Problem 1

1(a)

Beta Distribution

Given,

$$f(x) = \left[\frac{x^{\alpha-1}(1-x)^{\beta-1}(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!}\right] \tag{1}$$

and

$$\beta = 1 \tag{2}$$

Substituting (2) in (1), we get,

$$f(x) = \alpha x^{\alpha-1} \tag{3}$$

Likelihood function L is given by,

$$L = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \alpha x_i^{\alpha-1}$$

Then,

$$\begin{aligned} ln(L) = ln(\prod_{i=1}^{n} \alpha x_i^{\alpha-1}) &= \sum_{i=1}^{n} ln(\alpha) + \sum_{i=1}^{n}(\alpha-1)ln(x_i) \\ &= nln(\alpha) + \sum_{i=1}^{n}(\alpha-1)ln(x_i) \end{aligned}$$

$$\frac{\partial lnL}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} ln(x_i) = 0$$

Hence MLE of α is,

$$\hat{\alpha} = -\frac{n}{\sum_{i=1}^{n} ln(x_i)}$$

Normal Distribution

Given,

$$f(x) = \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right] \tag{4}$$

and

$$\mu = \sigma^2 = \theta \tag{5}$$

Substituting (5) in (4), we get,

$$f(x) = \left[\frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-\theta)^2}{2\theta}}\right] \tag{6}$$

Likelihood function L is given by,

$$L = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x_i-\theta)^2}{2\theta}}\right]$$

Then,

$$\begin{aligned} ln(L) = ln(\prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x_i-\theta)^2}{2\theta}}\right]) &= -\frac{1}{2}\sum_{i=1}^{n} ln(2\pi\theta) - \sum_{i=1}^{n}\left(\frac{x_i^2}{2\theta}\right) - \sum_{i=1}^{n}\frac{\theta}{2} + \sum_{i=1}^{n} x_i \\ &= -\frac{n}{2} ln(2\pi\theta) - \sum_{i=1}^{n}\left(\frac{x_i^2}{2\theta}\right) - \frac{n\theta}{2} + \sum_{i=1}^{n} x_i \end{aligned}$$

$$\frac{\partial lnL}{\partial\theta} = -\frac{n}{2\theta} + \sum_{i=1}^{n}\left(\frac{x_i^2}{2\theta^2}\right) - \frac{n}{2} = 0$$

$$\Leftrightarrow n\theta^2 + n\theta - \sum_{i=1}^{n} x_i^2 = 0 \tag{7}$$

By solving the quadratic equation (7), the MLE of θ is,

$$\hat{\theta} = -\frac{-n \pm \sqrt{n^2 - 4n\sum_{i=1}^{n} x_i^2}}{2n}$$

1(b) Bias of KDE

Given,

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h} K\left(\frac{x - X_i}{h}\right) \tag{8}$$

Step 1:

$$\begin{aligned} E_{X1,\cdots,X_n}[\hat{f}(x)] &= \frac{1}{n}\sum_{i=1}^{n} E\left[\frac{1}{h} K\left(\frac{x - X_i}{h}\right)\right] \\ &= E\left[\frac{1}{h} K\left(\frac{x - X}{h}\right)\right] \\ &= \frac{1}{h}\int K\left(\frac{x-t}{h}\right) f(t) dt \end{aligned}$$

Hence,

$$E_{X1,\cdots,X_n}[\hat{f}(x)] = \frac{1}{h}\int K\left(\frac{x-t}{h}\right)f(t)dt \tag{9}$$

Step 2:
Let $z = \frac{x-t}{h} \Rightarrow t = x - hz$
Applying this in eqn (9) and by using Taylor's theorem, we get,

$$\begin{aligned} E[\hat{f}(x)] &= \int K(z)f(x-hz)dz \\ &= \int K(z)dz\left[f(x) - hzf'(x) + h^2z^2\frac{f''(x)}{2} - h^3z^3\frac{f'''(x)}{3} + \dots\right] \\ &= \int f(x)K(z)dz - \int hzf'(x)K(z)dz + \int h^2z^2\frac{f''(x)}{2}K(z)dz + \dots \\ &\left\{Using \int K(z)dz = 1, \int zK(z)dz = 0 \text{ and } \int z^2K(z)dz = \sigma_K^2\right\} \\ &= f(x) + \frac{h^2\sigma_K^2 f''(x)}{2} + o(h^2) \end{aligned}$$

$$\Rightarrow E[\hat{f}(x)] = f(x) + \frac{h^2\sigma_K^2 f''(x)}{2} + o(h^2) \tag{10}$$

Step 3:
From eqn (10), Bias is,

$$E[\hat{f}(x)] - f(x) = \frac{h^2\sigma_K^2 f''(x)}{2} + o(h^2)$$

Problem 2

2(a)

Normalize data

Based on the given data,
The Features are : x and y coordinates.

