22F CS-559 A HW 1

Jiangrui Zheng

September 26, 2022

1 Problem 1

1. The problem states that we should find the probability of select a CS student $P(CS)$. We know that:

$$P(S1) = 0.2, P(S2) = 0.2, P(S3) = 0.6\,. \tag{1}$$

$P(S1) = 0.2, P(S2) = 0.2, P(S3) = 0.6$

So we can get:

$$\begin{aligned} P(CS|S1) &= 6/(6+8+6) = 0.3 && (2)\\ P(CS|S2) &= 10/20 = 0.5 && (3)\\ P(CS|S3) &= 6/20 = 0.3 && (4)\\ & && (5) \end{aligned}$$

Conditional probability:

$$\begin{aligned} P(CS) &= P(CS|S1)P(S1) + P(CS|S2)P(S2) + P(CS|S3)P(S3) && (6)\\ &= 0.3 \times 0.2 + 0.5 \times 0.2 + 0.3 \times 0.6 && (7)\\ &= 0.34 && (8) \end{aligned}$$

Thus, the probability of select a student majored in CS is 0.34.

2. We need find $P(S3|STAT)$.
First, using Conditional probability:

$$\begin{aligned} P(STAT) &= P(STAT|S1)P(S1) + P(STAT|S2)P(S2) + P(STAT|S3)P(S3) && (9)\\ &= 0.4 \times 0.2 + 0.5 \times 0.2 + 0.3 \times 0.6 && (10)\\ &= 0.36 && (11) \end{aligned}$$

Using Bayes' theorem,

$$\begin{aligned} P(S3|STAT) &= \frac{P(STAT|S3)P(S3)}{P(STAT)} && (12)\\ &= \frac{0.3 \times 0.6}{0.36} && (13)\\ &= 0.5 && (14) \end{aligned}$$

Thus the probability $P(S3|STAT)$ is 0.5.

2 Problem 2

1. Assume that weights of students are normally distributed, we know the probability density function of the Univariate Gaussian is that:

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{15}$$

Also we know there are 10 students so n is 10, then the likelihood function is that:

$$\begin{aligned} P(x|\mu, \sigma^2) &= \prod_{n=1}^{10} N(x_n|\mu, \sigma^2) &(16)\\ &= (\frac{1}{\sqrt{2\pi\sigma^2}})^{10} e^{\sum_1^{10} -\frac{(x_n-\mu)^2}{2\sigma^2}} &(17)\\ &= (2\pi\sigma^2)^{-5} e^{\sum_1^{10} -\frac{(x_n-\mu)^2}{2\sigma^2}} &(18) \end{aligned}$$

Thus we can give the probability of the data set given the two parameters.

2. We want to derive and calculate the solution for both μ and σ^2 using Maximum Likelihood Estimation. The corresponding Log-likelihood function is:

$$\begin{aligned} \ln P(x|\mu, \sigma^2) &= \ln \left(2\pi\sigma^2\right)^{-5} \exp\left(-\sum_{n=1}^{10} \frac{(x_n-\mu)^2}{2\sigma^2}\right) &(19)\\ &= \ln \left(2\pi\sigma^2\right)^{-5} + \ln\exp\left(-\sum_{n=1}^{10} \frac{(x_n-\mu)^2}{2\sigma^2}\right) &(20)\\ &= -5\ln 2\pi\sigma^2 - \sum_{n=1}^{10} \frac{(x_n-\mu)^2}{2\sigma^2} &(21)\\ &= -5\left(\ln 2\pi + \ln\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=1}^{10}(x_n-\mu)^2 &(22)\\ &= -5\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=1}^{10}(x_n-\mu)^2 &(23) \end{aligned}$$

Maximizing Log-likelihood with respect to μ and σ^2 by setting their derivatives to 0. By equation 23, we can make:

$$\frac{\partial \ln P}{\partial \mu} = \frac{1}{\sigma^2}\sum_{n=1}^{10}(x_n-\mu) = 0 \tag{24}$$

$$\sum_{n=1}^{10}(x_n-\mu) = 0 \tag{25}$$

$$\mu = \frac{1}{10}\sum_{i=1}^{10} x_n = \bar{x} \tag{26}$$

So we can see μ is mean.
Then let's see derivative of σ^2:

$$\frac{\partial \ln P}{\partial \sigma^2} = -\frac{5}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=1}^{10}(x_n-\mu)^2 = 0 \tag{27}$$

$$\sum_{n=1}^{10}(x_n-\mu)^2 = 10\sigma^2 \tag{28}$$

$$\sigma^2 = \frac{1}{10}\sum_{n=1}^{10}(x_n-\mu)^2 \tag{29}$$

So we can state that:

$$\mu = \bar{x}, \quad \sigma^2 = \frac{1}{10}\sum_{n=1}^{10}(x_n-\mu)^2$$

Import weights of students, using calculator:
Then we get $\mu = 140.9$ and $\sigma^2 = 346.69$.

