Work Origination Certification

By submitting this document, I, the author of this deliverable, certify that

1. I have reviewed and understood the Academic Honesty section of the current version of FITs Student Handbook available at http://www.fit.edu/studenthandbook/, which discusses academic dishonesty (plagiarism, cheating, miscellaneous misconduct, etc.)

2. The content of this report reflects my personal work and, in cases it is not, the source(s) of the relevant material has/have been appropriately acknowledged after it has been first approved by the courses instructor.

3. In preparing and compiling all this report material, I have not collaborated with anyone and I have not received any type of help from anyone but from the courses instructor.

Shreenidhi Sudhakar

Full Name (please PRINT)

Shreenidhi

Signature

10/03/2016

Date

Task 1

Let

- C_1 : class 1 Labels
- C_2 : class 2 Labels
- pFlip : probability of flip
- $\underline{x}$: feature space.

a) Prior probabilities :

$$p(C_1) = 0.5(1-pFlip) + 0.5 \cdot pFlip = 0.5$$
$$p(C_2) = 0.5(1-pFlip) + 0.5 \cdot pFlip = 0.5$$

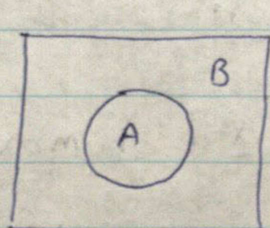
class conditional densities :

$$f(\underline{x}|C_1) = \begin{cases} 2 * (1-pFlip) & , \underline{x} \in A \\ 2 * pFlip & , \underline{x} \in B \end{cases}$$

where,

A = area of Circle $\Rightarrow$ Region 1 (R1)

B = area of square - area of circle $\Rightarrow$ Region 2 (R2)



$\text{111}'\text{y}$

$$f(\underline{x}|\omega_2) = \begin{cases} 2 \cdot pFlip & , \underline{x} \in A \\ 2 * (1 - pFlip) & , \underline{x} \in B \end{cases}$$

b) <u>Posterior</u> <u>Class</u> <u>Probability</u>

$$P(c_1/\underline{x}) = \frac{f(\underline{x}|c_1)\,P(c_1)}{f(\underline{x})} \qquad [\because \text{Bayes Theorem}]$$

$$= \frac{f(\underline{x}|c_1)\,P(c_1)}{f(\underline{x}|c_1)\,P(c_1) + f(\underline{x}|c_2)\,P(c_2)}$$

$$= \frac{(x \in A)(1 - pFlip) + (x \in B)\,pFlip}{(x \in A)\cdot(1 - pFlip) + (x \in B)\,pFlip + (x \in A)\,pFlip + (x \in B)(1 - pFlip)}$$

$$\therefore P(c_1/x) = \begin{cases} 1 - pFlip & , \text{if } \underline{x} \in A \\ pFlip & , \text{if } \underline{x} \in B \end{cases}$$

<u>Optimal</u> <u>region</u>:

$$R_1 = \{ \underline{x} \mid \operatorname{argmax} P(c_1|\underline{x}) = 1 \}$$

$$\Rightarrow R_1 = \{ \underline{x} : x \in A \}$$

III'ly - Similarly

III'ly,

$$P(C_2|\underline{x}) = \frac{f(\underline{x}|C_2)\,P(C_2)}{f(\underline{x})} \quad [\because \text{Bayes Theorem}]$$

$$= \frac{f(\underline{x}|C_2)\,P(C_2)}{f(\underline{x}|C_1)\,P(C_1) + f(\underline{x}|C_2)\,P(C_2)}$$

$$= \frac{(x \in A) * pFlip + (x \in B) * (1 - pFlip)}{(x \in A)(1 - pFlip) + (x \in B)\,pFlip + (x \in A)(pFlip) + (x \in B)(1 - pFlip)}$$

$$\therefore P(C_2|x) = \begin{cases} pFlip & , \text{ if } \underline{x} \in A \\ 1 - pFlip & , \text{ if } \underline{x} \in B \end{cases}$$

<u>Optimal</u> <u>Region</u>

$$R_2 = \{ \underline{x} \mid \operatorname{argmax} P(C_2|x) = 1 \}$$

$$\Rightarrow R_2 = \{ \underline{x} : x \in B \}$$

d) $P(\text{error}) = P(\hat{c}_1|c_2)\,P(c_2) + P(\hat{c}_2|c_1)\,P(c_1)$

$= (pFlip * 0.5) + (pFlip * 0.5)$

$\therefore P(\text{error}) = \underline{pFlip}$

Thus, $\boxed{\text{Bayes error} = pFlip}$

Task 6

We know that

CDF of Gaussian function is given by :

$$\Phi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} * \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \cancel{\frac{1}{\sqrt{2\pi}\,\sigma} * \frac{1}{2}\left(1 + \operatorname{erf}\left(x/\sqrt{2}\right)\right)}$$

$$= \cancel{1/\sqrt{2\pi}\,\sigma} * \frac{1}{2}\left[1 + \operatorname{erf}\frac{(x-\mu)}{\sqrt{2} * \sigma}\right]$$

where,

erf = error function.

