Homework Assignment 1

Problem 1:

1. $p(S1) = 0.2 \qquad p(S2) = 0.2 \qquad p(S3) = 0.6$

$$P(CS|S1) = \frac{6}{20} = \frac{3}{10}$$

$$\cancel{P(STAT)}\ P(CS|S2) = \frac{10}{20} = \frac{1}{2}$$

$$P(CS|S3) = \frac{6}{20} = \frac{3}{10}$$

$$P(CS) = P(CS|S1)P(S1) + P(CS|S2)P(S2) + P(CS|S3)P(S3)$$

$$= \frac{3}{10} \times 0.2 + \frac{1}{2} \times 0.2 + \frac{3}{10} \times 0.6$$

$$= 0.34$$

2. Using Baye's theorem,

$$P(S3|STAT) = \frac{P(STAT|S3)\,P(S3)}{P(STAT)}$$

$$P(STAT) = P(STAT|S1)P(S1) + P(STAT|S2)P(S2) + P(STAT|S3)P(S3)$$

$$= \frac{4}{10} \times 0.2 + \frac{1}{2} \times 0.2 + \frac{3}{10} \times 0.6$$

$$= 0.36$$

$$P(S3|STAT) = \frac{\frac{3}{10} \times 0.6}{0.36}$$

$$= 0.5$$

Problem 2:

1) The probability of the data set given μ and σ^2 (the likelihood function):

$$P(x|\mu,\sigma^2) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu,\sigma^2)$$

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

$$= (2\pi\sigma^2)^{-N/2} \exp\left(\frac{-1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right)$$

2) The log-likelihood function:

$$\log(P(x|\mu,\sigma^2)) = \log\left((2\pi\sigma^2)^{-N/2}\exp\left(\frac{-1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right)\right)$$

$$= \log\left((2\pi\sigma^2)^{-N/2}\right) + \log\left(\exp\left(-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right)\right)$$

$$= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2$$

$$= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{n}(x_n-\mu)^2$$

Maximizing log-likelihood with respect to μ and σ^2

$$\max_{\mu,\sigma^2} \log\left(P(x|\mu,\sigma^2)\right)$$

The first order conditions for a maximum are

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

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

The partial derivative of the log-likelihood with respect to the mean is

$$\frac{\partial}{\partial \mu} \log\left(P(x|\mu,\sigma^2)\right) = \frac{\partial}{\partial \mu}\left(-\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right)$$

$$= \frac{\partial}{\partial \mu}\left(-\frac{N}{2}\log(2\pi)\right) + \frac{\partial}{\partial \mu}\left(-\frac{N}{2}\log(\sigma^2)\right) + \frac{\partial}{\partial \mu}\left(-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right)$$

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

$$= \frac{1}{\sigma^2}\left(\sum_{n=1}^{N}x_n - N\mu\right)$$

This is equal to zero only if

$$\sum_{n=1}^{N} x_n - N\mu = 0$$

$$\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$$

The partial derivative of the log-likelihood with respect to the variance is

$$\frac{\partial}{\partial\sigma^2}\left(\log(P(x|\mu,\sigma^2))\right) = \frac{\partial}{\partial\sigma^2}\left(-\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right)$$

$$= \frac{\partial}{\partial\sigma^2}\left(-\frac{N}{2}\log(2\pi)\right) + \frac{\partial}{\partial\sigma^2}\left(-\frac{N}{2}\log(\sigma^2)\right) + \frac{\partial}{\partial\sigma^2}\left(-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right)$$

$$= -\frac{N}{2\sigma^2} - \left[-\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right]\frac{\partial}{\partial\sigma^2}\left(\frac{1}{\sigma^2}\right)$$

$$= -\frac{N}{2\sigma^2} - \left[\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right]\left(-\frac{1}{(\sigma^2)^2}\right)$$

$$= -\frac{N}{2\sigma^2} + \left[\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right]\frac{1}{(\sigma^2)^2}$$

$$= \frac{1}{2\sigma^2}\left[\frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - N\right]$$

This is equal to zero only if

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

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

Maximizing log-likelihood with respect to μ and σ^2:

$$\mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

$$\sigma^2_{ML} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu_{ML})^2$$

Problem 2:

