QUESTION 3

ABHAY SHANKAR K: CS21BTECH11001 & KARTHEEK TAMMANA: CS21BTECH11028

(I) Derive the expression of likelihood and prior for a heteroscedastic setting for a single data point with input $\mathbf{x_n}$ and output t_n.

Consider the following formula for the target variable:

$$t = y(\mathbf{x}, \mathbf{w}) + \epsilon(\mathbf{x})$$

where ϵ is a zero mean Gaussian random variable with precision $\beta(\mathbf{x})$ and y is a deterministic function. Due to heteroscedasticity, the Gaussian noise is dependent on the input $\mathbf{x}$.

Due to the properties of Gaussian distribution, t is also normal, with its distrbution i.e. the likelihood function given by:

$$p\left(t_n|\mathbf{x_n}, \mathbf{w}, \boldsymbol{\beta}\right) = \sqrt{\frac{\beta_n}{2\pi}} \exp\left\{-\frac{\beta_n}{2}\left(t_n - y(\mathbf{x_n}, \mathbf{w})\right)^2\right\}$$

The prior is given by:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{m_0}, \mathbf{S_0})$$

Here, m_0 is the mean of the prior distribution and S_0 is the covariance matrix.

In a homoscedastic setting, we may assume that $S_0 = \beta^{-1}I$, where I is the identity matrix. However, in a heteroscedastic setting, we obviously cannot, and instead S_0 will be an arbitrary diagonal matrix (because the variables are independent, but not identically distributed).

(II) Provide the expression for the objective function that you will consider for the ML and MAP estimation of the parameters considering a data set of size N.

With $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, the objective function for ML estimation is given by:

$$\begin{aligned} p\left(\mathbf{t}|\mathbf{X}, \mathbf{w}, \boldsymbol{\beta}\right) &= \prod_{n=1}^{N} \sqrt{\frac{\beta_n}{2\pi}} \exp\left\{-\frac{\beta_n}{2}\left(t_n - y(\mathbf{x_n}, \mathbf{w})\right)^2\right\} && (1) \\ &= \mathcal{N}(\mathbf{t}|\mathbf{\Phi^T w}, \mathbf{R}) && (2) \\ \equiv \ln p &= \text{constant} + \frac{1}{2}\sum_{n=1}^{N} \beta_n \left(t_n - y(\mathbf{x_n}, \mathbf{w})\right)^2 && (3) \end{aligned}$$

Defining the matrix $\mathbf{R}$ with $R_{ii} = \beta_i$.

Given

- $$p(\mathbf{w}) = \mathcal{N}(\mathbf{m_0}, \mathbf{S_0})$$
- $$p(t|w) = \mathcal{N}(\mathbf{\Phi^T w}, \mathbf{R^{-1}})$$

The objective function for MAP estimation is given by the normal distrbution with mean $\boldsymbol{\mu}$ and variance $\mathbf{\Sigma}$ where

$\mathbf{\Sigma} = (\mathbf{S_0}^{-1} + \mathbf{\Phi^T R \Phi})^{-1}$ $\boldsymbol{\mu} = \mathbf{\Sigma}(\mathbf{S_0}^{-1}\mathbf{m_0} + \mathbf{\Phi^T R t})$

Date: October 7, 2023.

Thus, we have

$$p(\mathbf{w}|\mathbf{t}) = \mathcal{N}\left(\mathbf{w}|(\mathbf{S_0}^{-1} + \mathbf{\Phi^T R \Phi})^{-1}(\mathbf{S_0}^{-1}\mathbf{m_0} + \mathbf{\Phi^T R t}), (\mathbf{S_0}^{-1} + \mathbf{\Phi^T R \Phi})^{-1}\right)$$

(III) Show

$$E_{\mathcal{D}} = \sum_{n=1}^{N} r_n \left\{t_n - \mathbf{w^T}\phi(\mathbf{x_n})\right\}^2$$

and find $\mathbf{w}$ that minimizes $E_{\mathcal{D}}$.

Reframing 3 in the style of the homoscedastic case, we have

$$-\ln p(\mathcal{D}|\mathbf{w}) = \frac{N}{2}\ln\beta_n - \frac{N}{2}\ln(2\pi) - E_{\mathcal{D}}$$

whence it is evident that

$$E_{\mathcal{D}} = \frac{1}{2}\sum_{n=1}^{N} \beta_n \left\{t_n - \mathbf{w^T}\phi(\mathbf{x_n})\right\}^2$$

Setting $r_n = \frac{\beta_n}{2}$, we obtain the desired equation.

Now, we may obtain the $\mathbf{w}$ that minimizes $E_{\mathcal{D}}$ by differentiating **(III)** and setting the derivative to zero like so:

$$\sum_{n=1}^{N} \beta_n \left\{t_n - \mathbf{w^T}\phi(\mathbf{x_n})\right\} \phi\left(\mathbf{x_n}\right)^{\mathbf{T}} = 0$$

Whence

$$\sum_{n=1}^{N} t_n \beta_n \phi_n^T = \mathbf{w^T} \sum_{n=1}^{N} \beta_n \phi\left(\mathbf{x_n}\right)\phi\left(\mathbf{x_n}\right)^{\mathbf{T}}$$

Let the matrix $\mathbf{R}$ be a diagonal matrix with $R_{ii} = \beta_i$. Then, we have

$$\Phi^T R\mathbf{t} = (\Phi^T R \Phi)w \tag{4}$$

$$\implies \mathbf{w_{ML}} = (\mathbf{\Phi^T R \Phi})^{-1}\mathbf{\Phi^T R t} \tag{5}$$