CS534 Written Assignment 2

Aashish Adhikari

Q.1) Solution:

The question is analogous to putting n balls in k bins and thus the distribution is multinomial.

pmf for the parameters can be taken from multinomial distribution.

$$\therefore \text{Likelihood}(p_A, p_C, p_G, p_T) = p_A^{n_A} \times p_C^{n_C} \times p_G^{n_G} \times p_T^{n_T}$$

We have the constraint $p_A + p_C + p_G + p_T = 1$

In such cases where we have constraints, for optimization, we have Lagrange Multiplier.

We need to find maximum of Likelihood (p_A, p_C, p_G, p_T) subject to $p_A + p_C + p_G + p_T = 1$.

$$\Rightarrow \therefore F(p_A, p_C, p_G, p_T) = \mathcal{L}(p_A, p_C, p_G, p_T) - \lambda\left(p_A + p_C + p_G + p_T - 1\right)$$

$$\text{or, } F(p_A, p_C, p_G, p_T) = \left[p_A^{n_A} \times p_C^{n_C} \times p_G^{n_G} \times p_T^{n_T}\right] - \lambda p_A - \lambda p_C - \lambda p_G - \lambda p_T + \lambda$$

Now, we take partial derivatives wrt p_A, p_C, p_G, p_T & λ and set them to zero to maximize.

$\therefore \frac{\partial(F(\ ))}{\partial p_A}$ = To make computation easier, lets take log of the likelihood.

$$F'() = \log_e(\mathcal{L}) = \log_e\left(p_A^{n_A}\right) + \log_e p_C^{n_C} + \log_e p_G^{n_G} \times \log_e p_T^{n_T}$$

Lets represent a by 1, c by 2, g by 3, t by 4.

$$\Rightarrow \log_e(\mathcal{L}) = \sum_{i=1}^{4} \log_e p_i^{n_i}$$

New Link
Date
Page

$$= \sum_{i=1}^{4} n_i \log p_i$$

Also, $\lambda(p_a + p_c + p_g + p_t - 1)$ becomes $\lambda\left[\left[\sum_{i=1}^{4} p_i\right] - 1\right] = \lambda G$

$$\therefore \frac{\partial[F() - \lambda G]}{\partial p_{ai}} = \frac{n_i \times p_i^{n_i - 1}}{(p_i^{n_i})} - \lambda\left[1\right]$$

Setting it to 0;

$$n_i \times p_i^{n_i - 1 - n_i} = \lambda$$

or, $\frac{n_i}{p_i} = \lambda$

$$\Rightarrow p_i = \frac{n_i}{\lambda}$$

Also, $\frac{\partial}{\partial \lambda}[F() - \lambda G) = 0 - [4\lambda + 1]$

Setting to 0 $\Rightarrow 4\lambda + 1 = 0$

$\Rightarrow \lambda = -1/4$

Also, $\frac{\partial}{\partial \lambda}[F() - \lambda G] = 0 - p_a - p_c - p_g - p_t + 1$

Setting to 0 $\Rightarrow 0 = \frac{n_a}{\lambda} + \frac{n_g}{\lambda} + \frac{n_c}{\lambda} + \frac{n_t}{\lambda} - 1$

or, $+\lambda = n_a + n_g + n_c + n_t = n_{total}$

or, $\lambda = + n_{total}$

$\therefore p_i = \frac{n_i}{n_{total}}$

maximum likelihood estimates are:

$\Rightarrow p_a = \frac{n_a}{(n_a + n_c + n_g + n_t)}$

$p_c = \frac{n_c}{n_a + n_c + n_g + n_t}$

$p_g = \frac{n_g}{n_a + n_c + n_g + n_t}$

$p_t = \frac{n_t}{n_a + n_c + n_g + n_t}$.

Q.2. Solution

Naive Bayes Classifier uses

$$P(A|B) = \frac{P(B|A) \times P(B)}{P(A)}$$

$$= \frac{P(B|A) \times P(B)}{\sum_B P(A|B) P(B)}$$

Here, we have 3 features A, B, C & output Y.

a) $\therefore P(Y) = 1/2$ for $Y = 1 \Rightarrow P(Y=0) = 1/2$ & $P(Y=1) = 1/2$

$P(Y) = 1/2$ for $Y = 0$

$P(A=0|y=0) = 2/3$
$P(B=0|y=0) = 1/3$
$P(C=0|y=0) = 1/3$
} for $y=0$

$P(A=0|y=1) = 1/3$
$P(B=0|y=1) = 1/3$
$P(C=0|y=1) = 2/3$
} for $y=1$

Since the assumption is that the features are independent, probabilities like $P(A=0, B=0, C=1|y=0)$ can be calculated by multiplying individual probabilities such as $P(A=1|y=0)$, $P(B=0|y=0)$, and $P(C=1|y=0)$.
Hence, above 8 probabilities are all that we need to estimate.

