Siddharth Pansuria

CS 559 A : Machine Learning

Homework Assignment 1

* Problem 1 : Probability

Major	Session 1	Session 2	Session 3
CS	6	10	6
Stat	8	10	6
Mg	6	0	8
	$p(S1) = 0.2$	$p(S2) = 0.2$	$p(S3) = 0.6$

1)

*

$$p(CS) = \sum_{n=1}^{3} p(CS, \text{session } n)$$

$$= p(CS, S1) + p(CS, S2) + p(CS, S3)$$

We know that,

$$p(CS \mid S1) = \frac{p(CS, S1)}{p(S1)}$$

$$\therefore p(CS, S1) = p(CS \mid S1) \cdot p(S1)$$

$p(CS \mid S1)$ (probability of student being CS major given that the session 1 is already selected) $= \dfrac{\text{Total CS students in S1}}{\text{Total S1 students}}$

$$= \frac{6}{20}$$

Similarly,

$$p(CS/S2) = \frac{10}{20} \qquad p(CS/S3) = \frac{6}{20}$$

$$\therefore \; P(CS, S1) = p(CS/S1) \cdot p(S1)$$
$$= \frac{6}{20} \cdot \frac{2}{10} = \frac{6}{100}$$

$$P(CS, S2) = p(CS/S2) \cdot p(S2)$$
$$= \frac{10}{20} \cdot \frac{2}{10} = \frac{10}{100}$$

$$p(CS, S3) = p(CS/S3) \cdot p(S3)$$
$$= \frac{6}{20} \cdot \frac{6}{10} = \frac{18}{100}$$

$$\therefore \; p(CS) = p(CS, S1) + p(CS, S2) + p(CS, S3)$$
$$= \frac{6}{100} + \frac{10}{100} + \frac{18}{100}$$
$$= \frac{34}{100}$$

$$\boxed{p(CS) = 0.34}$$

P.T.O.

2) According to Baye's Rule ;

$$p(Stat \mid S3) = \frac{p(S3 \mid stat) \cdot p(stat)}{p(S3)}$$

$$\therefore \quad p(S3 \mid Stat) = \frac{p(stat \mid S3) \cdot p(S3)}{p(stat)}$$

$$p(stat \mid S3) = \frac{\text{Total Stat students in S3}}{\text{Total students in S3}}$$

$$= \frac{6}{20}$$

$$p(Stat) = \sum_{m=1}^{3} p(stat, Sm)$$

$$p(stat) = p(stat, S1) + p(stat, S2) + p(stat, S3)$$

$$= p(stat \mid S1) \cdot p(S1) + p(stat \mid S2) \cdot p(S2) + p(stat \mid S3) \cdot p(S3)$$

$$= \frac{8}{20} \cdot \frac{2}{10} + \frac{10}{20} \cdot \frac{2}{10} + \frac{6}{20} \cdot \frac{6}{10}$$

$$= \frac{36}{100}$$

P.T.O.

$$\therefore \quad p(s3 \mid stat) = \frac{p(stat \mid s3) \cdot p(s3)}{p(stat)}$$

Substituting the values,

$$p(s3 \mid stat) = \frac{\frac{6}{20} \cdot \frac{6}{10}}{\frac{36}{100}} = \frac{1}{2} = 0.5$$

$$\boxed{p(s3 \mid stat) = 0.5}$$

\# Problem : 2 ML Estimation

112, 120, 131, 128, 145, 158, 157, 136, 148, 176

$\mu, \sigma^2 \rightarrow$ unknown mean and variance

i) The probability density function for every x_i is as follows:

$$P(x_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2} \frac{(x_i - \mu)^2}{\sigma^2}\right)$$

Hence the likelihood function can be defined as,

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2} \frac{(x_i - \mu)^2}{\sigma^2}\right)$$

$$\star \quad L(\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^5} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} \right)$$

2) taking log on both sides,

$$\log(L(\mu, \sigma^2)) = -5 \log 2\pi - 5 \log \sigma^2 - \frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2}$$

for maximing likelihood function with respect to μ, we derivate $\log(L(\mu, \sigma^2))$ with respect to μ and equate it with 0.

$$\frac{\partial}{\partial \mu}(\log(L(\mu, \sigma^2))) = \frac{\partial}{\partial \mu}\left(-5\log 2\pi - 5\log \sigma^2 - \frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2}\right)$$

$$= \frac{\partial}{\partial \mu}\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} \right)$$

(everything except μ and x_i are constant)

$$= -\frac{1}{2} \sum_{i=1}^{n} \frac{2(x_i - \mu)(-1)}{\sigma^2}$$

$$\frac{\partial}{\partial \mu}(\log(L(\mu, \sigma^2))) = 0$$

$$-\frac{1}{2} \sum_{i=1}^{n} \frac{2(x_i - \mu)(-1)}{\sigma^2} = 0$$

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

$$n\mu = \sum_{i=1}^{n} x_i$$

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Hence, maximum likelihood estimator of μ is as follows,

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\hat{\mu} = \frac{1}{10}(112 + 120 + 131 + 126 + 145 + 158 + 157 + 136 + 148 + 176)$$

$$= \frac{1}{10}(1409)$$

$$= 140.9$$

$$\boxed{\hat{\mu} = 140.9} \quad \text{(in pounds)}$$

For maximizing likelihood function, with respect to σ^2, we derivate $\log(L(\mu, \sigma^2))$ with respect to σ^2 and equate it with 0.

$$\frac{\partial}{\partial \sigma^2}(\log(L(\mu, \sigma^2))) = \frac{\partial}{\partial \sigma^2}\left(-5\log 2\pi - 5\log \sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$= \frac{\partial}{\partial \sigma^2}\left(-5\log \sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

P.T.O.

$$= \frac{-5}{\sigma^2} - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2 \frac{(-1)}{(\sigma^2)^2}$$

$$= \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{(\sigma^2)^2} - \frac{5}{\sigma^2}$$

$$\frac{\partial}{\partial \sigma^2}\left(\log\left(L(\mu,\sigma^2)\right)\right) = 0$$

$$\frac{5}{\sigma^2} = \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{(\sigma^2)^2}$$

$$\sigma^2 = \frac{1}{10}\sum_{i=1}^{n}(x_i-\mu)^2$$

Hence, maximum likelihood estimator of σ^2 is as follows,

$$\hat{\sigma}^2 = \frac{1}{10}\sum_{i=1}^{n}(x_i-\mu)^2$$

Now $\hat{\mu} = 140.9$, we will substitute this in the above equation.

$$\hat{\sigma}^2 = \frac{1}{10}\sum_{i=1}^{n}(x_i-140.9)^2$$

$$\boxed{\hat{\sigma}^2 = 346.69}$$

* Problem 3 : ML estimation

x	1	2	3	4
$p(x)$	$\frac{2q}{3}$	$\frac{q}{3}$	$\frac{2(1-q)}{3}$	$\frac{1-q}{3}$

$(4, 1, 3, 2, 4, 3, 2, 1, 3, 2)$

1) Now the likelihood function for the data set given,

$$L(q) = \prod_{x=1}^{m} P(I_x)$$

$$= P(4) \cdot P(1) \cdot P(3) \cdot P(2) \cdot P(4) \cdot P(3) \cdot P(2) \cdot P(1) \cdot P(3) \cdot P(2)$$

$$= \left(\frac{2q}{3}\right)^2 \left(\frac{q}{3}\right)^3 \left(\frac{2(1-q)}{3}\right)^3 \left(\frac{1-q}{3}\right)^2$$

In the above step, we substitute the probability of each event x from the table provided.

$$\therefore L(q) = \prod_{i=1}^{m} P(x_i) = \left(\frac{2q}{3}\right)^2 \left(\frac{q}{3}\right)^3 \left(\frac{2(1-q)}{3}\right)^3 \left(\frac{1-q}{3}\right)^2$$

Taking log on both sides,

$$\log(L(q)) = 2\log\frac{2q}{3} + 3\log\frac{q}{3} + 3\log\frac{2(1-q)}{3} + 2\log\frac{(1-q)}{3}$$

$$= 7\log 2 - 10\log 3 + 5\log q + 5\log(1-q)$$

$$\log(L(q)) = 7\log 2 - 10\log 3 + 5(\log q + \log(1-q))$$

Now for maximizing, we can derivate the log likelihood with respect to q,

$$\frac{\partial(\log(L(q)))}{\partial q} = \frac{\partial}{\partial q}(5(\log q + \log(1-q)))$$

$$= \frac{5}{q} + \frac{5(-1)}{1-q}$$

$$= \frac{5}{q} - \frac{5}{1-q}$$

$$\frac{\partial(\log(L(q)))}{\partial q} = 0$$

$$\frac{5}{q} = \frac{5}{1-q}$$

$$\boxed{q = \frac{1}{2} = 0.5}$$

* Problem 4 : MAP estimation

$$p(y|x, w, \beta) = N(y|f(x,w), \beta^{-1})$$

$$p(w|\alpha) = \left(\frac{\alpha}{2\pi}\right)^{\frac{(M+1)}{2}} \exp\left(-\frac{\alpha}{2} w^T w\right)$$

$$p(y|x, w, \beta) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\beta^{-1}}} \exp\left(-\frac{1}{2} \frac{(y_n - f(x_n, w))^2}{\beta^{-1}}\right)$$

