dark otherwise.

We Now do Normalization of the each of the column by calculating mean and variance and normalize using

Yi = [y1i, y2i]t = [(ki – M1)/S1, (ri – M2)/S2]t

We will now have the data in Nx2 Matrix where N is number of training samples.

Task – II Density Estimation

We now extract the distributions for the integer 3 and 7 and calculate their Maximum Likelihood estimations. As, the distribution is Normal distribution, the MLE of the distribution is mean and variance of the distribution

T=150

Mean of the Distribution of Number 3 : [-0.37951983 0.35854635]

Covariance of Distribution of Number 3 : [[0.8437565 -0.91737529]

[-0.91737529 1.10355626]]

Mean of Distribution of Number 7:   [0.37158471 -0.35104975]

Covariance of distribution of Number 7: [[0.87417217 -0.72010492]

[-0.72010492 0.64980588]]

T=200

Mean of the Distribution of Number 3   : [-0.37951983 0.30674367]

Covariance of Distribution of Number 3 : [[0.8437565 -0.91624532]

[-0.91624532 1.136679]]

Mean of Distribution of Number 7: [0.37158471 -0.30033017]

Covariance of distribution of Number 7: [[0.87417217 -0.72986478]

[-0.72986478 0.68416819]]

Task-III Classification and Bayes Error

We calculate classify the data and calculate the Bayesian error for each classification.

We first compute p(x|wi) using the normal distribution of density estimation of the each 3 and 7 and the use the formula : prior*likelihood/evidence

Evidence = Sum of all the prior*likelihood for all available classes

T=150

Type of Data	Prior 1	Prior 2	Bayes Error
Train	0.5	0.5	0.31409709381432627
Test	0.5	0.5	0.30647387097742845
Train	0.3	0.7	0.22237422922954495
Test	0.3	0.7	0.22088605489469224

T=200

Type of Data	Prior 1	Prior 2	Bayes Error
Train	0.5	0.5	0. 30569370083304287
Test	0.5	0.5	0. 3013463267107516
Train	0.3	0.7	0. 25196358680103814
Test	0.3	0.7	0. 2539460350502862

Conclusion, The performs best when the Error is less, the error is observed when prior 1=0.3 and prior 2=0.7 and for T=150