b. Solution

$$P(y=1|A=1, B=0, C=0)$$

$$= \frac{P(A=1, B=0, C=0)\, P(y=1)}{P(A=1, B=0, C=0)}$$

Since they are independent, multiply

$$= \frac{2/3 \times 1/3 \times 2/3 \times 1/2}{2/3 \times 1/3 \times 2/3 \times 1/2 + 1/3 \times 1/3 \times 1/3 \times 1/2}$$

$$=$$

$$= \frac{2/3 \times 1/3 \times 2/3 \times 1/2}{2/3 \times 1/3 \times 2/3 \times 1/2 + 1/3 \times 1/3 \times 1/3 \times 1/2}$$

$$= \frac{4 \times 54}{54 \times 5}$$

$$= 4/5$$

c. Solution:

No, it is not valid.

It is because Naive-Bayes classifier assumes conditional independence. However, the question has mentioned that A, B, C are independent and independence does not guaranteed conditional independence.

For example, consider picking a card from a deck of 52 cards and rolling a die.

A is an event of choosing an Ace

S is an event of rolling a 6.

probability of (A|S) = probability of A ← Independent

$$\Rightarrow P(A|S) = \frac{P(S|A)\ P(S)}{P(A)}$$

$$\text{or, } [P(A)]^2 = [P(S)]^2$$

$$\text{or, } \left(\frac{4}{52}\right)^2 = \left(\frac{1}{6}\right)^2$$ which is not true. ✗

Hence, invalid.

Q3 Solution:

p is the probability of head for a weighted coin

It has a Beta distribution

$p \sim \text{Beta}(2,2)$

$$\text{pdf}(p; 2,2) = \frac{1}{B(\alpha=2, \beta=2)} p^{2-1}(1-p)^{2-1}$$

$$= \frac{1}{B(2,2)} p \times (1-p)$$

Posterior distribution of p is given by

$$\frac{1}{B(n_1+2, n_0+2)} p^{n_1+1} \times (1-p)^{n_0+1}$$

where n_1 is the no of heads

n_0 is the no of tails

Case 1: $n_1 + n_0 = 5$

$n_1 = 2$.

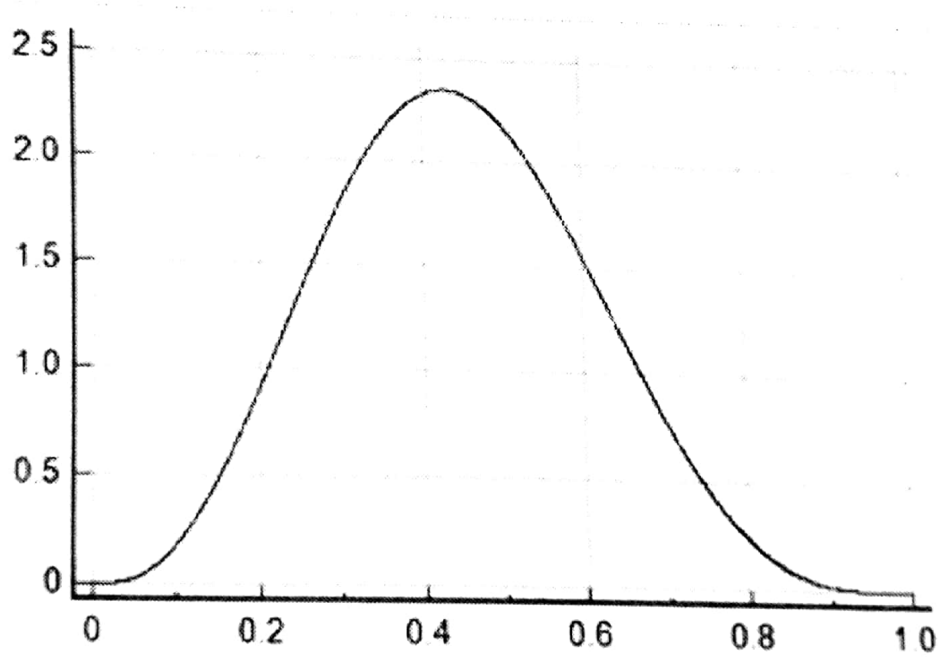
$$\Rightarrow \frac{1}{B(4,5)} p^3 \times (1-p)^4$$