* In the above step we just substituted values from the formula,

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right)$$

$$\therefore p(y|x, w, \beta) = \prod_{n=1}^{N} \frac{\sqrt{\beta}}{\sqrt{2\pi}} \exp\left(\frac{-\beta}{2} (y_n - f(x_n, w))^2\right)$$

$$p(w|\alpha) = \left(\frac{\alpha}{2\pi}\right)^{\frac{(M+1)}{2}} \exp\left(\frac{-\alpha}{2} w^T w\right)$$

Now, we know that according to Bayes theorem,

posterior $\propto$ likelihood $\times$ prior

$\therefore$ posterior = likelihood $\times$ prior $\times$ constant

$$p(w|x, y, \alpha, \beta) = k \times p(y|x, w, \beta) \times p(w|\alpha)$$

Now we have to maximize $p(w|x, y, \alpha, \beta)$ ie posterior. But the expressions on the right hand side are too complex. So taking log on both sides.

$$\log(p(w|x, y, \alpha, \beta)) = \log k + \log(p(y|x, w, \beta)) + \log(p(w|\alpha))$$

Now we know,

$$p(y|x, w, \beta) = \prod_{n=1}^{N} \frac{\sqrt{\beta}}{\sqrt{2\pi}} \exp\left(-\frac{\beta}{2}(y_n - f(x_n, w))^2\right)$$

$$\log(p(y|x, w, \beta)) = \log\left(\left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \exp\left(-\frac{\beta}{2}\sum_{n=1}^{N}(y_n - f(x_n, w))^2\right)\right)$$

$$= \frac{N}{2}\log\beta - \frac{N}{2}\log 2\pi - \frac{\beta}{2}\sum_{n=1}^{N}(y_n - f(x_n, w))^2$$

$$p(w|\alpha) = \left(\frac{\alpha}{2\pi}\right)^{\frac{M+1}{2}} \exp\left(-\frac{\alpha}{2} w^T w\right)$$

$$\log(p(w|\alpha)) = \log\left(\left(\frac{\alpha}{2\pi}\right)^{\frac{M+1}{2}} \exp\left(-\frac{\alpha}{2} w^T w\right)\right)$$

$$= \frac{(M+1)}{2}\log\alpha - \frac{(M+1)}{2}\log 2\pi - \frac{\alpha}{2} w^T w$$

P.T.O.

$\log(p(w|x, y, \alpha, \beta))$

$$= \log k + \log P(y|x, w, \beta) + \log(p(w|\alpha))$$

Substituting values we get,

$$= \log k + \frac{N}{2}\log\beta - \frac{N}{2}\log 2\pi - \frac{\beta}{2}\sum_{n=1}^{N}(y_n - f(x_n, w))^2$$

$$+ \frac{(M+1)}{2}\log\alpha - \frac{(M+1)}{2}\log 2\pi - \frac{\alpha}{2}w^Tw$$

$$= \underset{\text{(constant)}}{C} - \left(\frac{\beta}{2}\sum_{n=1}^{N}(y_n - f(x_n, w))^2 + \frac{\alpha}{2}w^Tw\right)$$

* We have to maximize posterior and so we can say we have to maximize log(posterior). Since the sum of squares is in negative, we can say that to maximize posterior we have to ~~minimize~~ minimize the term that is being deducted. Another way would be taking the negative logarithm and minimizing ~~maximizing~~ the positive part. Because if we need posterior to be maximum, then -log(posterior) has to be minimum ~~negative~~

$$\therefore -\log(\text{posterior}) = \frac{\beta}{2}\sum_{n=1}^{N}(f(x_n, w) - y_n)^2 + \frac{\alpha}{2}w^Tw - C$$

P.T.O.

Thus minimizing the $-\log(\text{posterior})$ would result in maximizing posterior.

Hence to minimize we have minimize the positive part of the right hand side.

$\therefore$ Minimizing $\frac{\beta}{2}\sum_{n=1}^{N}(f(x_n, w) - y_n)^2 + \frac{\alpha}{2} w^T w$

would result in maximising the posterior.