Introduction to Machine Learning - HW1

Professors: Abolghasemi & Erabi
Student: Mohamad Mahdi Samadi
Student ID: 810101465

Q3.

Given a dataset $D = \{\{x_1, y_1\}, \{x_2, y_2\}, ..., \{x_n, y_n\}\}$ in which $y_i = wx_i + \epsilon$ where ϵ comes from a Normal distribution with $m = 0$ and $\delta^2 = 1$ (Standard normal distribution), prove that maximizing the log-likelihood function is equivalent to minimizing sum of squared residuals.

$$argmax_w\, log(P(D|W) = argmin_w \sum_{i=1}^{n} (y_i - wx_i)^2$$

Given the Gaussian noise. The probability density function for each y_i is:

$$Normal\, pdf = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2} \rightarrow pdf\, \epsilon = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}\epsilon^2} \rightarrow P(y_i|x_i, w) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(y_i - wx_i)^2}$$

$$P(D|W) = \prod_{i=1}^{n} P(y_i, x_i, w) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp(\frac{-1}{2}(y_i - wx_i)^2)$$

After taking the logarithm, the log-likelihood function is:

$$log\, P(D|W) = log[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(y_i - wx_i)^2} = \sum_{i=1}^{n} log[\frac{1}{\sqrt{2\pi}} e^{(\frac{-1}{2}(y_i - wx_i)^2)}]$$

$$log\, P(D|W) = \sum_{i=1}^{n} log[\frac{1}{\sqrt{2\pi}}] + log[e^{\frac{-1}{2}(y_i - wx_i)^2)}] nlog[\frac{1}{\sqrt{2\pi}}] - \frac{1}{2}\sum_{i=1}^{n}(y_i - wx_i)^2$$

In order to maximize the log-likelihood function, we need to maximize the $- \sum_{i=1}^{n} (y_i - wx_i)^2$ term because the $\frac{1}{2}$ coefficient is constant. Also the $nlog[\frac{1}{\sqrt{2\pi}}]$ term is constant and the only thing changing with w is the $- \sum_{i=1}^{n} (y_i - wx_i)^2$ term.

$$argmax_w P(D|W) = argmax_w[- \sum_{i=1}^{n} (y_i - wx_i)^2] \equiv argmin_w[\sum_{i=1}^{n} (y_i - wx_i)^2] = argmin_w SSR$$