Case 2: $n_1 + n_0 = 50$

$n_1 = 20$

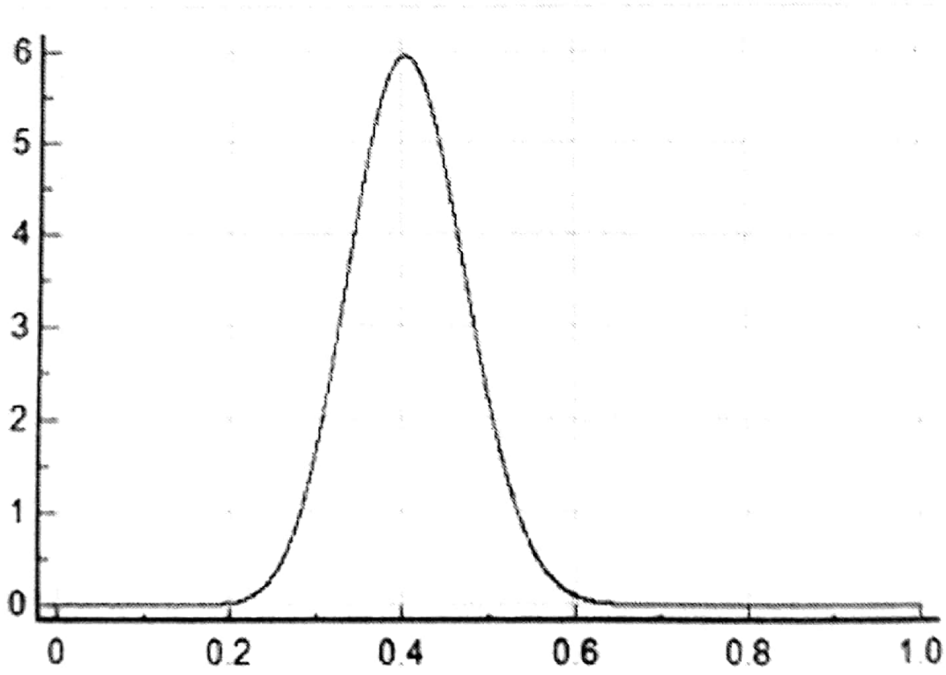
$$\Rightarrow \frac{1}{B(22,32)} p^{21}(1-p)^{31}$$

Graph



Function: PDFBeta(x .4 .5)

Graph

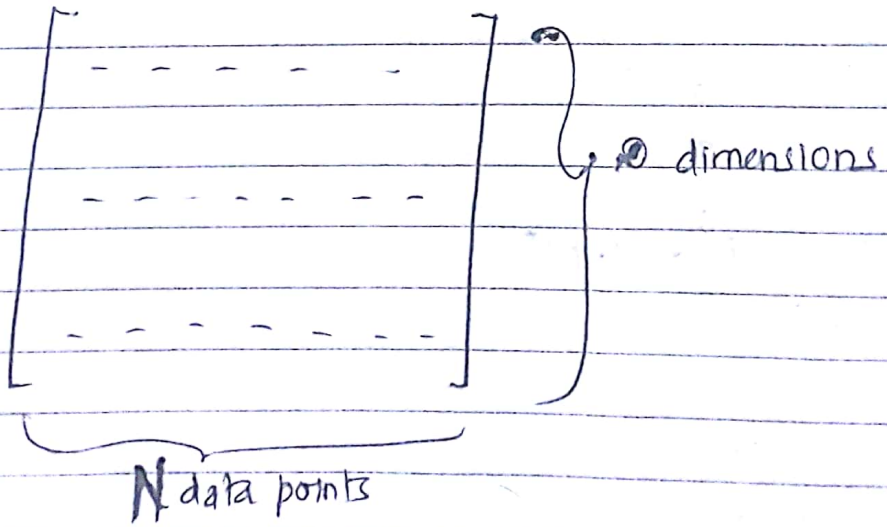


Function: PDFBeta(x ,22 ,32)

We can see that as the number of coin tosses from this coin increases, the probability gets nearer and nearer to 0.4. Thus as the number of coin tosses tends to a large number, the probability of getting a head ≈ 0.4 according to the distribution plot i.e., it reaches towards the real value.

Q.4 Solution

a) We have N data points in D dimensions.



We map each point $\vec{x_n}$ to $(D+N)$ dimensional point $\langle \vec{x_n}, e_n \rangle$ where e_n is a N-dimensional vector of all zeros but one 1 at the n^{th} position.

