CS 7650 : Problem set 1

1.)

P(A,B)

A \ B	0	1
0	0.2	0.5
1	0.1	0.2

(a.) $p(A=0) = p(A=0, B=0) + p(A=0, B=1)$

$= 0.2 + 0.5$

$= 0.7$

(b.) $p(B=0 \mid A=1) = \dfrac{p(B=0, A=1)}{p(A=1)}$

$= \dfrac{0.1}{0.3}$

$= 0.\overline{33}$

(c.) $p(A=B) = p(A=0, B=0) + p(A=1, B=1)$

$= 0.2 + 0.2$

$= 0.4$

2.) Assuming that X is conditionally independent of Y given Z, statements (a) and (c) are always true.

a.) $P(X,Y) = \sum_{c \in \mathcal{X}_z} p(X, Y, Z=c) \longrightarrow$ TRUE

b.) $P(X,Y,Z) = P(X) + P(Y) + P(Z=c),\ c \in \mathcal{X}_z \longrightarrow$ FALSE

c.) $P(X, Y \mid Z) = P(X \mid Z) \cdot P(Y \mid Z) \longrightarrow$ TRUE

because $P(X, Y \mid Z) = P(X \mid Y, Z) \cdot P(Y \mid Z)$

$= P(X \mid Z) \cdot P(Y \mid Z)$ from the conditional independence

d.) $P(X,Y,Z) = P(X) + P(Y) - P(Z) \longrightarrow$ FALSE

e.) $P(X, Y) = P(X) \cdot P(Y) \longrightarrow$ FALSE

since $P(X,Y) = P(X \mid Y) \cdot P(Y)$

$\hookrightarrow$ not equal to $P(X)$

3) $p(\text{thunderstorm}) = 0.1$

$p(\text{barking} \mid \text{thunderstorm}) = 0.75$

$p(\text{barking} \mid \text{no thunderstorm}) = 0.25$

From Bayes Rule : $p(B|A) = \dfrac{p(A|B) \cdot p(B)}{p(A)}$

~~or~~ $\cancel{p(B|A) = \dfrac{p(A,B)}{p(A)}}$ $= \dfrac{p(A|B) \cdot p(B)}{\sum_B p(A,B)}$

For convenience, I will be representing 'A' as the event of the dog barking, 'A^C' as the event of the dog not barking, 'B' as the event of the thunderstorm approaching and 'B^C' as the event of no thunderstorm approach.

$$\therefore p(B|A) = \frac{p(A|B) \cdot p(B)}{p(A|B) \cdot p(B) + p(A|B^C) \cdot p(B^C)}$$

$$= \frac{0.75 \times 0.1}{0.75 \times 0.1 + 0.25 \times 0.9}$$

$$= \frac{0.075}{0.075 + 0.225} = \frac{0.075}{0.3}$$

$$= 0.25$$

The probability that a thunderstorm is approaching, given that the dog is currently barking.

4)

		y	
		0	1
x	0	0	1/5
	1	2/5	2/5

$p(x=0) = 1/5 \quad p(x=1) = 4/5$

$p(y=0) = 2/5 \quad p(y=1) = 3/5$

(a.) $H(X) = -\sum_x p(x) \log_2 p(x)$

$= -p(x=0) \cdot \log_2 p(x=0) - p(x=1) \cdot \log_2 p(x=1)$

$= -\frac{1}{5} \cdot \log_2 \frac{1}{5} - \frac{4}{5} \log_2 \frac{4}{5}$

$= 0.721928$

(b.) $H(Y) = -\sum_y p(y) \cdot \log_2 p(y)$

$= -p(y=0) \cdot \log_2 p(y=0) - p(y=1) \cdot \log_2 p(y=1)$

$= -\frac{2}{5} \cdot \log_2 \frac{2}{5} - \frac{3}{5} \cdot \log_2 \frac{3}{5}$

$= 0.97095$

(c.) $H(X, Y) = -\sum_{x,y} p(x,y) \cdot \log_2 p(x,y)$

$= -p(x=0, y=0) \cdot \log_2 p(x=0, y=0)$
$- p(x=0, y=1) \cdot \log_2 p(x=0, y=1)$
$- p(x=1, y=0) \cdot \log_2 p(x=1, y=0)$
$- p(x=1, y=1) \cdot \log_2 p(x=1, y=1)$

$= -0 \log_2 0 - 0.2 \log_2 0.2 -$
$0.4 \log_2 (0.4) - 0.4 \log_2 0.4$

$= 0 + 0.4643 + 2(0.528)$

$= 1.5203$

5.)

(a.) PDF: $f(x) = \begin{cases} ce^{-x} & , x > 0 \\ 0 & , \text{otherwise} \end{cases}$

(i.) $f(x)$ is considered a valid PDF if the area under the curve is 1

$$\therefore \int_0^\infty ce^{-x}dx + \int_{-\infty}^0 0\,dx = 1$$

$$c \cdot \int_0^\infty e^{-x}dx = 1$$

$$\Rightarrow c \cdot \left[-e^{-x}\right]_0^\infty = 1$$

$$\Rightarrow c \cdot [-0 + 1] = 1$$

$$\Rightarrow c = 1$$

(ii.) $E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$

$u \to x$

$v \to e^{-x}$

$$= \int_0^\infty x \cdot e^{-x}\,dx$$

using integration by parts.

