CSCI-567: Assignment #1

Due on Tuesday, September 23, 2015

Saket Choudhary
2170058637

Contents

Problem # 1

Problem # 1: (a) 1

Given: $X_i \sim Beta(\alpha, 1)$ MLE for α:
Consider $X = (X_1, X_2, \dots, X_n)$ Likelihood function: $L(\alpha|X)$ $L(\alpha|X) = \Pi_{i=1}^{n} f(x_i)$ where

$$f(x_i) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(1)} x^{\alpha-1} \tag{1}$$

$$= \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)} x^{\alpha-1} \qquad = \alpha x^{\alpha-1} \tag{2}$$

$$L(\alpha|X) = \left(\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(1)}\right)^n \Pi_{i=1}^{n} (x_i)^{\alpha-1} \tag{3}$$

$$LL = \log(L(\alpha|X)) = n\log(\alpha) + (\alpha-1)\sum_{i=1}^{n} x_i \tag{4}$$

$$\frac{dLL}{d\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log(x_i) \tag{5}$$

$$\frac{dLL}{d\alpha} = 0 \implies \hat{\alpha} = \frac{n}{\sum_{i=1}^{n} log(1/x_i)} \tag{6}$$

Minima at $\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} log(1/x_i)}$ is guaranteed due to log being a concave function.

Problem # 1: (a) 2

Given: $x_i \sim N(\theta, \theta)$ i.e $f(x_i) = (2\pi\theta)^{-\frac{1}{2}} e^{-\frac{(x_i-\theta)^2}{2\theta}}$ MLE estimate for θ:

$$L(\theta|X) = (2\pi\theta)^{-\frac{N}{2}} e^{-\sum_{i=1}^{n} \frac{(x_i-\theta)^2}{2\theta}} \tag{7}$$

$$LL = \log(L(\theta|X)) = -\frac{N}{2}\log((2\pi\theta)) - \sum_{i=1}^{n} \frac{(x_i-\theta)^2}{2\theta} \tag{8}$$

$$\frac{dLL}{d\theta} = -\frac{N}{2}\left(\frac{1}{\theta}\right) + \frac{\sum_{i=1}^{n} x_i^2}{2\theta^2} - \frac{N\theta}{2} \tag{9}$$

$$\frac{dLL}{d\theta} = 0 \implies N\theta^2 + N\theta - \sum_{i=1}^{n} x_i^2 = 0 \tag{10}$$

The above equation is a quadratic and will have two solutions, Since, $\theta \geq 0$ (a constraint that comes from θ being the variance), the
$\theta = \frac{-N \pm \sqrt{N^2 + 4N\sum_{i=1}^{n} x_i^2}}{2N}$
Since, $\hat{\theta} \geq 0$, $\hat{\theta} = \frac{-N + \sqrt{N^2 + 4N\sum_{i=1}^{n} x_i^2}}{2N}$

Problem # 1: (b) 1

Given: $f(\hat{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K(\frac{x-X_i}{h})$ To show: $E_{X_1,X_2,...X_n}[f(\hat{x})] = \frac{1}{h}\int K(\frac{x-t}{h})f(t)dt$
Proof:

$$E[f(\hat{x})] = E[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K(\frac{x-X_i}{h})] \tag{11}$$

$$= \frac{1}{nh}E[K(\frac{x-X_i}{h})] \tag{12}$$

$$= \frac{1}{h}E[K(\frac{x-X_1}{h})] \qquad = \frac{1}{h}E[K(\frac{x-t}{h})] \tag{13}$$

where the penultimate equality comes from the fact that X_i are iid for all $i \in [1,n]$. and t X and hence.

$$E[f(\hat{x})] = \frac{1}{h}E[K(\frac{x-X_1}{h})] \qquad = \frac{1}{h}\int K(\frac{x-t}{h})f(t)dt = RHS \tag{14}$$

Problem # 1: (b) 2

Consider $z = \frac{x-t}{h} \implies t = x - hu$
Then,

$$E[f(\hat{x})] = \frac{1}{h}\int K(z)f(x-hz)dz \tag{15}$$

$$\tag{16}$$

$f(x-hz) = f(x) - f'(x)hz + \frac{1}{2}f''(x)\frac{(hz)^2}{2} - \frac{1}{3}f'''(x)\frac{(hz)^3}{3!} + \cdots + (-1)^n\frac{1}{n!}f^{(n)}(x)(\frac{(hu)^n}{n!})$
By definition, $\int k(z)dz = 1$. Also define an auxilllary variable $M_j = \int k(z)z^j dz$ for the j^{th} moment of the kernel function, and hence, $\int K(z)f(x-hz)dz = f(x) - hf'(x)M_1 + \frac{1}{2}(h^2)f^{''}(x)M_2 + \cdots + (-1)^n\frac{1}{n!}f^{(n)}M_n$
Now, $Bias = E[f(\hat{x})] - f(x) = -hf'(x)M_1 + \frac{1}{2}(h^2)f^{''}(x)M_2 + \cdots + (-1)^n\frac{1}{n!}f^{(n)}M_n$
And as $h \longrightarrow 0$, $Bias \longrightarrow 0$

Problem 2

Problem 2: (a)

