CSE 6363

HW1

SHREYASH
SHRIVASTAVA

1

a) $P_\lambda(x) = \lambda e^{-\lambda x}$

$D = \{K_1, K_2 \dots K_n\}$

MLE optimization for parameter λ

$$f(K_1 \dots K_n / \lambda) = \lambda e^{-\lambda K_1} \cdot \lambda e^{-\lambda K_2} \dots \lambda e^{-\lambda K_n}$$

$$L = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

$$\log L = \log\left(\prod_{i=1}^{n} \lambda e^{-\lambda x_i}\right)$$

$$= \sum_{i=1}^{n} \log(\lambda) + \sum_{i=1}^{n} \log(e^{-\lambda x_i})$$

$$= N \cdot \log(\lambda) - \sum_{i=1}^{n} \lambda x_i \ln(e)$$

$$= N \log(\lambda) - \lambda \cancel{N} \sum_{i=1}^{n} x_i$$

$$\boxed{\frac{1}{N}\sum_{i=1}^{n} x_i = \bar{x} \quad \sum_{i=1}^{n} x_i = N\bar{x}}$$

$$\log L = N \log(\lambda) - \lambda N \bar{x}$$

$$\frac{\partial L}{\partial \lambda} = 0 = \frac{\partial (N \log \lambda)}{\partial \lambda} - \frac{\partial (N \bar{x} \lambda)}{\partial \lambda}$$

$$= \frac{N}{\lambda} - N\bar{x} = 0$$

$$\frac{N}{\lambda} = N\bar{x}$$

$$\boxed{\lambda = \frac{1}{\bar{x}}} \quad (\text{~~ME~~ MLE optimization})$$

b) $\lambda = \frac{1}{\bar{x}}$

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

$$\bar{x} = \frac{1.5 + 3 + 2.5 + 2.75 + 2.9 + 3}{6}$$

$$\bar{x} = 2.6083$$

$$\boxed{\lambda = \frac{1}{2.6083} = 0.3833}$$

c) Optimization for a MAP approach

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

$$\theta_{MLE} = \arg\max_{\theta} \prod_{i=1}^{n} f(x_1 \cdots x_n \mid \theta)$$

conjugate prior

$$\theta_{MAP} = \arg_{\theta}\max f(\theta \mid x_1 \cdots x_n) = \arg\max_{\theta} \frac{f(x_1 \cdots x_n \mid \theta)\, g(\theta)}{h(x_1 \cdots x_n)}$$

$$= \arg_{\theta}\max \frac{\left(\prod_{i=1}^{n} f(x_1 \cdots x_n \mid \theta)\right) g(\theta)}{h(x_1 \cdots x_n)}$$

$$= \arg\max_{\theta} g(\theta) \prod_{i=1}^{n} \left(f(x_i \mid \theta)\right)$$

$\hookrightarrow$ prior distribution of θ

$\theta_{MAP} = \arg\max_{\theta} \left(\log(g(\theta)) \right) + \sum_{i=1}^{n} \log(f(x_i \mid \theta))$

$\cancel{g(\theta).}$

$\cancel{= \arg\max_{\theta} (\log}$

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

$\theta_{MAP} = \arg\max_{\lambda} \left(\log(P_{\alpha,\beta}(\lambda)) \right) + \sum_{i=1}^{n} \log(\lambda e^{-\lambda x_i})$

$= \arg\max_{\lambda} \left(\log\left(\frac{\beta^{\lambda}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \right) \right) + \sum_{i=1}^{n} \log(\lambda e^{-\lambda x_i})$

$\theta_{MAP} = \arg\max_{\lambda} \Big(\log(\beta^{\alpha}) + \log(\lambda^{\alpha-1}) + \log(e^{-\beta\lambda}) - \log(\Gamma(\alpha)) \Big)$

$+ \sum_{i=1}^{n} \log(\lambda e^{-\lambda x_i})$

$= \arg\max_{\lambda} \Big(\alpha \log(\beta) + (\alpha-1)\log(\lambda) - \beta\lambda\log(e) - \log\Gamma(\alpha) \Big)$

$+ \sum_{i=1}^{n} \left[\log(\lambda) - \lambda x_i \log(e) \right]$

$= \arg\max_{\lambda} \Big(\alpha\log(\beta) + (\alpha-1)\log(\lambda) - \beta\lambda\log(e) - \cancel{\log}\log(\Gamma(\alpha))$

$+ \sum_{i=1}^{n} \log\lambda - \lambda\log e \sum_{i=1}^{n} x_i$

Scanned with CamScanner

$$\frac{\partial \theta_{MAP}}{\partial \lambda} = 0$$

$$\frac{\partial \theta_{MAP}}{\partial \lambda} = \frac{\partial(\alpha \log(\beta))}{\partial \lambda} + \frac{(\alpha-1)\,\partial(\log(\lambda))}{\partial \lambda} - \beta \log e \frac{\partial(\lambda)}{\partial \lambda}$$

$$- \frac{\partial(\log(\Gamma(\alpha)))}{\partial \lambda} + N \cdot \frac{\partial(\log \lambda)}{\partial \lambda} - \log e \, N\bar{x} \frac{\partial \cancel{\log}(\lambda)}{\partial \lambda}$$