Calculation: Weights: 112, 120, 131, 126, 145, 158, 157, 136, 148, 176

Mean (μ) $= \frac{1}{N}\sum_{n=1}^{N} x_n$

$$= \frac{112+120+131+126+145+158+157+136+148+176}{10}$$

$$= \frac{1409}{10}$$

$$= 140.9$$

Variance (σ^2) $= \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2$

$$= \frac{(28.9)^2+(-20.9)^2+(-9.9)^2+(-14.9)^2+(4.1)^2+(17.1)^2+(16.1)^2+(-4.9)^2+(7.1)^2+(35.1)^2}{10}$$

$$= \frac{835.21+436.81+98.01+222.01+16.81+292.41+259.21+24.01+50.41+1232.01}{10}$$

$$= \frac{3466.9}{10}$$

$$= 346.69$$

Problem 3:

1) $$L(q|x) = \prod_{n=1}^{N} \text{Probability mean function}$$

$$= P(4).P(1).P(3).P(2).P(4).P(3).P(2)$$
$$. P(1).P(3).P(2)$$

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

2) The log-likelihood function:

$$\log(L(q|x)) = \log\left(\left(\frac{2q}{3}\right)^2 . \left(\frac{q}{3}\right)^3 . \left(\frac{2(1-q)}{3}\right)^3 \left(\frac{1-q}{3}\right)^2\right)$$

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

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

$$= 2\left[\log\frac{2}{3} + \log q\right] + 3\left[\log q + \log\frac{1}{3}\right] + 3\left[\log\frac{2}{3} + \log(1-q)\right]$$
$$+ 2\left[\log(1-q) + \log\frac{1}{3}\right]$$

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

$$= 5 \log q + 5 \log(1-q) + \text{const.}$$

Maximizing log likelihood w.r.t. q

$$\max_{q} \log(L(q|X))$$

The first order conditions for a maximum are

$$\frac{d}{dq} \log(L(q|X)) = 0$$

The derivative of the log likelihood w.r.t. q is

$$\frac{d}{dq} \log(L(q|X)) = \frac{d}{dq}(5 \log q + 5 \log(1-q) + \text{const.})$$

$$= \frac{d}{dq}(5 \log q) + \frac{d}{dq}(5 \log(1-q)) + \frac{d}{dq}(\text{const.})$$

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

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

This is equal to zero only if

$$\frac{5}{q} - \frac{5}{(1-q)} = 0$$

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

$$5(1-q) = 5q$$

$$5 - 5q = 5q$$

$$10q = 5$$

$$q = \frac{5}{10}$$

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

Problem 4 :

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

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

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

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

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

The log-likelihood function:

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

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

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

$$\log(p(y|x,w,\beta)) = \frac{N}{2}\log\beta - \frac{N}{2}\log 2\pi - \beta E_D(w)$$

where $E_D(w)$ is the sum of squares error function

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

Prior Gaussian distribution for w:

$$p(w|\alpha) = \cancel{\left(\frac{\alpha}{2\pi}\right)} \mathcal{N}(w|0,\alpha^{-1})$$

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

Using Baye's theorem, the posterior distribution for w:

$$p(w|x,y;\alpha,\beta) \propto p(y|x,w,\beta).p(w|\alpha)$$

Since the posterior is proportional to the product of likelihood and prior, the log of the posterior distribution is proportional to the sum of the log likelihood and the log of the prior

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

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

Taking the negative logarithm of maximum posterior

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

$$= -\left[\log(p(y|x,w,\beta)) + \log(p(w|\alpha))\right]$$

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

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

$$= \beta E_D(w) + \alpha E_w(w) + \text{Const.}$$

where $E_D(w)$ is defined by

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

and $E_w(w) = \frac{1}{2} w^T w$

Maximizing the log posterior w.r.t. w gives the maximum posterior (MAP) estimator of w.

Maximizing the log posterior is equivalent to minimizing the sum of squares error function E_D plus a quadratic regularization term E_W.

Problem 5:

Below are the performance metrics to predict the cooling load using the given Energy Efficiency Dataset (ENB2012_data.xlsx) as the training data for the following three models based on mean squared error (MSE):

1. Lasso regression
2. Ridge Regression
3. Elastic Net regression

Model	Mean Squared Error
Lasso regression	15.162
Ridge regression	10.891
Elastic Net regression	19.198

As we can see, using the 5-fold cross validation to compare model performance based on mean squared error (MSE), the Ridge regression model has the lowest mean squared error value. This implies that it is the best model to predict the cooling load using the given dataset.