After adding the dimensions

New Link
Date
Page

individual points

$$\begin{bmatrix} 1 & x_1 & x_2 & x_3 & \cdots & x_d & 1 & 0 & 0 & \cdots & 0 \\ 1 & x_1 & x_2 & x_3 & \cdots & x_d & 0 & 1 & 0 & \cdots & 0 \\ 1 & & & & & & & & & & \\ 1 & & & & & & & & & & \\ \vdots & & & & & & & & & & \\ 1 & x_1 & x_2 & x_3 & \cdots & x_d & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$\longleftarrow$ d+n+1 dimensions $\longrightarrow$

The perceptron loss is given as $\mathcal{L}_p(g(w,x),y) = \max\begin{pmatrix} 0 \\ -yw^Tx \end{pmatrix}$

So, when we increase the dimension by n, the new dimension is (d+n).

Let points $X_0, X_1, \ldots, X_n$ are in a (d+n) dimensional space as shown in the matrix above. These points, suppose come under 2 different sets i.e., are linearly separable into 2 sets A & B. for this, there must exist (d+n+1) real numbers $w_0, w_1, \ldots w_n, w_k$ such that each point X_i in A satisfies $\sum_{i=0}^{n} w_i x_i > k$ and

each point X_i in B satisfies $\sum_{i=0}^{n} w_i x_i < k$.

Consider we are adding 3 new dimensions.

$$X \begin{bmatrix} \cdots & 1 & 0 & 0 \\ \cdots & 0 & 1 & 0 \\ \cdots & 0 & 0 & 1 \end{bmatrix}$$

The new weights added are w_{e1}, w_{e2}, w_{e3}

We need to ~~find the~~ make sure that the $\sum w_i x_i$ is greater or less than k.

Thus, setting learning values of new weights such that each of the examples lies on one side of the linear classifier guarantees the solution.

~~Notice that for each example,~~

If we set all values upto dimension (1 to d+1) as 1 and a large positive or negative value for every other position in our new vector of weights, then the result is either we get $g(x,w)$ is +ve or -ve.

For example

$$\begin{bmatrix} w_0 & w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ 1 \\ 0 \end{bmatrix}$$

already existing input

added according to the question

$w_3 \times 0 = 0$

only w_2 is taken to be significant among the added dimensions.

So, setting w_2 to a large negative value can allow us to classify as a negative class.

Hence such a projection forces the data to be linearly separable.

New Link
Date
Page

b) This mapping does not allow us to generalize because
Solution: for each different length (dimension) of e_n, we need to address the weight uniquely. However, forcing the data to be linearly separable is possible independent of the size (dimension) of e_n.

Q. B. Solution

Given $K(x_i, x_j) = (x_i \cdot x_j + 1)^3$

Let $x_i = a$ & $x_j = b$.

$\therefore (a \cdot b + 1)^3$

$= a^3 + b^3 + 3a^2b + 3ab^2$

Assume we have p features for each training example

$$\sum_{x=1}^{p} (a_x)^3 + \sum_{x=1}^{p} (b_x)^3 + 3\sum_{x=1}^{p} (a_x)^2 b_x + 3\sum_{x=1}^{p} a_x (b_x)^2$$

$$= \sum_{x=1}^{p}\sum_{y=1}^{p}\sum_{z=1}^{p} a_x a_y a_z + \sum_{x=1}^{p}\sum_{y=1}^{p}\sum_{z=1}^{p} b_x b_y b_z$$

$$+ 3\sum_{x=1}^{p}\sum_{y=1}^{p} a_x a_y b_x + 3\sum_{x=1}^{p}\sum_{y=1}^{p} a_x b_x b_y$$

$$= \sum_{x=1}^{p}\sum_{y=1}^{p}\sum_{z=1}^{p} a_x a_y a_z + \sum_{x=1}^{p}\sum_{y=1}^{p}\sum_{z=1}^{p} b_x b_y b_z +$$

$$3\sum_{x=1}^{p}\sum_{y=1}^{p} a_x^2 b_x + 3\sum_{x=1}^{p}\sum_{y=x+1}^{p} a_x a_y b_x +$$

$$3\sum_{x=1}^{p}\sum_{y=1}^{p} a_x b_x^2 + 3\sum_{x=1}^{p}\sum_{y=x+1}^{p} a_x b_x b_y$$

∴ the ~~kernel~~ corresponding ϕ functions are:

$\{ a_x^3, b_x^3, \sqrt{3}\, a_x^2, \sqrt{3}\, b_x, \sqrt{3}\, a_x a_y, \sqrt{3}\, b_x^2, \sqrt{3}\, a_x, \sqrt{3}\, b_x b_y \}$

