UTA ID: 1001553524 CHAWAN, VARSHA RANI

ML - Homework 1

1. MLE and MAP

1.1

(a) Performance function :

~~given~~ For a poisson process the probability ~~of~~ the first event to occur at time x after a restart is described by an exponential distribution :

$$\therefore \quad P(x/\lambda) = \lambda e^{-\lambda x}$$

$$P(D/\lambda) = \lambda^n e^{-\lambda K_i} \qquad \text{where } D = \{K_1 \cdots K_n\}$$

optimization :

Maximum likelihood estimation

we need to find λ that maximize $P(D/\lambda)$

$$\hat{\lambda} = \operatorname{argmax} P(D/\lambda) = \operatorname{argmax} \ln(P(D/\lambda))$$

Converting to Log likelihood

$$= \sum_{i=1}^{n} \ln(\lambda e^{-\lambda K_i})$$

$$= \sum_{i=1}^{n} (\ln \lambda - \lambda K_i) = n \ln \lambda - \lambda \sum_{i=1}^{n} K_i$$

Take the derivative with respect to λ & equate to 0

$$\frac{d}{d\lambda} \ln(P(D/\lambda)) = 0 \Rightarrow \frac{d}{d\lambda}\left(n \ln \lambda - \lambda \sum_{i=1}^{n} K_i\right) = 0$$

$$\frac{n}{\hat{\lambda}} - \sum_{i=1}^{n} K_i = 0$$

$$\boxed{\Rightarrow \hat{\lambda} = \frac{n}{\Sigma K_i}} = \frac{1}{\bar{K}} \quad \text{—— (a)}$$

(b)

given $D = \{1.5, 3, 2.5, 2.75, 2.9, 3\}$

from the part (a)

$$\frac{d}{d\lambda} \ln(P(D|\lambda)) = \frac{1}{\bar{K}} = \frac{n}{\Sigma K_i} = \hat{\lambda} = \frac{6}{\sum_{i=1}^{6} K_i}$$

$$\therefore \hat{\lambda} = \frac{6}{1.5+3+2.5+2.75+2.9+3}$$

$$\boxed{\hat{\lambda} = 0.383} \quad \text{—— (b)}$$

(c) given the conjugate prior to $P(D/\lambda)$ (ie) exponential function is gamma distribution

The Gamma distribution is given as

$$P_{\alpha\beta}(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

Derivation : optimization for MAP approach

$$\therefore P(\lambda/K) = P(D/\lambda)\, P_{\alpha\beta}(\lambda)$$

$$P(\lambda/K) = \left(\lambda^{n} e^{-\lambda \sum_{i=1}^{n} K_i}\right)\left(\frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}\right)$$

$= \cancel{\lambda e}$

$$= \lambda^{n+\alpha-1} e^{-\lambda \sum_{i=1}^{n} (k_i+\beta)} \frac{\beta^\alpha}{\Gamma(\alpha)} \quad \text{--- } 1$$

$$\therefore \quad P(\lambda/k) \propto \lambda^{n+\alpha-1} e^{-\lambda \sum_{i=1}^{n} (k_i+\beta)}$$

$$P(\lambda/k) = \text{argmax}\left(P(D/\lambda)P(\lambda)\right)$$

applying log on both sides

$$\ln P(\lambda/k) = \ln\left(\lambda^{n+\alpha-1} e^{-\lambda \sum_{i=1}^{n} (k_i+\beta)} \frac{\beta^\alpha}{\Gamma(\alpha)}\right)$$

(ignore $\frac{\beta^\alpha}{\Gamma(\alpha)}$ as they are constant will be ignored in differentiation)

$$= -\lambda\left(\sum_{i=1}^{n} k_i + \beta\right) + (n+\alpha-1)\ln\lambda$$

Now to find $\hat{\lambda}$ that maximises $P(\lambda/k)$ equate the differentiation of above equation to 0

$$\therefore \frac{d}{d\lambda} P(\lambda/k) = 0$$

$$\Rightarrow \frac{d}{d\lambda}\left[-\lambda\left(\sum_{i=1}^{n} k_i+\beta\right) + (n+\alpha-1)\ln\lambda\right) = 0$$

$$\Rightarrow -\sum_{i=1}^{n} k_i - \beta + \frac{n+\alpha-1}{\lambda} = 0$$

$$\boxed{\Rightarrow \hat{\lambda} = \frac{(n+\alpha-1)}{\sum_{i=1}^{n} k_i + \beta}} \quad \text{--- } c$$

Scanned by CamScanner

result for $D = \{1.5, 3, 2.5, 2.75, 2.9, 3\}$

here $n = 6$, given $\alpha = 5$, $\beta = 10$

$\therefore$ from the equation C

$$\hat{\lambda} = \frac{(n+\alpha-1)}{\sum_{i=1}^{6} x_i + \beta}$$

$$= \frac{(6+5-1)}{(1.5+3+2.5+2.75+2.9+3)+10}$$

$$= \frac{10}{25.65}$$

$$\boxed{\hat{\lambda} = 0.389} \cong 0.39$$

2(a) given traing data D = { ((170, 57, 32), W) 5
((175, 72, 30), M), }

H	W	A	G
170	57	32	W
192	95	28	M
150	45	30	W
170	65	29	M
175	78	35	M
185	90	32	M
170	65	28	W
155	48	31	W
160	55	30	W
182	80	30	M
175	69	28	W
180	80	27	M
160	50	31	W
175	72	30	M