Mean $\bar{x} = \frac{1}{N}\sum x = \frac{1}{10}\sum_{i=1}^{10} x_i = 8.6$
Mean $\bar{y} = \frac{1}{N}\sum y = \frac{1}{10}\sum_{i=1}^{10} y_i = 19.6$
Standard deviation $x_{sd} = \sqrt{\frac{\sum_{i=1}^{10}(x_i-\bar{x})^2}{10-1}} = 21.3269$
Standard deviation $y_{sd} = \sqrt{\frac{\sum_{i=1}^{10}(x_i-\bar{x})^2}{10-1}} = 25.1960$
Student with unknown major: $(9, 18)$:
Normalised to : $(0.0187, -0.0635)$

ID	x	y	x_n	y_n	L1	L2
M1	10	49	0.0623	1.107	<u>1.2117</u>	1.168
M2	-12	38	-0.9163	0.6928	1.6871	1.1998
M3	-9	47	-0.7829	1.0317	1.8926	1.354
EE1	29	19	0.9074	-0.0226	<u>0.9272</u>	<u>0.8904</u>
EE2	32	31	1.0409	0.4292	1.5125	<u>1.1341</u>
EE3	37	38	1.2633	0.6928	1.9985	1.4554
CS1	8	9	-0.0267	-0.3991	<u>0.3834</u>	<u>0.3418</u>
CS2	30	-28	0.9519	-1.7922	2.6661	1.9678
CS3	-18	-19	-1.1832	-1.4534	2.5942	1.8394
CS4	-21	12	-1.3167	-0.2862	1.5605	1.3535

Procedure: We first normalise the data point with unknown major using the mean and standard deviation of the known points, and then calculated the L1 and L2 distances. L1 distance between two points (x_1, y_1) and (x_0, y_0) is defined as : $L1 = |x_1 - x_0| + |y_1 - y_0|$
L2 distance is defined as $L2 = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$ **For** $L1$:
$K = 1$: For $K = 1$ the nearest neighbor is CS1 and hence the unknown sample 'could' be a computer science
$K = 3$: For $K = 3$ the nearest neighbors are M1, EE1, CS1 and hence there is a 'tie'. Choosing the label of the least distance would again result in CS1 as $CS1 < EE1 < M1$.
For $L2$:
$K = 1$: For $K = 1$ the nearest neighbor is CS1 and hence the unknown sample 'could' be a computer science
$K = 3$: For $K = 3$ the nearest neighbors are M1, EE1, EE2. Since two nearest neighbors are from EE, we assign it the unknwon sample to be from Electrical engineering.
Comparison For $K = 1$, both $L1$ and $L2$ distance metric give the same results, however for $K = 3$, the $L1$ metric yields a tie, since the distances are similar but $L2$ metric being a square quantity of a number smaller than 1 further reduces the distances and ..... TODO

Problem 2: (b)

Total points: N
Total points with label class c: N_c
Given: $p(x|Y=c) = \frac{K_c}{N_c V}$ and $\sum K_c = K$ Class prior: $p(Y=c) = \frac{N_c}{N}$
Unconditional density $p(x) = \sum_c p(x|Y=c)p(Y=c) = \sum_c \frac{K_c}{N_c V} \times \frac{N_c}{N} = \sum_c \frac{K_c}{NV} = \frac{K}{NV}$
Posterior $P(Y=c|x) = \frac{P(x|Y=c) \times P(Y=c)}{P(x)} = \frac{\frac{K_c}{N_c V} \times \frac{N_c}{N}}{\frac{K}{NV}} = \frac{K_c}{K}$

Problem 3

Problem 3: (a)

Information gain $G = H[Y] - H[Y|X]$ where Y is the outcome variable and X is an attribute to be split. In our case $Y=$ 'Rains or not' In order to maximise gain for a fixed Y we need to minimise the conditional entropy $H[Y|X]$ $p_{rain} = \frac{9+5+6+3+7+2+3+1}{80} = frac3680 = 0.9$ and hence $p_{no-rain} = 0.1$

$$H[Y] = -p_{rain}\log(p_rain) - p_{no-rain}\log(p_{no-rain}) \tag{17}$$
$$= -(0.9\log_2(0.9) + 0.1\log_2(0.1)) \tag{18}$$

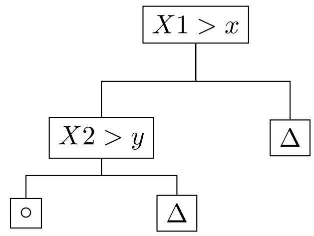
Temperature

Problem 3 continued on next page...

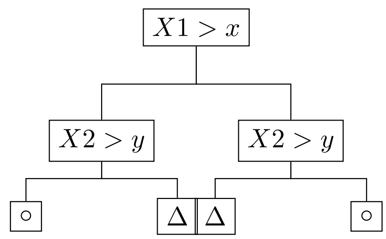
Problem 3: (b)

By default the rightmost branch correspoonds to the parent condition being a YES[FIX ME]

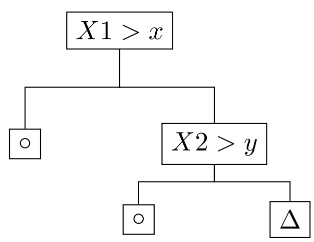
a:



b:



c:



d: