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:

$$t = \mathbf{w^T}\boldsymbol{\phi}(\mathbf{x}) + \epsilon(\mathbf{x})$$

where ϵ is a zero mean Gaussian random variable with precision $\beta(\mathbf{x})$ and $\boldsymbol{\phi}$ 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:

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

So, we can say

$$p\left(t_n|\mathbf{x_n}, \mathbf{w}, \boldsymbol{\beta}\right) = \mathcal{N}(t_n|\mathbf{w^T}\boldsymbol{\phi}(\mathbf{x}_n), \beta^{-1})$$

The prior is given by:

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

Here, $\mathbf{m_0}$ is the mean of the prior distribution and $\mathbf{S_0}$ is the covariance matrix.

(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.

(a) To express this more succinctly, we define the design matrix

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\phi}(\mathbf{x}_1) \\ \vdots \\ \boldsymbol{\phi}(\mathbf{x}_n) \end{pmatrix}$$

the data set $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and the weighing matrix $\mathbf{R}$ with $R_{ii} = \beta_i$.

Thus, the objective function for ML estimation is given by:

$$\begin{aligned} p\left(\mathbf{t}|\mathbf{X}, \mathbf{w}, \boldsymbol{\beta}\right) &= \prod_{n=1}^{N} p\left(t_n|\mathbf{x_n}, \mathbf{w}, \boldsymbol{\beta}\right) \\ &= \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w^T}\boldsymbol{\phi}(\mathbf{x}_n), \beta_n^{-1}) \\ &= \exp\left\{-\frac{1}{2}\sum_{n=1}^{N}\beta_n\left(t_n - \mathbf{w^T}\boldsymbol{\phi}(\mathbf{x}_n)\right)^2\right\} \\ &= \exp\left\{-\frac{1}{2}\left(\mathbf{t} - \boldsymbol{\Phi^T w}\right)^T \mathbf{R}\left(\mathbf{t} - \boldsymbol{\Phi^T w}\right)\right\} \\ &= \mathcal{N}(\mathbf{t}|\boldsymbol{\Phi}\mathbf{w}, \mathbf{R^{-1}}) \end{aligned} \tag{1}$$

Date: October 8, 2023.

(b) Given

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

and

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

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

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

This can be derived using the quantity $\mathbf{z} = \begin{pmatrix} \mathbf{w} \\ \mathbf{t} \end{pmatrix}$ as follows:

$$\begin{aligned} \ln p(\mathbf{z}) &= \ln p(\mathbf{w}) + \ln p(\mathbf{t}|\mathbf{w}) \\ &= -\frac{1}{2}(\mathbf{w} - \mathbf{m_0})^T \mathbf{S_0}^{-1}(\mathbf{w} - \mathbf{m}_0) \\ &\quad - \frac{1}{2}(\mathbf{t} - \mathbf{\Phi w})^T \mathbf{R}(\mathbf{t} - \mathbf{\Phi w}) + \text{constant} \end{aligned} \tag{2}$$

Furthermore, due to linearity of expectation, we have

$$\mathbb{E}[\mathbf{z}] = \begin{pmatrix} \mathbb{E}[\mathbf{w}] \\ \mathbb{E}[\mathbf{t}] \end{pmatrix}$$

from which

$$\text{cov}[\mathbf{z}] = \begin{pmatrix} \text{var}[\mathbf{w}] & \text{cov}[\mathbf{w}, \mathbf{t}] \\ \text{cov}[\mathbf{t}, \mathbf{w}] & \text{var}[\mathbf{t}] \end{pmatrix}$$

From 2, it is clear that $p(\mathbf{z})$ is a Gaussian distribution. Now we complete the square.

To find the covariance of $\mathbf{w}|\mathbf{t}$, we consider the single term of second order in $\mathbf{w}$ from 2:

$$\frac{1}{2}\mathbf{w^T}\mathbf{\Sigma}^{-1}\mathbf{w} = \frac{1}{2}\mathbf{w^T}(\mathbf{S_0}^{-1} + \mathbf{\Phi R \Phi^T})\mathbf{w}$$

We treat $\mathbf{t}$ as a constant.

Thus, the covariance is given by

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

Similarly, we may obtain $\boldsymbol{\mu}$ using the terms of 2 of first order in $\mathbf{w}$. We have

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

which yields

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

(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 1 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

$$\Phi^T R\mathbf{t} = (\Phi^T R \Phi) w \tag{3}$$

$$\implies \mathbf{w_{ML}} = (\mathbf{\Phi^T R \Phi})^{-1} \mathbf{\Phi^T R t} \tag{4}$$