Introduction to Machine larning

Majid Farhadloo – farha043@umn.edu

March 2020

1 Question3-programming

method: MultiGaussClassify$_WithFullMatrix, dataset: Boston50$
$method: MultiGaussClassify_WithFullMatrix$
dataset: Boston50

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	mean	std dev
0.1471	0.2376	0.1881	0.2475	0.2079	0.2056	0.0361

==============
method: MultiGaussClassify$_WithFullMatrix, dataset: Boston75$
$method: MultiGaussClassify_WithFullMatrix$
dataset: Boston75

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	mean	std dev
0.2647	0.2178	0.2772	0.2376	0.1980	0.2391	0.0292

==============
method: MultiGaussClassify$_WithFullMatrix, dataset: Digits$
$method: MultiGaussClassify_WithFullMatrix$
dataset: Digits

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	mean	std dev
0.0556	0.0361	0.0474	0.0585	0.0557	0.0506	0.0082

==============
method: MultiGaussClassify$_WithDiagonal, dataset: Boston50$
$method: MultiGaussClassify_WithDiagonal$
dataset: Boston50

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	mean	std dev
0.2157	0.2277	0.2178	0.2277	0.1584	0.2095	0.0260

==============
method: MultiGaussClassify$_WithDiagonal, dataset: Boston75$
$method: MultiGaussClassify_WithDiagonal$
dataset: Boston75

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	mean	std dev
0.2255	0.1881	0.2574	0.2376	0.2673	0.2352	0.0277

==============
method: MultiGaussClassify$_WithDiagonal, dataset : Digits$
$method : MultiGaussClassify_WithDiagonal$
dataset: Digits

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	mean	std dev
0.0583	0.0417	0.0279	0.0557	0.0446	0.0456	0.0109

==============
method: LogisticRegression, dataset: Boston50 method: LogisticRegression
dataset: Boston50

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	mean	std dev
0.2157	0.1683	0.1386	0.1386	0.1485	0.1619	0.0290

==============
method: LogisticRegression, dataset: Boston75
method: LogisticRegression
dataset: Boston75

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	mean	std dev
0.1078	0.0792	0.1485	0.0891	0.1386	0.1127	0.0270

==============
method: LogisticRegression, dataset: Digits
method: LogisticRegression
dataset: Digits

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	mean	std dev
0.0500	0.0278	0.0501	0.0279	0.0362	0.0384	0.0100

intro to ML

Q1.

(a) $p(x|\theta) = \frac{1}{\sqrt{2\pi}\ \theta} \exp\left(\frac{-x^2}{2\theta^2}\right)$

form likelihood

$$L = \prod_{i=1}^{n} \left(\frac{1}{2\pi}\right)^{1/2} \times \frac{1}{\theta} \times \exp\left(\frac{-x_i^2}{2\theta^2}\right) \quad (1)$$

$$L = \left(\frac{1}{2\pi}\right)^{n/2} \times \frac{1}{\theta^n} \times \exp\left(\frac{-\sum_{i=1}^{n} x_i^2}{2\theta^2}\right) \quad (2)$$

$$\text{Log}(2) \Rightarrow \text{Log}\, L = -\text{Log}\,(2\pi)^{n/2}\theta^n - \frac{x_i^2}{2\theta^2} \quad (3)$$

$$-n/2\, \text{Log}\, 2\pi - n\, \text{Log}\, \theta - \sum_{i=1}^{n} \frac{x_i^2}{2\theta^2} \quad (4)$$

↳ derivative of (4) and equal to zero w.r.t $\underline{\theta}$

$$\frac{\partial}{\partial\theta} \text{Log}\, L = \frac{-n}{\theta} + \frac{\sum_{i=1}^{n} x_i^2}{\theta^3} = 0 \quad \underline{(5) \times \theta}$$

$$-n + \frac{\sum_{i=1}^{n} x_i^2}{\theta^2} = 0$$

$$n = \frac{\sum_{i=1}^{n} x_i^2}{\theta^2} \rightarrow \theta = \underline{\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}} \quad (6)$$

(b) simillar to part (a) we form the likelihood formation and then take the Log and we will have the following:

$$P(x|\theta) = \frac{1}{\theta} \exp^{(-x/\theta)}$$

$$\frac{1}{\theta^n} \times \exp^{-\sum_{i=1}^{n} x_i / \theta} = -\mathrm{Log}\,\theta^n - \frac{\sum_{i=1}^{n} x_i}{\theta} \quad (1)$$

get the derivative of (1):

$$-\frac{n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = \frac{\partial}{\partial\theta} \mathrm{Log}^L = 0$$

$$-n + \frac{\sum_{i=1}^{n} x_i}{\theta} \Rightarrow \theta = \frac{\sum_{i=1}^{n} x_i}{n} = \text{avg}$$

(c) $p(x|\theta) = \theta x^{\theta-1}$ ~~same as~~ same as previous form the likelihood organization

$$\Rightarrow \theta^n * \prod_{i=1}^{n} x_i^{\theta-1} \rightarrow \mathrm{Log}^L = \mathrm{Log}\,\theta^n \sum_{i=1}^{n} \mathrm{Log}\, x_i^{\theta-1} \quad (1)$$

derivative of (1):

$$n\,\mathrm{Log}\,\theta + (\theta-1) \sum_{i=1}^{n} \mathrm{Log}\, x_i$$

$$\hookrightarrow \frac{n}{\theta} + \sum_{i=1}^{n} \mathrm{Log}\, x_i = 0 \quad (2)$$

$$\theta = \frac{-n}{\sum_{i=1}^{n} \log x_i}$$

(d) similar approaches, we will have the following:

$p(x|\theta) = \frac{1}{\theta} \quad 0 \leq x \leq \theta, \ \theta > 0$

$L = \prod_{i=1}^{n} \frac{1}{\theta} \rightarrow \quad L = \left(\frac{1}{\theta}\right)^n \rightarrow$ take the Log (1)

$\log L = -\log \theta^n \rightarrow \quad (2)$

derivative of (2) w.r.t. θ

$\frac{\partial}{\partial \theta} \log L = \frac{-n}{\theta} = 0 \quad (3)$

$\frac{-n}{\theta} = \frac{\partial L}{\partial \theta}$

↳ in order to maximize the function is through minimize θ, and based on provided interval $0 \leq x \leq \theta \quad \theta > 0$

$\theta = \max x_i$

the reason behind this as the derivative indicates, this is a decreasing function.

Q2. form the likelihood estimation:

$$L = \prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\tfrac{1}{2}(x_i-\mu)^T \Sigma^{-1} (x_i-\mu)\right] \quad (1)$$

$$\hookrightarrow L = \frac{1}{(2\pi)^{nd/2} |\Sigma|^{n/2}} \exp\left[\sum_{i=1}^{n} \left(-\tfrac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)\right)\right] \quad (2)$$

take Log (2)

$$\hookrightarrow \operatorname{Log} L = \frac{-nd}{2} \operatorname{Log} 2\pi - \frac{n}{2} \operatorname{Log}|\Sigma| - \frac{1}{2} \sum_{i=1}^{n} (x_i-\mu)^T \Sigma^{-1}(x_i-\mu) \quad (3)$$

in the following we need to take derivative twice once w.r.t μ and next time Σ.

derivative of (3) w.r.t μ

$$\hookrightarrow \frac{\partial L}{\partial \mu} = \frac{-1}{2} \sum_{i=1}^{n} -2 \Sigma^{-1} (x_i-\mu) = 0 \quad (4)$$

~~[illegible]~~ $-2 \times -\frac{1}{2}$ cancel out, and divid by Σ^{-1} we have following:

$$\sum_{i=1}^{n} (x_i-\mu) = 0 \rightarrow \sum_{i=1}^{n} x_i - n\mu \Rightarrow$$

$$\mu = \sum_{i=1}^{n} \frac{x_i}{n} \rightarrow \mu = \bar{x} \quad (5)$$

now get derivative w.r.t Σ^{-1}

$$\frac{\partial L}{\partial \Sigma^{-1}} = \frac{n}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{n}\left[(x_i-\mu)(x_i-\mu)^T\right] = 0$$

$$\Sigma = \frac{\sum_{i=1}^{n}\left[(x_i-\mu)(x_i-\mu)^T\right]}{n} \quad (6)$$

(b) estimation of $\hat{\mu}_n$:

$$E(\hat{\mu}_n) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) \rightarrow$$

$$\frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n} \times n\mu = \mu$$

it is unbiased estimate.

(c) $E[\hat{\Sigma}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T\right]$

$$\frac{1}{n}\sum_{i=1}^{n} E\left[(x_i-\mu)(x_i-\mu)^T\right] =$$

$$\rightarrow \frac{1}{n}\sum_{i=1}^{n} E(x_i x_i^T) - n E(\mu\mu^T)$$

$\frac{n-1}{n}\Sigma$ it is biased.

Q3.

(a)

compute the risk for C_1

λ xn

$p(C_1|test) \times n + p(C_2|test) \times n + p(C_3|test) + p(reject|test)$

$0 + 0.25 \times 10 + 0.25 \times 100 + 10 = \boxed{37.5}$

$2.5 + 25 + 10$

similar proceedure for C_2 and C_3 C_2

$1 \times 0.5 + 0 + 0.25 \times 100 + 10 =$

$0.5 + 25 + 10 = \boxed{35.5}$

$C_3 = 1 \times 0.5 + 10 \times 0.25 + 0 + 10 = \boxed{13}$

(b) similar proceedure as above we will have the following

$0 + 0.5 \times \overset{5}{\cancel{10}} + 0.1 \times \overset{10}{\cancel{100}} + 5 = C_1 = 20$

$0.4 + 0 + 0.1 \times \cancel{100} + 5 = C_2 = 15.4$

$0.4 + 10 \times \underset{5}{\cancel{0.5}} + 0 + 5 = C_3 = 10.4$

on part (a) base on the miss clasification cost and complexity of model & (rejection)

(a) → model C_3

(b) → mod → C_3