Let,

$P(C_1)$ = prior probability of Class 1

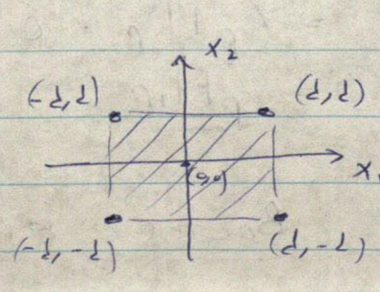
$P(C_2)$ = prior probability of Class 2

$P(C_3)$ = prior probability of Class 3

$P(C_4)$ = prior probability of Class 4

$\underline{x}$ = feature space in 2D ie, x1, x2.

Given,



→ misclassification region

Since, given Gaussians are isotropic and have common variance

$\Rightarrow P(C_1) = P(C_2) = P(C_3) = P(C_4) = 1/4$

and,

$$f(x|C_1) = f(x|C_2) = f(x|C_3) = f(x|C_4)$$

Thus,

Bayes error

$$= 1 - P(C_1) * f(x|C_1)$$

$$= 1 - \frac{1}{4} * \Phi(x_1; \mu_1, \sigma_1) * \Phi(x_2; \mu_2, \sigma)$$

$$= 1 - \frac{1}{4} * \Phi(x_1; d, \sigma_1) * \Phi(x_2; d, \sigma_2)$$

$$= 1 - \frac{1}{4} * \left[1 + \operatorname{erf}\left(\frac{-d}{\sigma * \sqrt{2}}\right)\right]\left[1 + \operatorname{erf}\left(\frac{-d}{\sigma\sqrt{2}}\right)\right]$$

$$\therefore \text{Bayes Error} = 1 - \frac{1}{4}\left[1 - \operatorname{erf}\frac{d}{\sqrt{2} * \sigma}\right]^2$$

In the Graphs

Class 1 is represented by Blue ‘*’ and Blue Decision Region

Class 2 is represented by Red ‘*’ and Red Decision Region

Black Region represents decision regions, in case of a tie.

Blue Circle Represents Theoretical Decision Boundary.

Note:

Results obtained for all questions may vary slightly on different runs of each Task. This is observed as we deal with random data in each run.

Task 1

Run Task1.m file to observe Graphs.

c**)** The constraint pFlip < 0.5 is to ensure that misclassification error (Bayes error) is kept to minimal. At pFlip = 0.5 the chance of misclassification is equal to 0.5 (approx.) as classifying labels based on pseudo random number generated to be <= 0.5 is half. When pFlip increase towards 1 from 0.6, misclassification error increases. This is because condition to flip the labels with respect to pFlip, is met most of the times.

e) Using MATLAB for calculations, results in the table are obtained:

Let

x = Feature Space

c1 = Class1, c2 = Class 2

R1 = Region 1, R2 = Region2

pFlip	**0.0**	**0.1**	**0.2**	**0.3**	**0.4**
p(c1)	0.5	0.5	0.47	0.52	0.60
p(c2)	0.5	0.5	0.53	0.48	0.40
p(c1|x,R1)	1.00	0.94	0.83	0.69	0.60
p(c2|x,R1)	0.00	0.06	0.17	0.31	0.40
p(c1|x,R2)=0.00	0.00	0.16	0.23	0.33	0.42
p(c2|x,R2)=1.00	1.00	0.84	0.77	0.67	0.57
Bayes Error	0.00	0.06	0.17	0.31	0.40

It’s found that, Theoretical **Bayes Error** Results approximately equals Predicted **Bayes Error** Output.

Task 2

Run Task2.m file to observe Graphs.

Let

N = Training Examples Count

a) It is seen that as N increases, Decision regions are defined better. As a result, occurrence of cases of tie decreases.

b) It is seen that as N increase, Decision regions are defined better. Compared to K =1, here when K = 5, # of decision regions obtained are less. This shows that as # of K Nearest Neighbors increase, Misclassification error decrease.

c) Results are as below:

- It is seen that as # of training examples increase, misclassification error decreases and # of sub-divisions of decision regions formed also decrease (occurrence of tie decrease).
- When N=10 and norm equals L-inf, no decision region is observed. This is because L-inf norm always takes the distance of maximum value. Since # of training patterns are less, KNN always ends up classifying the data which is farthest away from it as class label. Thus, no decision region is seen when N = 10.

Task 3

Run Task3.m file to observe Graphs.

In the Graphs,

X-Axis represents Training Sample Size in log 10 scale

Y-Axis represents Misclassification Error

Let

N = Training Examples Count

a) It's observed that average error rate is around 0.38. The error values vary between 0.36 - 0.43.

b) It's seen that as # of training patterns increase, misclassification error decrease. In the graph, when training pattern size = 1000, average error rate is observed to be 0.18. This is a sharp decrease when compared to average error of 0.38 obtained for training pattern size = 10.

c) Results are as below :

- It's observed that average error rate is around 0.46 when training pattern size = 10. The error values vary between 0.41- 0.48.
- It's seen that as # of training patterns increase, misclassification error decrease. In the graph, when training pattern size = 1000, average error rate is observed to be 0.13.
- In contrast to K=1, we have seen that error variation with respect training pattern count is more here. This is expected as we compare our result with K-neighbors to classify test data. As observed,

misclassification error decreases more rapidly when compared to case of K=1. This states that as confidence of vote from K-neighbor increase, KNN algorithm accuracy also increase.