Data points (155, 40, 35) (170, 70, 32) (175, 70, 35)
(180, 90, 20)

The step's to solve KNN

1. Calculate the cartesian distance between a data point and the data set [every element of dataset)
2. Then rank the minimum distance [(ie) in ascending order of distances]
3. check if it's included in the K^{th} neighbour & determine the gender

For dataset (155, 40, 35)

H	W	A	G	Distance	Rank	Included in nearest neighbour K_1	K_3	K_5
170	57	32	W	$(155-170)^2+(57-40)^2+(32-35)^2 = 523$	5	NO	N	YES
192	95	28	M	$(192-155)^2+(95-40)^2+(28-35)^2 = 4443$	14	NO	N	N
150	45	30	W	$(150-155)^2+(45-40)^2+(30-35)^2 = 75$	1	Yes	YES	YES
170	65	29	M	$(170-155)^2+(65-40)^2+(29-35)^2 = 886$	6	NO	N	N
175	78	35	M	$(175-155)^2+(78-40)^2+(35-35)^2 = 1844$	10	NO	N	N
185	90	32	M	$(185-155)^2+(90-40)^2+(32-35)^2 = 3409$	13	NO	N	N
170	65	28	W	$(170-155)^2+(65-40)^2+(28-35)^2 = 899$	7	NO	N	N
155	48	31	W	$(155-155)^2+(48-40)^2+(31-35)^2 = 80$	2	NO	YES	YES
160	55	30	W	$(160-155)^2+(55-40)^2+(30-35)^2 = 275$	4	NO	N	YES
182	80	30	M	$(182-155)^2+(80-40)^2+(30-35)^2 = 2354$	12	NO	N	N
175	69	28	W	$(175-155)^2+(69-40)^2+(28-35)^2 = 1290$	8	NO	N	N
180	80	27	M	$(180-155)^2+(80-40)^2+(27-35)^2 = 2289$	11	NO	N	N
160	50	31	W	$(160-155)^2+(50-40)^2+(31-35)^2 = 141$	3	NO	YES	YES
175	72	30	M	$(175-155)^2+(72-40)^2+(30-35)^2 = 1449$	9	NO	N	N

$\therefore$ when $K=1$, $G = W$; $K=3$, $G = \{w, w, w\} = W$; $K=5$, $G = \{w, w, w, w, w\} = W$

[here I used squares instead of square root for easy calculation purpose]

Similarly calculate distance for the other data sets & determine gender

			(170, 70, 32)					(175, 70, 35)					(180, 90, 20)					
H	W	A	Distance	Rank	K=1	K=3	K=5	Distance	Rank	K=1	K=3	K=5	Distance	Rank	K=1	K=3	K=5	G
170	57	32	169	6	N	N	N	~~169~~ 203	8				1333	10				W
192	95	28	1125	14	N	N	N	~~1125~~ 963	13				233	4			M	M
150	45	30	1029	13	N	N	N	1027	14				3025	14				W
170	65	29	34	2	N	M	M	86	4			M	806	9				M
175	78	35	98	5	N	N	M	64	3		M	M	394	5			M	M
185	90	32	625	11	N	N	N	509	10				169	2		M	M	M
170	65	28	41	3	N	W	W	99	5			W	789	8				W
155	48	31	710	12	N	N	N	900	12				2510	13				W
160	55	30	329	9	N	N	N	475	9				1725	11				W
182	80	30	248	8	N	N	N	174	6				204	3		M	M	W
175	69	28	42	4	N	N	W	50	2		W	W	530	7				M
180	80	27	225	7	N	N	N	189	7				149	1	M	M	M	W
160	50	31	501	10	N	N	N	641	11				2121	12				M
175	72	30	33	1	M	M	M	29	1	M	M	M	449	6				
classification:			K=1; M		K=3; M		K=5; M	K=1; M		K=3; M		K=5; M	K=1; M		K=3; M		K=5; M	

distance is calculated as:

$$D = \sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2} = \sqrt{(H_1-H_0)^2+(W_1-W_0)^2+(A_1-A_0)^2}$$

for simplicity purpose, I just calculated square of distance (D^2), this will not impact result.

Results:

Test data

① (155, 40, 35)

	Predictions	Result
K=1	W	W
K=3	W, W, W	W
K=5	W W W W W	W

② (170, 70, 32)

	predictions	Result
K = 1	M	M
K = 3	M W M	M
K = 5	M M W W M	M

③ (175, 70, 35)

	predictions	Result
K = 1	M	M
K = 3	M W M	M
K = 5	M M W W M	M

④ (180, 90, 20)

	predictions	Result
K = 1	M	M
K = 3	M M M	M
K = 5	M M M M M	M