$x = \{10, -12, -9, 29, 32, 37, 8, 30, -18, -21\}$
$y = \{49, 38, 47, 19, 31, 38, 9, -28, -19, 12\}$
We compute Mean and Variance of each feature as follows:
For x

$$\mu_x = \frac{1}{n}\sum_n x_n = \frac{10-12-9+29+32+37+8+30-18-21}{10} = \mathbf{8.6}$$

$$\sigma_x^2 = \frac{1}{n-1}\sum_n (x_n - \mu_x)^2 = 505.38$$

$$\sigma_x = \sqrt{505.38} = \mathbf{22.48}$$

For y

$$\mu_y = \frac{1}{n}\sum_n y_n = \frac{49+38+47+19+31+38+9-28-19+12}{10} = \mathbf{19.6}$$

$$\sigma_y^2 = \frac{1}{n-1}\sum_n (y_n - \mu_y)^2 = \mathbf{705.38}$$

$$\sigma_y = \sqrt{705.38} = \mathbf{26.56}$$

Hence, the normalized and scaled coordinates are :

Mathematics	0.062	1.107
Mathematics	-0.92	0.69
Mathematics	-0.78	1.03
Electrical Engineering	0.91	-0.02
Electrical Engineering	1.04	0.43
Electrical Engineering	1.26	0.69
Computer Science	-0.027	-0.4
Computer Science	0.95	-1.79
Computer Science	-1.18	-1.45
Computer Science	-1.32	-0.29

Classify a student

Given coordinates of unclassified student : $x = 9, y = 18$
Using the previous data, the normalized coordinates are :
$x = 0.018, y = -0.06$
We calculate $L1$ and $L2$ using the formulae:

$$L1 = |x - x_n| + |y - y_n| \tag{11}$$

$$L2 = \sqrt{|x - x_n|^2 + |y - y_n|^2} \tag{12}$$

ID	CLASS	X-Coordinate(x_n)	Y-Coordinate(y_n)	L1 Distance	L2 Distance
1	Mathematics	0.062	1.107	1.211	1.168
2	Mathematics	-0.92	0.69	1.688	1.2
3	Mathematics	-0.78	1.03	1.888	1.351
4	Electrical Engineering	0.91	-0.02	0.932	0.893
5	Electrical Engineering	1.04	0.43	1.512	1.134
6	Electrical Engineering	1.26	0.69	1.992	1.451
7	Computer Science	-0.027	-0.4	0.385	0.344
8	Computer Science	0.95	-1.79	2.662	1.965
9	Computer Science	-1.18	-1.45	2.588	1.835
10	Computer Science	-1.32	-0.29	1.568	1.358

Using the above table,

For $K = 1$ and $L1$:
$nn(x) = \{ID(7)\}$
Hence the unknown Student belongs to **Computer Science**.

For $K = 3$ and $L1$:
$3nn(x) = \{ID(7), ID(4), ID(5)\}$
$\upsilon_{computerscience} = 1$, $\upsilon_{electricalengineering} = 2$
$y = argmax\upsilon = ElectricalEngineering$
Hence the unknown Student belongs to **Electrical Engineering**.

For $K = 1$ and $L2$:
$nn(x) = \{ID(7)\}$
Hence the unknown Student belongs to **Computer Science**.

For $K = 3$ and $L2$:
$3nn(x) = \{ID(7), ID(4), ID(1)\}$
$\upsilon_{computerscience} = 1$, $\upsilon_{electricalengineering} = 1$, $\upsilon_{mathematics} = 1$
Since we have a tie, we go for closest point which is Computer Science.
Hence the unknown Student belongs to **Computer Science**.

Comparision:
We see that in case of K=1, we did get the answer as Computer Science which is the class of the first nearest neighbor irrespective of L1 or L2 distance metric.

But in case of K=3, the second and third nearest neighbors changed depending on L1 or L2 metric. Hence we got different values.

2(b)

Calculate p(x)

By Total probability theorem, we have,

Given:

$$\sum_{n} K_c = K \tag{13}$$

$$p(x|y = c_i) = \frac{K_c}{N_c V} \tag{14}$$

$$p(y = c_i) = \frac{N_c}{N} \tag{15}$$

$$p(x) = \sum_{i=1}^{n} p(x|y = c_i) \times p(y = c_i)$$

$$= \frac{K_{c_1}}{N_{c_1} V} \times \frac{N_{c_1}}{N} + \frac{K_{c_2}}{N_{c_2} V} \times \frac{N_{c_2}}{N} + \cdots + \frac{K_{c_n}}{N_{c_n} V} \times \frac{N_{c_n}}{N}$$

$$= \frac{K_{c_1} + K_{c_2} + \cdots + K_{c_n}}{NV} = \frac{K}{NV}$$

$$\Leftrightarrow \mathbf{p(x)} = \frac{\mathbf{K}}{\mathbf{NV}} \tag{16}$$

Calculate p(Y=c|x)