$$\int uv = u\int v - \int u' \int v$$

$$E[X] = \left[x \cdot (-e^{-x}) - \int (1) \cdot -e^{-x}dx\right]_0^\infty$$

$$= \left[-xe^{-x} + \int e^{-x}dx\right]_0^\infty$$

$$= \left[-xe^{-x} - e^{-x}\right]_0^\infty$$

$$= \lim_{x \to \infty} (-xe^{-x} - e^{-x}) - (0 - 1)$$

$$= 0 - 0 - 0 + 1$$

$$= 1$$

$\boxed{2}\ \boxed{1}\ \boxed{1} = 2$

(b.) 3 locks, 3 keys

$$p(\geq 1 \text{ match}) = 1 - p(0 \text{ matches})$$

$$= 1 - \frac{2}{3!}$$

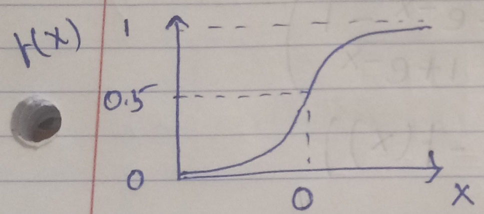
no. of outcomes where no matches result

total number of ways of assigning a key to a lock

$$= 1 - \frac{2}{6}$$

$$= \frac{4}{6}$$

6.) $f(x) = \dfrac{1}{1+e^{-x}}$ (logistic / sigmoid)



Plot (c) matches this equation

- The minimum value of the sigmoid function is 0 (when $x \to -\infty$). On the other hand, the maximum value is 1 (when $x \to \infty$). We can interpret the output of a sigmoid function as a probabilistic value (since it lies between 0 and 1), and we can use it in the final layer of a machine learning model (such as logistic regression) to ~~output a probabilistic value~~. output the predicted probability of an instance belonging to the positive class.

$\cdot \frac{\partial}{\partial x} f(x) = \frac{\partial}{\partial x}\left(\frac{1}{1+e^{-x}}\right)$ $\frac{\partial}{\partial x}(u/v) = \frac{vu' - uv'}{v^2}$

$= \frac{1+e^{-x}(0) - 1(-e^{-x})}{(1+e^{-x})^2}$

$= \frac{e^{-x}}{(1+e^{-x})^2}$ $\cancel{= \frac{1+e^{-x}-1}{(1+e^{-x})^2}}$

$= \frac{1}{(1+e^{-x})} \cdot \frac{e^{-x}}{(1+e^{-x})}$

$= f(x)\left(\frac{1+e^{-x}-1}{1+e^{-x}}\right)$

$= f(x) \cdot (1 - f(x))$

7.)

(a.) $f(t) = 100(60+5t)^{2/3}$

$\frac{\partial}{\partial t} f(t) = \frac{2}{3} \cdot 100 \, (60+5t)^{-1/3} \times \frac{\partial}{\partial t}(60+5t)$

$= \frac{200}{3} \times 5 \times \frac{1}{(60+5t)^{1/3}}$

$\left.\frac{\partial}{\partial t} f(t)\right|_{t=0} = \frac{1000}{3} \times \frac{1}{(60)^{1/3}}$

$= \frac{1000}{3 \times 3.9148}$

$= 85.145$

(b)

(i) $f(x) = C^T x$

$$\frac{\partial f(x)}{\partial x} = C^T + x \cdot \frac{\partial}{\partial x} C^T$$

if C is a function of x

$$\frac{\partial f(x)}{\partial x} = C^T$$

if C is not a function of f.

(ii) $g(x) = \frac{1}{2} x^T H x$

$$\frac{\partial g(x)}{\partial x} = \frac{1}{2} x^T (H + H^T)$$

assuming that H doesn't depend on x

If H is a symmetric matrix, then

$$\frac{\partial g(x)}{\partial x} = x^T H$$

(iii) $h(x) = \frac{1}{2} x^T H x + C^T x$

$$H = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$$\frac{\partial h(x)}{\partial x} = 0 \Rightarrow \frac{1}{2} x^T (H + H^T) + C^T = 0$$

$$\Rightarrow \frac{1}{2} x^T \cdot \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix} + \begin{bmatrix} 1 & 4 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 4 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} 2x_1 + 1, & 4x_2 + 4 \end{bmatrix} = 0$$

$\therefore$ $x_1 = -1/2$ and $x_2 = -1$

$$x^T = \begin{bmatrix} -1/2 & -1 \end{bmatrix} \quad \text{or} \quad x = \begin{bmatrix} -1/2 \\ -1 \end{bmatrix}$$

Using the second derivative test, we can figure out if $x = [-1/2 \ -1]^T$ is a local maxima, local minima or a saddle point.

$$h(x) = \frac{1}{2}[x_1 \ x_2]\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [1 \ 4]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \frac{1}{2}[x_1 \ x_2]\begin{bmatrix} 2x_1 \\ 4x_2 \end{bmatrix} + x_1 + 4x_2$$

$$= x_1^2 + 2x_2^2 + x_1 + 4x_2$$

$\cancel{h'(x) = 2x_1 + 4x_2 +}$

$h_{x_1}(X) = 2x_1 + \cancel{2x_2^2} + 1 + \cancel{4x_2}$

$h_{x_2}(X) = \cancel{2x_1} \ 4x_2 + 4$

$h_{x_1^2 x_1}(X) = 2$

$h_{x_2^2 x_2}(X) = 4$

$h_{x_1 x_2}(X) = 0$

$$|H| = h_{x_1^2 x_1}(X) \cdot h_{x_2^2 x_2}(X) - h_{x_1 x_2}(X)^2$$

$$= 2 \times 4 - 0$$

$$= 8$$

Since the determinant of the Hessian is greater than 0 and the second derivative is greater than 0 at $[-1/2 \ -1]^T$, this point is a minimum point.