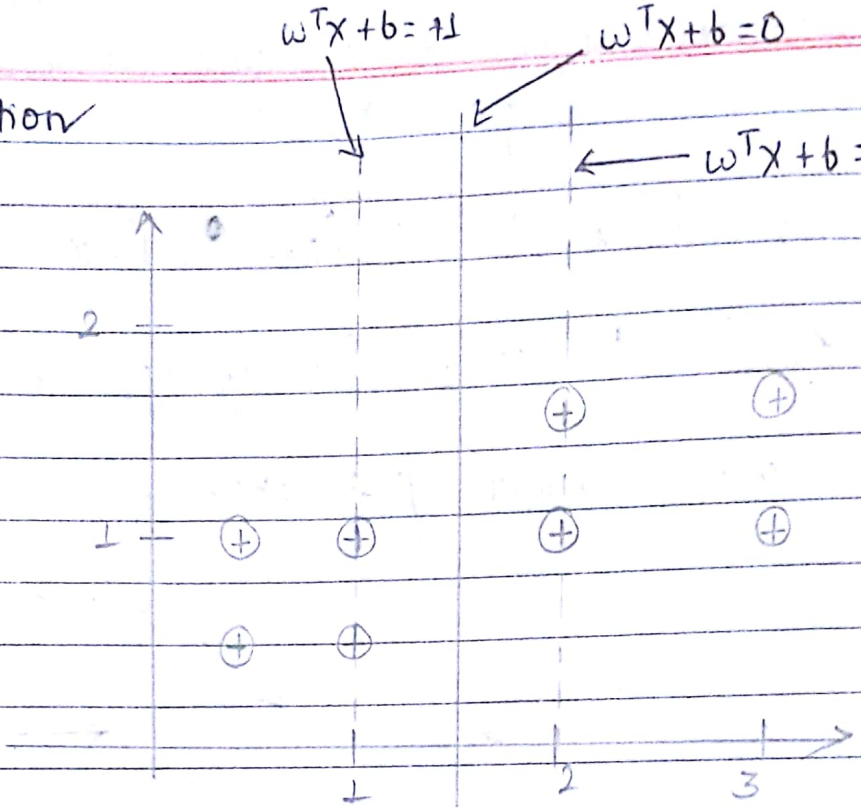
New Link
Date
Page

Changing indices x,y,z to $1,2,3$ and $a = x_i$, $b = x_j$
Converting in terms of x_i & x_j.

$$\left\{ x_{i1}^3,\ x_{j1}^3,\ \sqrt{3}(x_{i1})^2,\ \sqrt{3}(x_{j1})^2,\ \sqrt{3}\,x_{i1}x_{j2},\ \sqrt{3}\,(x_{j1})^2,\ \sqrt{3}\,x_{i1},\ \sqrt{3}\,x_{j1}x_{j2} \right\}$$

which is our required ϕ function.

Q 6 (a) Solution



$w^T x + b = 0$ is the decision boundary that lies in the middle of 1 & 2 on X axis.

$\Rightarrow$ point (1.5, 0) lies on the decision boundary.

$\Rightarrow$ points (1, 0.5) & (1, 1) lie on $w^T x + b = +1$

$\Rightarrow$ points (2, 1) & (2, 1.5) lie on $w^T x + b = -1$

$\therefore w^T x + b = 0$

we have the point (1.5, 0) on this boundary

(b) Taking weight vector $w = \begin{pmatrix} w \\ 0 \end{pmatrix}$ & $w^T = \begin{bmatrix} w & 0 \end{bmatrix}$

$$\begin{bmatrix} w & 0 \end{bmatrix} \begin{bmatrix} 1.5 \\ 0 \end{bmatrix} + b = 0$$

Gives $b = -1.5w$.

For $w^T x + b = -1$, there are 2 points (2,1) & (2,1.5)
For $w^T x + b = 1$, we have (1, 0.5] & (1,1).

$$\begin{bmatrix} w & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} + b = 1 \qquad \begin{bmatrix} w & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + b = 1$$

$w + b = 1$ $\qquad$ $w + b = 1$

(I)

put value for $w^T x + b = 1$

$$\begin{bmatrix} w & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} + b = -1 \qquad \begin{bmatrix} w & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1.5 \end{bmatrix} + b = -1$$

$2w + b = -1$ $\qquad$ $2w + b = -1$

(II)

Solve (I) & (II)

$$2w + b = -1$$
$$- w + b = -1$$
$$\overline{w = -2}$$

put value in (II) $\Rightarrow$ $b = -1 - 2\times(-2)$
$= 3.$

$\therefore b = 3, w = -2$ #