Task 4

a) Run Task4.m and observe Figures from 1 - 5 :

- When Training Pattern = 10

Spread	**Decision region**	**Error Rate**
0.001	Doesn't cover all Data Sample	High
0.1	Covers all Data Sample	Medium
0.75	Covers all Data Sample	Medium

- When Training Pattern = 30

Spread	**Decision region**	**Error Rate**
0.001	Covers all Data Sample	Medium
0.1	Covers all Data Sample	Low
0.75	Covers all Data Sample	High

- When Training Pattern = 50

Spread	Decision region	Error Rate
0.001	Covers all Data Sample	Medium
0.1	Covers all Data Sample	Low
0.75	Doesn't covers any Data Sample	Medium

- When Training Pattern = 75

Spread	Decision region	Error Rate
0.001	Covers all Data Sample	Medium
0.1	Covers all Data Sample	Low
0.75	Doesn't covers any Data Sample	High

- When Training Pattern = 100

Spread	Decision region	Error Rate
0.001	Covers all Data Sample	Medium
0.1	Covers all Data Sample	Low
0.75	Doesn't cover any Data Sample	High

It’s observed from the graphs that when spread values and training data samples are low, misclassification error of PWC Model. Here, we obtain a good PWC model when Spread = 0.1 and training Data sample is between 50 and 100, preferably 100.

b) Run Task4.m and observe Figures from 6 - 10 :

- When Training Pattern = 10

Spread	Decision region	Error Rate
0.001	Covers all Data Sample	Medium
0.1	Covers all Data Sample	Medium
0.75	Covers all Data Sample	Medium

- When Training Pattern = 30

Spread	Decision region	Error Rate
0.001	Covers all Data Sample	Low
0.1	Covers all Data Sample	Low
0.75	Doesn’t covers all Data Sample	High

- When Training Pattern = 50

Spread	Decision region	Error Rate
0.001	Covers all Data Sample	Low
0.1	Covers all Data Sample	Low
0.75	Doesn’t cover any Data Sample	High

- When Training Pattern = 75

Spread	Decision region	Error Rate
0.001	Covers all Data Sample	Low
0.1	Covers all Data Sample	Low
0.75	Covers all Data Sample	Medium

- When Training Pattern = 100

Spread	Decision region	Error Rate
0.001	Covers all Data Sample	Low
0.1	Covers all Data Sample	Low
0.75	Doesn’t cover any Data Sample	High

It’s observed from the graphs that when spread values and training data samples are low, misclassification error of PWC Model. Here, we obtain a good PWC model when Spread = 0.1 and training Data sample is between 50 and 100, preferably 100.

Results obtained for Gaussian and Square Sinc Kernel matches under optimal condition of spread and training sample values we have found. In different set of combination of these values, result varies. Square Sinc Kernel works well to an extent for low spread value compared to Gaussian Kernel.

c) Run Task4.m and observe Figures from 1 - 5 after making the below changes on Line 17 of Task4.m:

Replace the code from : `spreadValues = [.005 0.1 0.75];`
To : `spreadValues = [.25 0.5 0.75]`

It's observed that the spread of 3D plot obtained corresponds to classification of Class 1 labels. If the decision region is not formed (when spread value is very small or large) then 3D surface spread formed corresponds to a flat spread indicating that classification error is high. If decision region is formed, then 3D surface spread formed corresponds to distribution of Class 1 labels under the spread of its area.

Task 5

Run Task5.m file to observe the graphs.

a) Results are:

- Champion K-NN model is obtained when K = 5 with error = 0.29.
- Large Confidence Interval is obtained when K is between 4 – 6
- Small Confidence Interval is obtained when K > 6
- Training Set error is greater than Validation Set Error. This is because we use only 30 training patterns to predict the output. In Supervised Learning, Training Set accuracy is found to be zero as we test accuracy of a particular model only after training it. This behavior can be seen when we implement algorithms based on SVM and Neural Network.
- Error for best KNN model obtained is greater than Bayes error.
  - Theoretical Value : Bayes Error = 0.1 = pFlip
  - Obtained Least Error = 0.29

b) Results are as below:

- Champion PWC model is obtained when spread = 0.101 with error = 0.24.
- Training Set error is greater than Validation Set Error. This is because we use only 30 training patterns to predict the output. In Supervised Learning, Training Set accuracy is found to be zero as we test accuracy of a particular model only after training it. This behavior can be seen when we implement algorithms based on SVM and Neural Network.
- Error for best PWC model obtained is greater than Bayes error.
  - Theoretical Value : Bayes Error = 0.1 = pFlip
  - Obtained Least Error = 0.24