Machine Learning

Assignment - I

by

Mohammed Ilyas Habeeb
(mih278)

and

Faiyaz Sundrani
(fs1459)

Machine Learning

Assignment - 1

Part - I Written Exercises

1. Given:

Ans- Company A

$\mu_1 = 9.6\ g$

$\sigma_1 = 1.6\ g$

Company B

$\mu_2 = 9.2\ g$

$\sigma_2 = 1.8\ g$

Received shipment from one of two companies, containing 3 mice

(a) Given that:

$$P(A) = 2P(B)$$

Since, we know that:

$$P(A) + P(B) = 1 \quad (\because \text{priors sum upto } 1)$$

$$\Rightarrow 2P(B) + P(B) = 1$$

$$\Rightarrow 3P(B) = 1 \Rightarrow P(B) = 1/3$$

$$\therefore P(A) = 2(1/3) = \frac{2}{3}$$

$$\boxed{\therefore P(A) = \frac{2}{3} \; ; \; P(B) = \frac{1}{3}}$$

b)

Ans- Given that

Wt. of Mice 1 = 9.2 g $- x_1$

Wt. of Mice 2 = 8.9 g $- x_2$ = Data

Wt. of Mice 3 = 9.9 g $- x_3$

We know from (a) that :

$$P(A) = \frac{2}{3}, \; P(B) = \frac{1}{3}$$

Posterior probability according to Bayes' Rule is:

$$\underbrace{P(A/D)}_{\text{prob. that mice are from company A, given data}} = \frac{P(A) \cdot p(D/A)}{p(D)}$$

where $P(A)$ = prior

$p(D/A)$ = likelihood

$p(D)$ = Evidence (Total probability)

$= P(A) \cdot p(D/A) + P(B) \cdot p(D/B)$

Since, we know that the weights of mice are distributed according to a Gaussian Distribution; we can use the probability density function to calculate the likelihood:

p.d.f of Gaussian Distribution

$$= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

where μ = mean,
σ = standard deviation

Now,

$$p(D=x_1,x_2,x_3/A) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x_1-\mu_1)^2/2\sigma_1^2} * \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-(x_2-\mu_1)^2}{2\sigma_1^2}} * \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-(x_3-\mu_1)^2}{2\sigma_1^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma_1}\right)^3 e^{\frac{-1}{2\sigma_1^2}\left[(x_1-\mu_1)^2+(x_2-\mu_1)^2+(x_3-\mu_1)^2\right]}$$

$$= \left(\frac{1}{\sqrt{2*3.14*1.6}}\right)^3 e^{\frac{-1}{2*(1.6)^2}\left[(9.2-9.6)^2+(8.9-9.6)^2+(9.9-9.6)^2\right]}$$

$$= 0.0134$$

Similarly, $p(D/B) = 0.00995$

$$p(D) = P(A)\cdot p(D/A) + P(B)\cdot p(D/B)$$

$$= \left(\frac{2}{3} * 0.0134\right) + \left(\frac{1}{3} * 0.0095\right) = 0.01226$$

$$\therefore\ p(A|D) = \frac{P(A)\cdot p(D/A)}{p(D)}$$

$$= \frac{\left(\frac{2}{3} * 0.0134\right)}{0.01226}$$

$$= \underline{\underline{0.73}}$$

Therefore, posterior probability that the mice are from Company A is $\underline{\underline{0.73}}$

(c)

Ans - Likelihood of data given mice are from Company A = $p(D/A)$

" " " " B = $p(D|B)$

From (b), we know that:

$$p(D/A) = 0.0134$$
$$p(D/B) = 0.00995$$

Since $p(D/A) > p(D/B)$,

$p(D/A)$ is the Maximum Likelihood

Therefore,

the Maximum likelihood hypothesis is that the weights of the mice are from company A.

