Max Shi

CS 559: Machine Learning

Homework 1

2/9/2022

Problem 1

1) $P(CS\ student) = P(S1) * P(CS|S1) + P(S2) * P(CS|S2) + P(S3) * P(CS|S3)$
$$= 0.2 * \frac{6}{20} + 0.2 * \frac{10}{20} + 0.6 * \frac{6}{20}$$
$$= 0.34$$

2) $P(S3|STAT) = \frac{P(STAT|S3) * P(S3)}{P(STAT)}$
$$P(STAT) = P(S1) * P(STAT|S1) + P(S2) * P(STAT|S2) + P(S3) * P(STAT|S3)$$
$$= 0.2 * \frac{8}{20} + 0.2 * \frac{10}{20} + 0.6 * \frac{6}{20} = 0.36$$
$$P(STAT|S3) = \frac{6}{20}$$
$$P(S3) = 0.6$$
$$P(S3|STAT) = \frac{0.6 * \frac{6}{20}}{0.36} = 0.5$$

Problem 2

1) Because this is a normal distribution, the likelihood function is $P(x|\mu, \sigma^2) = \prod_{n=1}^{N} N\,(x_n|\mu, \sigma^2)$
Which, expanded, is $P(x|\mu, \sigma^2) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{1}{2}\right)\left(\frac{x_n-\mu}{\sigma}\right)^2}$

2) Taking the log of that function, we obtain:
$$\ln\big(P(x|\mu, \sigma^2)\big) = \prod_{n=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{1}{2}\right)\left(\frac{x_n-\mu}{\sigma}\right)^2}\right)$$
$$= \sum_{n=1}^{N} (\ln\left((2\pi\sigma^2)^{-\frac{1}{2}}\right) + \ln\left(e^{\left(-\frac{1}{2}\right)\left(\frac{x_n-\mu}{\sigma}\right)^2}\right))$$
$$= \sum_{n=1}^{N} \ln\left((2\pi\sigma^2)^{-\frac{1}{2}}\right) + \sum_{n=1}^{N} \left(-\frac{1}{2}\right)\left(\frac{x_n - \mu}{\sigma}\right)^2 * \ln(e)$$
$$= -\frac{N}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2$$
To calculate both parameters, we take the derivative with respect to mu and sigma, and set them equal to zero.
$$\frac{d}{d\mu}\left(-\frac{N}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2\right) = \frac{1}{2\sigma^2} * -1 \sum_{n=1}^{N} 2(x_n - \mu)$$
$$0 = \frac{1}{\sigma^2}(\sum_{n=1}^{N} x_n - N * \mu)$$

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

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

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

$$-\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=1}^{N}(x_n-\mu)^2 = \frac{1}{2\sigma^2}\left(-N + \frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right)$$

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

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

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

Now, using the data points from the problem, we obtain:

$$\sum_{n=1}^{N} x_n = 1409$$

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n = \frac{1}{10}(1409) = 140.9$$

$$\sigma_{ML}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2 = \frac{1}{10}\sum_{n=1}^{N}(x_n-140.9)^2 = \frac{1}{10} * 3466.9 = 346.69$$

Problem 3

1) The likelihood function of this dataset is $P(x) = \prod_{n=1}^{10} P(X_n) = \left(\frac{2q}{3}\right)^2\left(\frac{q}{3}\right)^3\left(\frac{2(1-q)}{3}\right)^3\left(\frac{1-q}{3}\right)^2$

$$P(x) = \frac{4q^5 * 8(1-q)^5}{3^{10}}$$

2) Taking the log of this function, we obtain:

$\log P(x) = \log(4q^5 * 8(1-q)^5) - \log(3^{10})$

$= \log 4 + \log(q^5) + \log 8 + \log(1-q)^5 - \log(3^{10})$

$= \log 4 + 5\log(q) + \log 8 + 5\log(1-q) - \log(3^{10})$

$\frac{d}{dq}(\log 4 + 5\log(q) + \log 8 + 5\log(1-q) - \log(3^{10})) = \frac{5}{q} - \frac{5}{1-q}$

$0 = \frac{5}{q} - \frac{5}{1-q} = \frac{5(1-q)-5q}{q(1-q)} = \frac{5-10q}{q(1-q)}$

$10q = 5 \rightarrow q = \frac{1}{2}$

Problem 4

Using Bayes theorem, the posterior distribution for w, $p(w|x, y, \alpha, \beta)$, is proportional to $p(y|x, w, \beta) * p(w|\alpha)$, where:

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

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

Combining these two together we obtain:

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

Taking the negative log of this whole expression, assuming that M+1 ≈ N, we obtain:

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

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

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

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

Canceling out terms yields:

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

As the parameters alpha and beta do not change, minimizing this function involves the second and fourth terms in this expression. Thus, this is equal to minimizing the regularized sum of squares error function, which is:

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