Ravisutha Sakrepatna Srinivasamurthy

PART1: ANALYSIS OF RESULTS FROM TAKE-HOME #1

The result from the take-home #1 is analyzed in this section. The correct format of the output "3 1 2 3 2 1" is compared with the obtained result. Confusion matrix is tabulated as shown in table 1. Probability of error is found to be 8.8%.

a. Confusion Matrix

	A	**B**	**C**
A	4536	313	151
B	581	4393	26
C	210	40	4750

Table 1. Results from take-home #1

b. Probability of Error

$$Classify\ x\ as\ A \wedge x\ belongs$$
$$Classify\ x\ as\ A \wedge x\ belongs$$
$$Classify\ x\ as\ B \wedge x\ belongs$$
$$Classify\ x\ as\ B \wedge x\ belongs$$
$$Classify\ x\ as\ C \wedge x\ belongs$$
$$Classify\ x\ as\ C \wedge x\ belongs$$
$$P(error) = P(¿B) + P(¿C) + P(¿A) + P(¿C) + P(¿A) + P(¿B)$$

$$P(error) = \left(\frac{(464+607+250)}{15000}\right) * 100 = 8.80$$

Ravisutha Sakrepatna Srinivasamurthy

PART2: HO-KASHYAP HYPERPLANAR CLASSIFIER

The Ho-Kayshap iterative algorithm is used to obtain the three hyperplanes by considering two classes at once. The three hyperplanes are as follows:

a. Hyperplane seperating ClassA and ClassB

$w1 = [0.0076, -0.0042, 0.0009, -0.0166, -0.7432] * [x4, x3, x2, x1, 1]^T$

b. Hyperplane separating ClassA and ClassC

$w2 = [0.0011, -0.0029, -0.0389, 0.0000, -0.2101] * [x4, x3, x2, x1, 1]^T$

c. Hyperplane separating ClassB and ClassC

$w3 = [-0.0035, -0.0014, -0.0564, 0.0130, -0.0806] * [x4, x3, x2, x1, 1]^T$

After calculating the hyperplanes, the training set is subjected to classification. The results are tabulated in table 2. The probability of error is found to be 18.66%.

Confusion matrix for training set H

	A	B	C
A	3985	603	412
B	1156	3844	0
C	603	25	4372

Table2. Results for training set using Ho-Kashyap

Probability of Error for training input (H)

$$P(errror) = \left(\frac{1015 + 1156 + 628}{15000}\right) * 100 = 18.66$$

Now, the testing set is subjected to classification. The results are tabulated in table 3 and probability of error is found to be 18.78%.

Confusion matrix for testing set S_T

	A	B	C
A	3918	647	435
B	1153	3847	0
C	560	22	4418

Table3. Results for testing set using Ho-Kashyap

Probability of Error for testing input (S_T)

$$P(error) = \left(\frac{(1082+1153+582)}{15000} \right) * 100 = 18.78$$

PART3: K-NNR STRATEGIES

For K = 1

To determine a vector that is closest to given vector, distance from given vector to all vectors in training set is calculated (Brute Force). The vector which is closest to given vector is picked. Class of given vector is same as the closest vector. Result is tabulated in the confusion matrix.

a. **Confusion Matrix**

	A	**B**	**C**
A	4101	631	268
B	654	4264	82
C	289	63	4648

Table 4. NNR with K = 1

b. **Probability of Error**

$$P(error) = \left(\frac{(899+736+352)}{15000} \right) * 100 = 13.24$$

For K = 3

Three closest vectors are picked and voting decides the class of the given vector.

a. **Confusion Matrix**

	A	**B**	**C**
A	4305	482	213
B	659	4293	48
C	265	42	4693

Table 5. NNR with K = 3

$$P(error) = \left(\frac{(695+707+307)}{15000} \right) * 100 = 11.39$$

b. **Probability of Error**

For K = 5

a. **Confusion Matrix**

	A	B	C
A	4368	431	201
B	614	4340	46
C	243	51	4706

Table 6. NNR with K = 5

b. **Probability of Error**

$$P(error) = \left(\frac{(632+660+294)}{15000}\right) * 100 = 10.57$$

PART4: PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA is used to reduce the number of features in the given data set. The way to choose best features is to select features having maximum covariance. It is difficult to interpret the variance when there is correlation between features. Diagonalization of covariance matrix comes in handy to pick the features with maximum covariance. The eigen vectors of covariance matrix with max eigen values are made as the new set of basis. Hence, the input data is also subjected to 'change of basis'. This results in reduced number of features (2 features). This is now classified using Bayesian classifier. Results are tabulated in table 3. Probability of error is found to be 25.62% which is not bad given that only two features are considered.

For testing set (S_T)

a. **Confusion Matrix**

	A	B	C
A	4016	381	603
B	1063	3561	376
C	71	1350	3579

Table 7. Results using reduced data set

b. **Probability of Error**

$$P(error) = \left(\frac{(984+1439+1421)}{15000}\right) * 100 = 25.62$$

Ravisutha Sakrepatna Srinivasamurthy

Comparing different classification approaches

From Part1, Bayesian classifier is better compared to all other approaches (considered up to this point) with an error percentage of 8.80. But this method requires estimation of parameters. Representing the given data in a closed form is not always possible.

In Non-parametric approach, the NNR with K = 5 is better than other approaches. But this approach is computationally inefficient. As K increases, the error percentage decreases.

Even though Ho-Kayshap's error percentage is not as good as Bayesian, this method is more practical. This is because there is no need for closed form expression as is the case with Bayesian.

PCA offers dimensionality reduction. Redundant features can be eliminated using this method. This results in low resource consumption and faster classification times. However, error percentage will take a hit with this procedure.