i.e., $P(D/A)$ is the M.L hypothesis.

```python
import numpy as np
import math

#Company A
prior_A = 2/3
mean_A = 9.6
stddeviation_A = 1.6

#Company B
prior_B = 1/3
mean_B = 9.2
stddeviation_B = 1.8

#values
x = [9.2, 8.9, 9.9]

def Evidence_Calculation(values, mean, stddeviation):
    '''
    i/p: values - list, mean - float, stddeviation - float
    o/p: Likelihood - float
    '''
    res = 0
    for i in values:
        int_res = (i - mean)**2
        res = res + int_res

    int_exp = ( -1/(2 * (stddeviation) **2 ) ) * res
    exp = np.exp(int_exp)
    Likelihood =( ( 1/( math.sqrt(2*np.pi) * stddeviation ) )**len(values) ) * exp

    return Likelihood

Likelihood_of_A = Evidence_Calculation(x, mean_A, stddeviation_A)
Likelihood_of_B = Evidence_Calculation(x, mean_B, stddeviation_B)

print("Prior of Company A is:", prior_A)
print("Prior of Company B is:", prior_B)

print("Likelihood of Company A is:", Likelihood_of_A)
print("Likelihood of Company B is:", Likelihood_of_B)

Evidence = (prior_A * Likelihood_of_A) + (prior_B * Likelihood_of_B)
print("Evidence is:", Evidence)

posterior_A = (prior_A * Likelihood_of_A) / Evidence
posterior_B = (prior_B * Likelihood_of_B) / Evidence
print("Posterior of Company A is:", posterior_A)
print("Posterior of Company B is:", posterior_B)
```

```
Prior of Company A is: 0.6666666666666666
Likelihood of Company A is: 0.0134153227386
Prior of Company B is: 0.3333333333333333
Likelihood of Company B is: 0.00995498580723
Evidence is: 0.0122618770948
Posterior of Company A is: 0.729378415984
Posterior of Company B is: 0.270621584016
```

2.

H_1 : has M-disease

H_2 : does not have M-disease

D_1 : +ve

D_2 : +ve

		Predicted yes	Predicted no	Total
Actual	yes	T.P	F.N (10%)	P
	no	F.P (25%)	T.N	N
		P'	N'	P+N

Given : outcomes are independent

Formula : $$P(H_1/D_1, D_2) = \frac{P(D_1/H_1)\,P(D_2/H_1)\cdot P(H_1)}{P(D_1, D_2)}$$

where : $P(D_1, D_2)$ is the Total Probability

$$= P(H_1)\cdot P(D_1|H_1)\cdot P(D_2|H_1) + P(H_2)\cdot P(D_1|H_2)\cdot P(D_2|H_2)$$

Priors :

$P(H_1) = 0.003$ (Since 0.3% of population has M-Disease)

$P(H_2) = 0.997$ (since priors sum up to 1)

$P(D_1|H_1) = 1 - F.N$
$= 1 - 10\%$
$= 1 - 0.1$
$= \underline{\underline{0.9}}$

$P(D_1|H_2) = F.P = \cancel{1-F.N}$
$= 0.25$

$P(D_2|H_1) = \underline{\underline{0.9}}$

$P(D_2|H_2) = 0.25$

$$P(H_1|D_1, D_2) = \frac{\overbrace{(0.9)(0.9)\cdot 0.003}^{=0.00243}}{[0.003\cdot(0.9)\cdot(0.9)] + [0.997\cdot(0.25)(0.25)]}$$

$$= \frac{0.00243}{0.0647425} = 0.037533305\text{o}2$$

$$= \underline{\underline{3.75\%}}$$

$$P(H_2 | D_1, D_2) = \frac{P(D_1|H_2)P(D_2|H_2) \cdot P(H_2)}{[P(D_1|H_1)P(D_2|H_1) \cdot P(H_1)] + [P(D_2|H_2)P(D_2|H_2) \cdot P(H_2)]}$$

$$= \frac{\overbrace{0.25 * 0.25 * 0.997}^{0.0623125}}{[0.9 * 0.9 * 0.003] + [0.25 * 0.25 * 0.997]}$$

$$= \frac{0.0623125}{0.0647425}$$

$$= 0.9625053095 = 96.25\%$$

$(\cancel{96.25\%})$

$P(c|x)$

(a) MAP hypothesis: patient does not have M-disease

$\hookrightarrow 0.0623125$

(b) M.L hypothesis $\hat{P}(x|c)$: patient has M-disease

(0.81) vs 0.0625

(c) Posterior probability that patient has M-Disease : 3.75%

(or)

0.0375

3.

Ans- I am defining:

$X = 1$, if head comes up

$X = 0$, if tails comes up

X takes two values, 1 and 0, with probabilites p and $1-p$.

Freq. Function of X with parameter 'θ':

$$p(x) = \begin{cases} \theta^{x}(1-\theta)^{1-x}, & \text{for } x \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$$

(a) As a function of θ, what is $P(HTHH \mid \theta)$?

Ans-

$$L(\theta) = P(HTHH \mid \theta) = \prod_{t=1}^{N} \theta^{(x^t)}(1-\theta)^{(1-x^t)}$$

$$= \theta^{1}(1-\theta)^{(1-1)} * \theta^{0}(1-\theta)^{1-0} * \theta^{1}(1-\theta)^{1-1} * \theta^{1}(1-\theta)^{1-1}$$

$$= \theta * (1-\theta) * \theta * \theta$$

$$= (\theta)^{3}(1-\theta)$$

(b) As a function of θ, what is $\log P(HTHH|\theta)$?

Ans -

$$l(\theta) = \log P(HTHH|\theta) = \log \prod_{t=1}^{N} \theta^{(x^t)} (1-\theta)^{(1-x^t)}$$

$$= \sum_t x^t \log\theta + \left(N - \sum_t x^t\right) \log(1-\theta)$$

$$= (1\cdot\log\theta + 0\cdot\log\theta + 1\cdot\log\theta + 1\cdot\log\theta) + (4 - (1+0+1+1)) \log(1-\theta)$$

$$= \log\theta + \log\theta + \log\theta + (4-3)\log(1-\theta)$$

$$= 3\log\theta + \log(1-\theta)$$

where $a = 3$,

$b = 1$

(c)

Ans - M.L.E of θ ~~is~~, i.e, $\arg\max_\theta P(HTHH|\theta)$

is same as $\arg\max_\theta \log P(HTHH|\theta)$

$\Rightarrow \arg\max_\theta \log P(HTHH)$ is nothing but the partial ~~derative~~ derivative of $l(\theta)$

(where $l(\theta) = \cancel{\arg\max_\theta} \log P(HTHH|\theta)$)

$= \cancel{\partial l}$

Since from (b), we know that:

$$l(\theta) = 3\log\theta + \log(1-\theta)$$

$$\frac{\partial l(\theta)}{\partial\theta} \Rightarrow \frac{3}{\theta} - \frac{1}{1-\theta} = 0$$

$$\Rightarrow \frac{3(1-\theta) - \theta}{\theta(1-\theta)} = 0$$

$$\Rightarrow 3(1-\theta) - \theta = 0$$

$$\Rightarrow 3 - 3\theta - \theta = 0$$

$$\Rightarrow \cancel{+}4\theta = \cancel{+}3$$

$$\Rightarrow \hat{\theta} = \frac{3}{4}$$

Alternatively, M.L.E of Bernoulli Density $\Rightarrow \hat{\theta} = \frac{\sum_t x^t}{N}$ (or) (Freq. Estimation)

$$= \frac{1+0+1+1}{4}$$

$$= 3/4$$

4.