By Bayes theorem, we have,

$$p(y = c|x) = \frac{p(x|y = c) \times p(y = c)}{p(x)}$$

$$= \frac{\left[\frac{K_c}{N_c V} \times \frac{N_c}{N}\right]}{\frac{K}{NV}} = \frac{K_c}{K}$$

$$\Leftrightarrow \mathbf{p(y = c|x)} = \frac{\mathbf{K_c}}{\mathbf{K}}$$

Problem 3

3(a)

We use 3 entropy formulas:

$$H[X] = -\sum_{k=1}^{K} P(x_k) log P(x_k) \tag{17}$$

$$H[Y|X] = \sum_{k} P(x_k) H[Y|x_k] \tag{18}$$

$$IG = H[Y] - H[Y|X] \tag{19}$$

Step 1 : Calculate H[Rainy,NotRainy]

$$H[Rainy, NotRainy] = -\left[\frac{36}{80} \times log\frac{36}{80} + \frac{44}{80} \times log\frac{44}{80}\right] = 0.6881 \tag{20}$$

Step 2 : Calculate IG for each attribute:
Temperature:
H[Rainy|Temperature]=

$$-\left[\frac{40}{80} \times \left[\frac{23}{40} \times log\frac{23}{40} + \frac{17}{40} \times log\frac{17}{40}\right] + \frac{40}{80} \times \left[\frac{13}{40} \times log\frac{13}{40} + \frac{27}{40} \times log\frac{27}{40}\right]\right] = 0.6561$$

$$\mathbf{IG[Temperature] = 0.6881 - 0.6561 = 0.0320}$$

Humidity:
H[Rainy|Humidity]=

$$-\left[\frac{40}{80} \times \left[\frac{23}{40} \times log\frac{23}{40} + \frac{17}{40} \times log\frac{17}{40}\right] + \frac{40}{80} \times \left[\frac{13}{40} \times log\frac{13}{40} + \frac{27}{40} \times log\frac{27}{40}\right]\right] = 0.6561$$

$$\mathbf{IG[Humidity] = 0.6881 - 0.6561 = 0.0320}$$

Sky Condition:
H[Rainy|Sky Condition]=

$$-\left[\frac{40}{80} \times \left[\frac{25}{40} \times log\frac{25}{40} + \frac{15}{40} \times log\frac{15}{40}\right] + \frac{40}{80} \times \left[\frac{11}{40} \times log\frac{11}{40} + \frac{29}{40} \times log\frac{29}{40}\right]\right] = 0.6248$$

$$\mathbf{IG[SkyCondition] = 0.6881 - 0.6248 = 0.0633}$$

At depth 1, *Sky Condition* has Highest IG. So, *Sky Condition* is the predictor feature.

Depth 2

If $SkyCondition = Cloudy$:

H[Rainy, NotRainy| Sky Condition = Cloudy] =

$$-\left[\frac{25}{40} \times log\frac{25}{40} + \frac{15}{40} \times log\frac{15}{40}\right] = 0.6616 \tag{21}$$

Temperature:
H[Rainy|Temperature, SkyCondition = Cloudy]=

$$-\left[\frac{20}{40} \times \left[\frac{15}{20} \times log\frac{15}{20} + \frac{5}{20} \times log\frac{5}{20}\right] + \frac{20}{40} \times \left[\frac{10}{20} \times log\frac{10}{20} + \frac{10}{20} \times log\frac{10}{20}\right]\right] = 0.6277$$

$$\mathbf{IG[Temperature, SkyCondition = Cloudy] = 0.6616 - 0.6277 = 0.0339}$$

Humidity:
H[Rainy|Humidity, SkyCondition = Cloudy]=

$$-\left[\frac{20}{40} \times \left[\frac{16}{20} \times log\frac{16}{20} + \frac{4}{20} \times log\frac{4}{20}\right] + \frac{20}{40} \times \left[\frac{9}{20} \times log\frac{9}{20} + \frac{11}{20} \times log\frac{11}{20}\right]\right] = 0.5943$$

$$\mathbf{IG[Humidity, SkyCondition = Cloudy] = 0.6616 - 0.5943 = 0.0673}$$

At depth 2 Cloudy, *Humidity* has Highest IG. So, *Humidity* is the predictor feature.

If $SkyCondition = Clear$:

H[Rainy, NotRainy| Sky Condition = Clear] =

$$-\left[\frac{11}{40} \times log\frac{11}{40} + \frac{29}{40} \times log\frac{29}{40}\right] = 0.5882 \tag{22}$$

Temperature:
H[Rainy|Temperature, SkyCondition = Clear]=

$$-\left[\frac{20}{40} \times \left[\frac{17}{20} \times log\frac{17}{20} + \frac{3}{20} \times log\frac{3}{20}\right] + \frac{20}{40} \times \left[\frac{8}{20} \times log\frac{8}{20} + \frac{12}{20} \times log\frac{12}{20}\right]\right] = 0.5479$$

$$\mathbf{IG[Temperature, SkyCondition = Clear] = 0.5882 - 0.5479 = 0.0403}$$

Humidity:
H[Rainy|Humidity, SkyCondition = Clear]=

$$-\left[\frac{20}{40} \times \left[\frac{7}{20} \times log\frac{7}{20} + \frac{13}{20} \times log\frac{13}{20}\right] + \frac{20}{40} \times \left[\frac{4}{20} \times log\frac{4}{20} + \frac{16}{20} \times log\frac{16}{20}\right]\right] = 0.5739$$

$$\mathbf{IG[Humidity, SkyCondition = Clear] = 0.5882 - 0.5739 = 0.0143}$$

At depth 2 Clear, *Temperature* has Highest IG. So, *Temperature* is the predictor feature.