$$\frac{\partial \theta_{MAP}}{\partial \lambda} = \cancel{\pm} \, 0 + \frac{(\alpha-1)}{\lambda} - \beta - 0 + N - N\bar{x}$$

$$= \frac{(\alpha-1)}{\lambda} - \beta + N - N\bar{x} = 0$$

$$(\alpha-1) - \beta\lambda + N\lambda - N\bar{x}\lambda = 0$$

$$\begin{bmatrix} \alpha = 5 \\ \beta = 10 \end{bmatrix}$$

$$(5-1) - 10\lambda + N\lambda - N\bar{x}\lambda = 0$$

$$4 - \lambda(10 \cancel{+} - N + N\bar{x}) = 0$$

$$\boxed{\lambda = \frac{4}{10 - N + N\bar{x}}}$$

2a) KNN

(155,40,35) (170,70,32) (175,70,25) (180,90,20)

By computing the distance from each point

	(155,40,35)	(170,70,32)	(125,70,26)	(180,90,20)
170, 57, 32, W	22.869	13	14.2	36.07
192, 95, 28, M	66.65	32.54	31.03	15.264
150, 45, 30, W	8.66	32.07	35.7	58
170, 65, 29, M	29.76	5.83	9.27	28.34
175 78 35 M	42.94	9.39	8.0	29.84
185, 90, 32 M	58.386	25	22.56	13.0
170 65 28 W	29.98	6.4	9.4	28.029
155 48 31 W	8.94	26.6	30	58
160 55 30 W	16.5	18.13	21.79	46.023
182 80 30 M	48.51	15.74	13.19	14.28
175 69 28 W	35.9	6.4	7.07	23.021
180 80 27 M	47.8	15	13.74	12.2
180 80 31 W	11.374	22.30	28.31	46.02

F=1	W	M	M	M
F=2	W	M	M	M
F=5	W	M	M	M

With 2 parameters

	(185, 40, 35)	(170, 70, 32)	(128, 70, 35)	(130, 90, 20
k = 1	W	M	W	M
k = 3	W	W	M	M
k = 5	W	M	M	M

If the parameters get reduced, the accuracy goes down

3a)

Parameters for Gaussian Naive Bayes

1. (155, 40, 35)

$P(height/M) = 7.42\,e^{-5}$
$P(weight/M) = 4.92\,e^{-6}$
$P(Age/M) = 0.0234$

$P(M|X) = 4.190\,e^{-12}$
$P(W|X) = 6.71\,e^{-8}$

$P(height/W) = 0.030$
$P(weight/W) = 0.0080$
$P(Age/W) = 0.0054$

Predicted as: 'female'

170, 70, 32

2. $P(height/M) = 0.02048$
$P(weight/M) = 0.02409$
$P(Age/M) = 0.1216$

$P(M|X) = 3.003\,e^{-5}$

$P(height/W) = 0.033$
$P(weight/W) = 0.010$
$P(age/W) = 0.103$

$P(W|X) = 1.73807\,e^{-5}$

Predicted as: 'male'

3. $\underline{175, 70, 35}$

$P(height \mid M) = 0.0454$

$P(weight \mid M) = 0.024$

$P(age \mid M) = 0.02348$

$P(M \mid x) = 1.286e^{-5}$

$P(height \mid W) = 0.0166$

$P(weight \mid W) = 0.0103$

$P(age \mid W) = 0.00054$

$P(W \mid x) = 4.711e^{-8}$

Predicted as 'male'

4. $\underline{180, 90, 20}$

$P(height \mid M) = 0.0587$

$P(weight \mid M) = 0.0240$

$P(age \mid M) = 3.619e^{-5}$

$P(M \mid x) = 2.516e^{-8}$

$P(height \mid W) = 0.0058$

$P(weight \mid W) = 7.3633e^{-6}$

$P(age \mid W) = 3.91771e^{-12}$

$P(W \mid x) = 8.507e^{-20}$

predicted as 'male'

8c)

155, 40

P(height | M) = 7.24 e^{-5}

P(weight | M) = 4.927 e^{-6}

P(M | x) = 1.73 e^{-10}

P(height | W) = 0.0306

P(weight | W) = 0.00804

P(W | x) = 0.000123

Predicted as female

170, 70

P(height | M) = 0.02048

P(weight | M) = 0.020409

P(M | x) = 0.00024

P(height | W) = 0.0330

P(weight | W) = 0.0103

P(W | x) = 0.00017)

Predicted as 'Male'

175, 70

P(height | M) = 0.045

P(weight | M) = 0.024

P(M | x) = 0.00054

P(height | W) = 0.0166

P(weight | W) = 0.0130

P(W | x) = 8651 e^{-5}

Predicted as 'male'

180, 90

P(height | M) = 0.058

P(weight | M) = 0.024

P(M | X) = 0.00070

P(height | W) = 0.00589

P(weight | W) = 7.3633e-6

P(W | X) = 2.121 e-8

Predicted as 'male'

3d) The Naive Bayes classifier is better compared to KNN doesn't know which attributes are more important.

However, Naive Bayes considers absolute independence of data parameters which might not always be the case.