Classes: $C_1 = +$
$C_2 = -$

$x_1 \in \{High, Medium, Low\}$
$x_2 \in \{Yes, No\}$
$x_3 \in \{Red, Green\}$

$k=3$

(a) $\hat{p}(x_1 = High \mid +) = \frac{1}{2}$

$= \frac{(\#\text{trials with outcome 'v'}) + m}{(\#\text{trials}) + km}$ where m = smoothing constant, k = no. of different types of attribute

$= \frac{1+0.2}{2+3(0.2)}$

$= \frac{1.2}{2.6} = \frac{12}{26} = 0.4615$. ~~or 0.4285, if k=4~~

(b) Classify:

$\hat{p}(x_1 = High, x_2 = Yes, x_3 = Green \mid C_1 = '+') * P(+)$

$= [\hat{p}(High/+) * \hat{p}(Yes \mid +) * \hat{p}(Green \mid +)] * P(+)$

$= \left[\frac{1+0.2}{2+k(0.2)} * \frac{0+0.2}{2+k(0.2)} * \frac{1+0.2}{2+k(0.2)}\right] * \frac{2}{5}$

$= \left[\frac{1.2}{2+0.2k} * \frac{0.2}{2+0.2k} * \frac{1.2}{2+0.2k}\right] * \frac{2}{5}$

$\left[\frac{1.2}{2.6} * \frac{0.2}{2.4} * \frac{1.2}{2.4}\right] * \frac{2}{5} = 0.00769$

$= 0.769\%$

$= 0.4615 * 0.0833 * 0.5 * 0.4$

$\hat{p}(-/High, Yes, Green)$

$$= \hat{p}(High, Yes, Green \mid -) * P(-)$$

$$= [\hat{p}(High \mid -) * \hat{p}(Yes \mid -) * \hat{p}(Green \mid -)] * P(-)$$

$$= \left[\frac{1+0.2}{3+k(0.2)} * \frac{2+0.2}{3+k(0.2)} * \frac{2+0.2}{3+k(0.2)}\right] * \frac{3}{5}$$

$$= \left[\frac{1.2}{3+0.2k} * \frac{2.2}{3+0.2k} * \frac{2.2}{3+0.2k}\right] * \frac{3}{5}$$

~~Since~~

$$= \left[\frac{1.2}{3+0.2(3)} * \frac{2.2}{3+0.2(2)} * \frac{2.2}{3+0.2(2)}\right] * \frac{3}{5}$$

$$= \left[\frac{1.2}{3.6} * \frac{2.2}{3.4} * \frac{2.2}{3.4}\right] * \frac{3}{5}$$

$$= 0.333 * 0.647 * 0.647 * 0.6 = 0.0837$$
$$= 8.37\%$$

Since

$$V_{NB} = \underset{v_j \in V}{\operatorname{argmax}} [\hat{p}(a_1 \mid v_j) * \hat{p}(a_2 \mid v_j) \cdots \hat{p}(a_n \mid v_j)] * P(v_j)$$

and $\hat{p}(- \mid High, Yes, Green) > \hat{p}(+ / High, Yes, Green)$

i.e, $0.0837 > 0.00769$

we get $V_{NB} = 0.0837$ and $\therefore$, the resulting classification For (High, Yes, Green) is '$-$' (minus)

5.

a)

i. Estimated value of P(C) for C = 1 is **0.745**

ii. Estimated value of P(C) for C = 2 is **0.255**

iii. $(\hat{\mu}, \hat{\sigma^2})_{RI,Class1} = (1.518662, 0.000010)$

iv. $(\hat{\mu}, \hat{\sigma^2})_{Ca,Class2} = (9.060196, 2.509274)$

v.

	Predicted_Class_GNB
indextoUse	
20	1
60	1
100	1
140	2
180	2

vi. % Training Error = 10.0 %

b)

i. % 5-fold cross-validation error = 10.0 %

ii.

	Predicted_Class_CV_GNB
indextoUse	
20	1
60	1
100	1
140	2
180	2

c)

Accuracy using Zero-R = 74.5 %

————————x————x——————-x————