3(b)

Given:

$$GiniIndex = \sum_{K=1}^{K} p_K(1 - p_K) \tag{23}$$

$$CrossEntropy = -\sum_{K=1}^{K} p_K log(p_k) \tag{24}$$

We need to prove that Gini Index is always less than or equal to entropy
OR

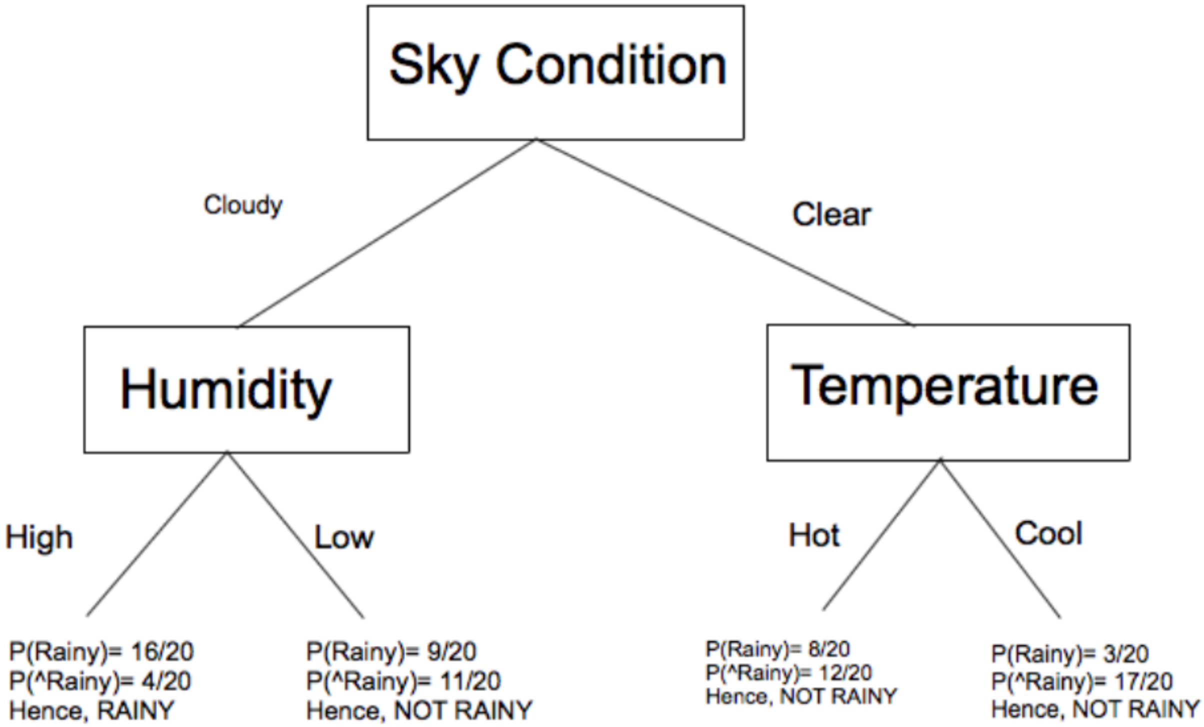

Figure 1: Gaussian Kernel

$$\sum_{K=1}^{K} p_K(1-p_K) \quad \leq \quad -\sum_{K=1}^{K} p_K log(p_k)$$

$$\Leftrightarrow \sum_{K=1}^{K}[p_K(1-p_K+log(p_K)] \leq 0$$

$$So \text{ for each } k \in K,$$

$$p_k(1-p_k+log(p_k) \leq (1-p_k+log(p_k) \; \{\because \mathbf{0 \leq p_k \leq 1}\}$$

$$\leq log(p_k) \; \{\because \mathbf{1 - p_k \geq 0}\}$$

$$\leq 0 \; \{\because \mathbf{log(p_k) \leq 0 \; \forall \; 0 \leq p_k \leq 1}\}$$

Hence it is proved that **Gini Index** $\leq$ **Cross-Entropy**

3(c)

