FOML Assignment1

SADINENI ABHINAY - CS21BTECH11055
NIMMALA AVINASH - CS21BTECH11039
Group 64

October 2023

Question 3

Let us consider general expression for error of model $\epsilon_i = y_i - w^T.\phi(x_i)$ where $\epsilon_i \sim N(0, \sigma_i^2)$

(a) (i) **Likelihood:** Let us assume the errors are part of Gaussian distribution with each distribution having variance σ_i^2 So $t_i \sim w^T.\phi(x_i) + N(0, \sigma_i^2)$ from that we get $t_i | x_i, \sigma_i^2 \sim N(w^T.\phi(x_i), \sigma_i^2)$

$$L = P(t_i | x_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{\frac{-1}{2\sigma_i^2}(w^T.\phi(x_i) - t_i)^2} \tag{1}$$

(ii) **Prior:** Lets assume that w has Gaussian distribution

$$Prior = P(w) = \frac{1}{\sqrt{2\pi\tau^2}} e^{\frac{-1}{2\tau^2} w^T w} \tag{2}$$

(b) (i) **ML estimation:** For ML estimation we will find w for which maximum the likelihood function , the objective function for negative log-likelihood:

$$NLL(w) = -log(L(w)) \tag{3}$$

$$= -\log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2\sigma_i^2}(w^T.\phi(x_i) - t_i)^2}\right) \tag{4}$$

$$= \sum_{i=1}^{n}\left(-\log\left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right) + \frac{1}{2\sigma_i^2}(w^T.\phi(x_i) - t_i)^2\right) \tag{5}$$

$$= \frac{1}{2}\sum_{i=1}^{n}\left(log(2\pi\sigma_i^2) + \frac{1}{\sigma_i^2}(w^T.\phi(x_i) - t_i)^2\right) \tag{6}$$

$$\hat{w}_{ML} = minarg_w \left\{\frac{1}{2}\sum_{i=1}^{n}\left(log(2\pi\sigma_i^2) + \frac{1}{\sigma_i^2}(w^T.\phi(x_i) - t_i)^2\right)\right\} \tag{7}$$

(ii) **MAP estimation:** For MAP estimate we have to maximum the negative posterior $P(w|t, \Phi, \sigma^2) \propto P(t|\Phi, w, \sigma^2) \cdot P(w)$

$$\begin{aligned}
MAPF &= -log(P(w|t, \Phi, \sigma^2)) \\
&= -log(P(t|\Phi, w, \sigma^2)) - log(P(w)) + constant \\
\hat{w}_{MAP} &= minarg_w\{-log(P(t|\Phi, w, \sigma^2)) - log(P(w))\} \\
&= minarg_w \left\{\frac{1}{2}\sum_{i=1}^{n}\left(log(2\pi\sigma_i^2) + \frac{1}{\sigma_i^2}(w^T.\phi(x_i) - t_i)^2\right) + \frac{1}{2\tau^2} w^T w\right\}
\end{aligned}$$

(c) Lets us try to find the ML estimate and compare it weighted least squares solution

$$\nabla NLL(w) = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(w^T.\phi(x_i) - t_i)(\phi(x_i))^T \tag{8}$$

Setting this gradient to zero gives

$$0 = \sum_{i=1}^{N} t_n \frac{1}{\sigma_i^2} \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} - \mathbf{w}^{\mathrm{T}} \left(\sum_{i=1}^{N} \boldsymbol{\phi}(\mathbf{x}_n) \frac{1}{\sigma_i^2} \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \right). \tag{9}$$

Solving for $\mathbf{w}$ we obtain

$$\mathbf{w} = \left(\boldsymbol{\Phi}^T \Sigma^{-1} \boldsymbol{\Phi} \right)^{-1} \boldsymbol{\Phi}^T \Sigma^{-1} \mathbf{t}$$

Σ^{-1} is a diagonal matrix with weights $\frac{1}{\sigma_i^2}$ as diagonals.Here Φ is an N×M matrix, called the design matrix, whose elements are given by $\Phi_{nj} = \phi_j(x_n)$, so that Now lets see the weighted least squares

$$E_D(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} r_n \left\{ t_n - \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \right\}^2 \tag{10}$$

Similarly for WLS we will get solution

$$\mathbf{w} = \left(\boldsymbol{\Phi}^T R \boldsymbol{\Phi} \right)^{-1} \boldsymbol{\Phi}^T R \mathbf{t} \tag{11}$$

R is a diagonal matrix with weights r_i as diagonals. Upon comparing we can conclude $\Sigma^{-1} = R$ that implies $r_i = \frac{1}{\sigma_i^2}$ and proved that $r_i > 0$ The expression for w for minimizing the least squares is

$$\boxed{\mathbf{w} = \left(\boldsymbol{\Phi}^T R \boldsymbol{\Phi} \right)^{-1} \boldsymbol{\Phi}^T R \mathbf{t}} \tag{12}$$

where R is diagonal matrix with $r_i = \frac{1}{\sigma_i^2}$