3 Problem 3

1. Suppose that X is a discrete random variable, there are 10 independent observations.
   Two random variables are independent if the joint p.m.f. is the product of the marginals:

   $$P(X=x, Y=y) = P(X=x)P(Y=y)$$

   By Maximum Likelihood Estimation

   $$\begin{aligned} P(X \mid \Theta = q) &= \prod_{i=1}^{N} P\left(X = x_i \mid \Theta = q\right) &(30)\\ &= P(1)^2P(2)^3P(3)^3P(4)^2 &(31)\\ &= \left(\frac{2q}{3}\right)^2\left(\frac{q}{3}\right)^3\left(\frac{2(1-q)}{3}\right)^3\left(\frac{1-q}{3}\right)^2 &(32)\\ &= \frac{32q^5\left(1-q\right)^5}{59049} &(33)\end{aligned}$$

   So we can get the probability of the data set given the parameter q.

2. The corresponding Log-likelihood function is:

   $$\begin{aligned} \ln P(X \mid \Theta = q) &= \ln\frac{32q^5\left(1-q\right)^5}{59049} &(34)\\ &= \ln\left(32q^5\left(1-q\right)^5\right) - \ln\left(59049\right) &(35)\\ &= \ln\left(32\right) + 5\ln\left(q\right) + 5\ln\left(1-q\right) - \ln\left(59049\right) &(36)\end{aligned}$$

Maximizing Log-likelihood with respect to q by setting their derivatives to 0.

$$\frac{\partial \ln P}{\partial q} = \frac{\partial}{\partial\, q}\left(\ln\left(q^5\,(1-q)^5\right)\right) = 0 \tag{37}$$

$$\frac{1}{q^5\,(1-q)^5}\frac{\partial}{\partial\, q}\left(q^5\,(1-q)^5\right) = 0 \tag{38}$$

$$\frac{5\,(-2q+1)}{q\,(1-q)} = 0 \tag{39}$$

$$q = \frac{1}{2} \tag{40}$$

Hence we derive and calculate the Maximum Likelihood estimate of q that $q = \frac{1}{2}$.

4 Problem 4

1. Referring the slides:
   Using our training data to determine the unknown parameters w, β by maximum likelihood:

$$p(y \mid x, w, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(y_n \mid f\left(x_n, w\right), \beta^{-1}\right)$$

Log Likelihood:

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

Prior probability distribution is :

$$p(w \mid \alpha) = \mathcal{N}\left(w \mid 0, \alpha^{-1}\right) = \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2}\exp\left(-\frac{\alpha}{2}w^T w\right)$$

Using Bayes's theorem, combining the prior together with the likelihood term:

$$p(w \mid x, y, \alpha, \beta) \propto p(y \mid x, w, \beta)p(w \mid \alpha)$$

Calculate Log on both sides:

$$\ln p(w \mid x, y, \alpha, \beta) \tag{41}$$

$$\propto \ln p(y \mid x, w, \beta) + \ln p(w \mid \alpha) \tag{42}$$

$$\propto -\frac{\beta}{2}\sum_{n=1}^{N}\left(f\left(x_n, w\right) - y_n\right)^2 + \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi) + \frac{m+1}{2}\ln\left(\frac{a}{2\pi}\right) - \frac{\alpha}{2}w^T w \tag{43}$$

$$\propto -\frac{\beta}{2}\sum_{n=1}^{N}\left(f\left(x_n, w\right) - y_n\right)^2 - \frac{\alpha}{2}w^T w + c \tag{44}$$

Since c is a constant, it does not affect after derivative to get max.
Hence, taking the negative logarithm, we can see that Maximum posterior is equivalent to minimizing the regularized sum-of-squares error function:

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

5 Problem 5

By using 5-fold cross validation to compare their performance based on mean squared error(MSE), we observe that MSE of Elastic Net is smallest. Hence we can state that Elastic model is best of them. See code for details.