Considering the first example we can draw the following decision tree For case (a): there is one split point for x denoted as s_x and one split point for y denoted as s_y For x-coordinate values greater than s_x the data falls into Y=t. so This is one level of decision tree. We check when x-coordinate of data falls lesser thans_x. On the next level for y-coordinate values greater than s_y the data falls into Y=t and for y coordinate values lesser s_y the data falls into Y=c. Thus a Decision tree of depth 2 is needed. For case (b): there is one split point for x denoted as s_x and one split point for y denoted as s_y For x-coordinate values of data greater than s_x then we check for y-coordinate values. If y-coordinate value is lesser than s_y then Y=t, and if y-coordinate value is greater than s_y then Y=c. Similarly, for x-coordinate values lesser than s_x then we check for y-coordinate values. If y-coordinate value is lesser than s_y then Y=c, and if y-coordinate value is greater than s_y then Y=t. Thus a Decision tree of depth 2 is needed. For case (c): The decision boundary is a line which cannot be determined by an out the split points. Moreover, a function describes the decision boundary and there is no definite decision tree which can determine it, especially not of depth lesser than 6. For case(d): There are two split points for x denoted by s_{x_1} and s_{x_2} (s_{x_1} ¡ s_{x_2}). Similarly, There are two split points for y denoted by s_{y_1} and s_{y_2} (s_{y_1} ¡ s_{y_2}) For x-coordinate values of data lesser than s_{x_1}, the data falls into Y=t. So this is the first level. For x coordinate values of data greater than s_{x_1}, we check it with s_{x_2}. If x-coordinate values of data greater than s_{x_2}, the data falls into Y=t. This is the second level of the decision tree. If x-coordinate values of data lesser than s_{x_2}, then we check the y-coordinate of the data. If the y-coordinate is lesser than s_{y_1}, then the data falls into Y=t. This is the third level of the decision tree. Now we check if y-coordinate is greater than s_{y_1}, then we check if the y-coordinate is greater than s_{y_2}. If so then it falls into Y=t and If the y-coordinate is lesser than s_{y_2}, then data falls into Y=c. This is the last level of the decision tree. Thus a decision tree of depth 4 is needed. Hence, (a), (b) and (d) can be modeled with a decision tree of depth lesser than 6.

Problem 4

4(a)

Given,

$$p_k = P(Y = k)\forall k \in 1, 2, \ldots, K \tag{25}$$

$$P(x_j|Y = y_k, \mu_{jk}, \sigma_{jk}) = \mathcal{N}(\mu_{jk}, \sigma_{jk}) = \frac{1}{\sqrt{2\pi\sigma_{jk}}} e^{-\frac{(x_j-\mu_{jk})^2}{2\sigma_{jk}}} \tag{26}$$

Lets calculate MLE of parameters using (19) and (20),
p_k:

μ_{jk}:

$$L = P(x_j|y_k, \mu_{jk}, \sigma_{jk}) = \prod_{j=1}^{D} \left(\frac{1}{\sqrt{2\pi\sigma_{jk}}} \right) e^{-\frac{(x_j-\mu_{jk})^2}{2\sigma_{jk}}}$$

$$ln(L) = ln \left[\prod_{j=1}^{D} \left(\frac{1}{\sqrt{2\pi\sigma_{jk}}} \right) e^{-\frac{(x_j-\mu_{jk})^2}{2\sigma_{jk}}} \right]$$

$$= -\frac{1}{2} \sum_{j=1}^{D} ln(2\pi\sigma_{jk}) - \sum_{j=1}^{D} \frac{(x_j-\mu_{jk})^2}{2\sigma_{jk}}$$

$$\frac{\partial ln(L)}{\partial \mu_{jk}} = \sum_{j=1}^{D} \frac{x_j - \mu_{jk}}{\sigma_{jk}} = 0$$

$$\Leftrightarrow \hat{\mu}_{\mathbf{jk}} = \mathbf{\frac{1}{D} \sum_{j=1}^{D} x_j}$$

σ_{jk}:

$$\frac{\partial ln(L)}{\partial \sigma_{jk}} = \frac{-D}{2\sigma_{jk}} + \sum_{j=1}^{D} \frac{(x_j-\mu_{jk})^2}{2\sigma_{jk}^2} = 0$$

$$\Leftrightarrow \hat{\sigma}_{\mathbf{jk}} = \mathbf{\frac{1}{D} \sum_{j=1}^{D} (x_j - \mu_{jk})^2}$$

4(b)

Given,

$$P(X_j|Y = y_k) = \theta_{jk}^{X_j} (1-\theta_{jk})^{1-X_j} \tag{27}$$

$$P(Y = 1) = \pi \tag{28}$$

From Naive Bayes theorem,

$$P(Y = 1|X) = \frac{P(X|Y = 1)P(Y = 1)}{\sum_{i=1}^{K} P(X|Y = y_i)P(Y = y_i)}$$

$$= \frac{P(X|Y = 1)P(Y = 1)}{P(X|Y = 1)P(Y = 1) + P(X|Y = 0)P(Y = 0)}$$

$$= \frac{1}{1 + \frac{P(X|Y=0)P(Y=0)}{P(X|Y=1)P(Y=1)}}$$

$$\Leftrightarrow P(Y=1|X) = \frac{1}{1+\frac{P(X|Y=0)P(Y=0)}{P(X|Y=1)P(Y=1)}} = \frac{1}{1+M} \tag{29}$$

From (23),

$$M = \frac{P(X|Y=0)P(Y=0)}{P(X|Y=1)P(Y=1)}$$

Applying log on both sides, we get:

$$ln(M) = ln(P(X|Y=0)P(Y=0)) - ln(P(X|Y=1)P(Y=1))$$

$$\Leftrightarrow ln(M) = ln\left(\prod_{j=1}^{D} \theta_{j0}^{X_j}(1-\theta_{j0})^{1-X_j}(1-\pi)\right) - ln\left(\prod_{j=1}^{D} \theta_{j1}^{X_j}(1-\theta_{j1})^{1-X_j} \times \pi\right)$$

$$= \sum_{j=1}^{D}[ln(1-\pi) - ln(\pi) + X_j ln(\theta_{j0}) + (1-X_j)ln(1-\theta_{j0}) - X_j ln(\theta_{j1}) - (1-X_j)ln(1-\theta_{j1})]$$

$$= -\sum_{j=1}^{D}\left[ln\left(\frac{(1-\pi)(1-\theta_{j1})}{\pi(1-\theta_{j0})}\right)\right] + \sum_{j=1}^{D} X_j\left[ln\left(\frac{\theta_{j0}(1-\theta_{j1})}{\theta_{j1}(1-\theta_{j0})}\right)\right]$$

$$\Leftrightarrow M = \exp^{-\sum_{j=1}^{D}\left[ln\left(\frac{(1-\pi)(1-\theta_{j1})}{\pi(1-\theta_{j0})}\right)\right] + \sum_{j=1}^{D} X_j\left[ln\left(\frac{\theta_{j0}(1-\theta_{j1})}{\theta_{j1}(1-\theta_{j0})}\right)\right]} \tag{30}$$

From (23) and (24) above,

$$P(Y=1|X) = \frac{1}{1+\exp^{-\sum_{j=1}^{D}\left[ln\left(\frac{(1-\pi)(1-\theta_{j1})}{\pi(1-\theta_{j0})}\right)\right] + \sum_{j=1}^{D} X_j\left[ln\left(\frac{\theta_{j0}(1-\theta_{j1})}{\theta_{j1}(1-\theta_{j0})}\right)\right]}} \tag{31}$$

$$\tag{32}$$

Where,

$$\omega_{\mathbf{o}} = \sum_{\mathbf{j=1}}^{\mathbf{D}}\left[\mathbf{ln}\left(\frac{(\mathbf{1}-\pi)(\mathbf{1}-\theta_{\mathbf{j1}})}{\pi(\mathbf{1}-\theta_{\mathbf{j0}})}\right)\right] \tag{33}$$

$$\mathbf{w} = \sum_{\mathbf{j=1}}^{\mathbf{D}}\left[\mathbf{ln}\left(\frac{\theta_{\mathbf{j0}}(\mathbf{1}-\theta_{\mathbf{j1}})}{\theta_{\mathbf{j1}}(\mathbf{1}-\theta_{\mathbf{j0}})}\right)\right] \tag{34}$$

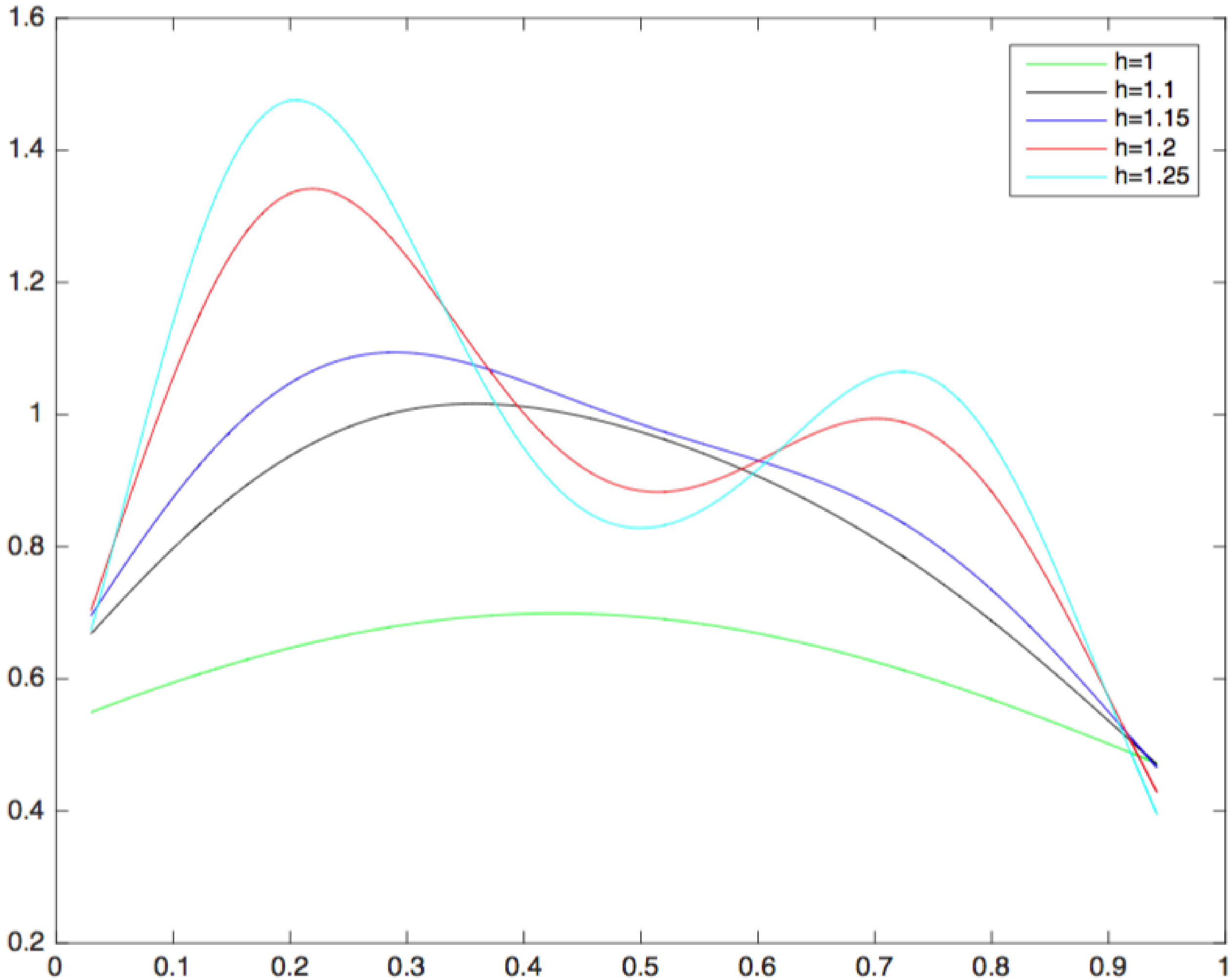

Figure 2: Gaussian Kernel

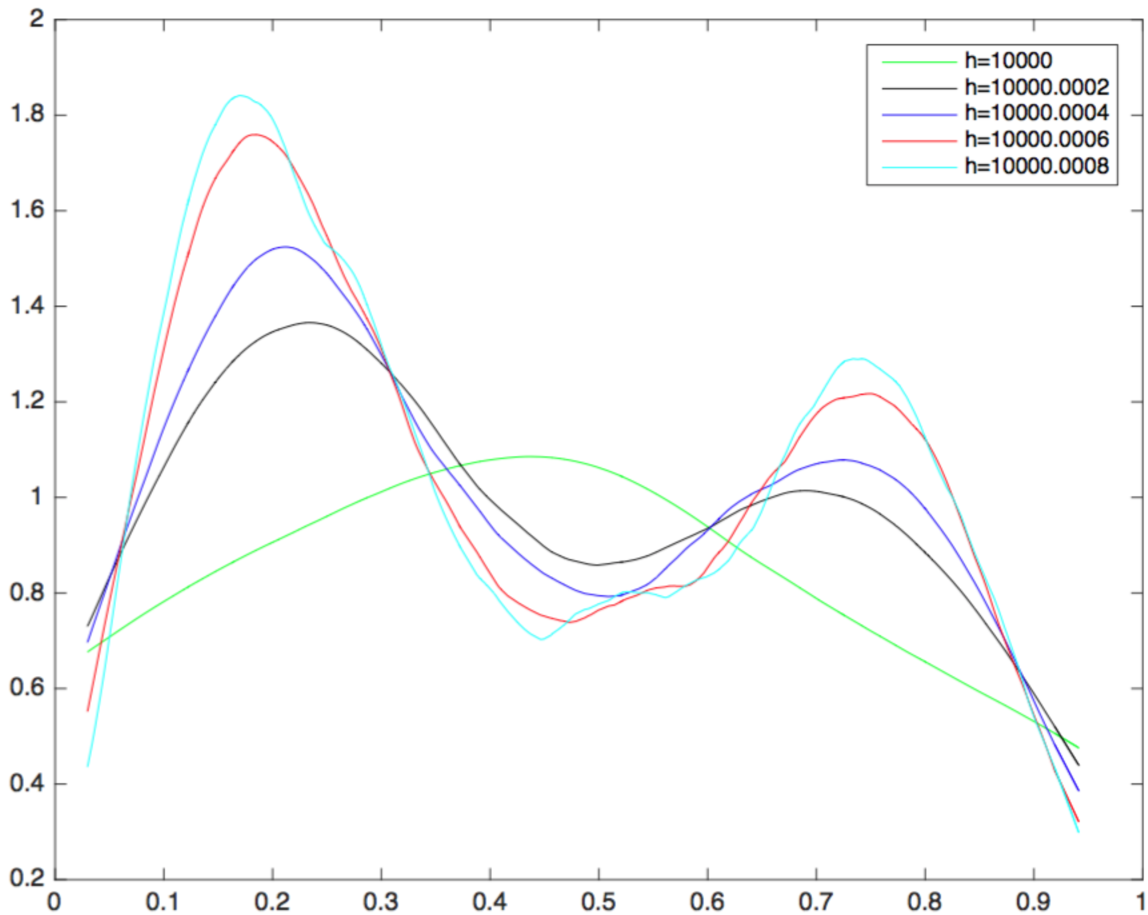

Figure 3: Epanechnikov Kernel

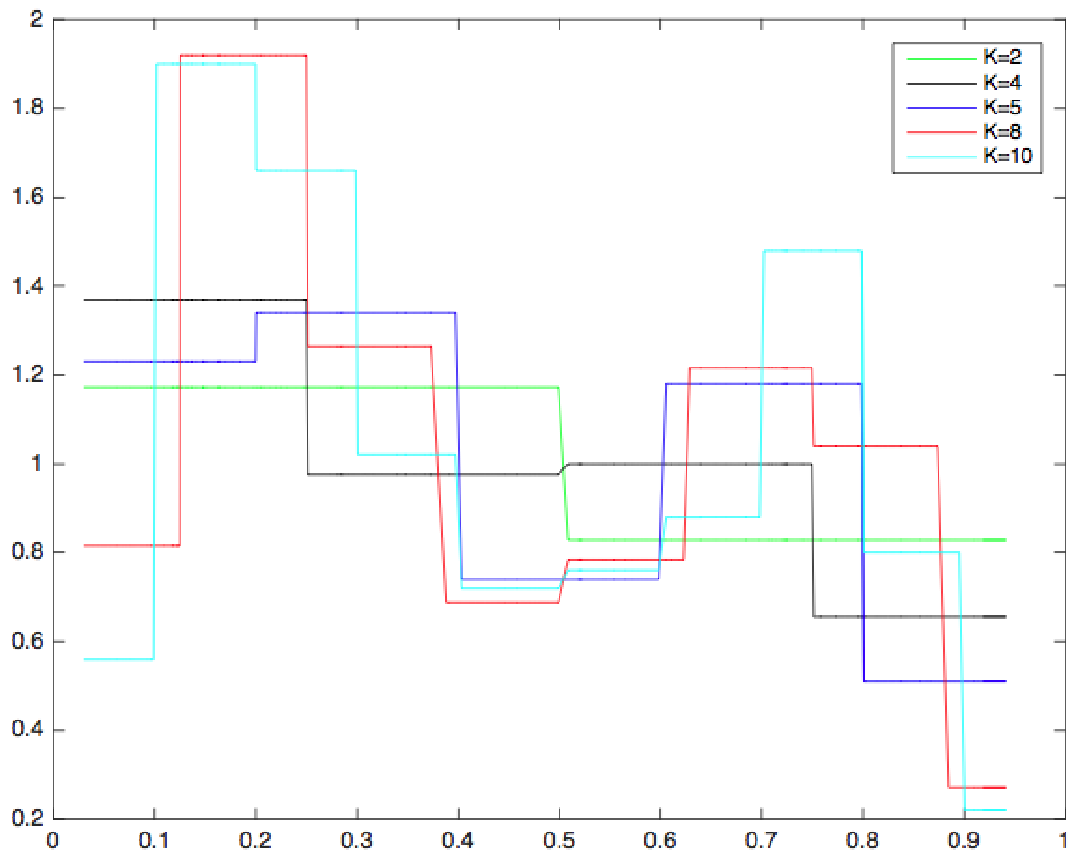

Figure 4: Histogram

Problem 5

5.1

(a)

5.2

(d)

K-NN Accuracies

K	Training data	Validation data	Test data
1	0.69	0.69	0.65
3	0.72	0.71	0.69
5	0.80	0.80	0.73
7	0.84	0.84	0.82
9	0.84	0.85	0.85
11	0.79	0.79	0.74
13	0.71	0.72	0.67
15	0.68	0.69	0.65

Decision Tree Accuracies

i	Gini Index	CrossEntropy
1	0.95, 0.86	0.95, 0.87
2	0.95, 0.86	0.95, 0.87
3	0.95, 0.86	0.95, 0.87
4	0.95, 0.88	0.95, 0.88
5	0.94, 0.87	0.94, 0.87
6	0.93, 0.88	0.93, 0.88
7	0.92, 0.89	0.92, 0.89
8	0.91, 0.89	0.92, 0.90
9	0.88, 0.88	0.92, 0.86
10	0.91, 0.85	0.88, 0.87

Naive Bayes

	Tic tac toe	Nursery
Training	0.66	
Validation	0.67	
Test	0.63	