QUESTION 3

ABHAY SHANKAR K: CS21BTECH11001 & KARTHEEK TAMMANA: CS21BTECH11028

(1) Derive the expression of likelihood and prior for a heteroscedastic setting for a single data point with input 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 $\beta(\mathbf{x})$.

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:

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

The prior is given by:

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

which is identical to the homoscedastic setting, because the prior distribution of weights is not affected by the variance of the input.

(2) 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.

The objective function for ML estimation is given by:

$$p\left(\mathcal{D}|\mathbf{w}\right) = \prod_{n=1}^{N} \sqrt{\frac{\beta_n}{2\pi}} \exp\left\{-\frac{\beta_n}{2}\left(t_n - y(x_n, \mathbf{w})\right)^2\right\} \tag{1}$$

$$\equiv -\ln p(\mathcal{D}|\mathbf{w}) = \frac{N}{2}\ln\beta_n - \frac{N}{2}\ln(2\pi) - \frac{1}{2}\sum_{n=1}^{N}\beta_n\left(t_n - y(x_n, \mathbf{w})\right)^2 \tag{2}$$

which is equivalent to minimizing the weighted sum of squared errors.

The objective function for MAP estimation, $p\left(\mathbf{w}|\mathcal{D}\right) = p\left(\mathcal{D}|\mathbf{w}\right)p\left(\mathbf{w}\right)$ is given by:

Date: October 4, 2023.

(3)
$$p\left(\mathcal{D}|\mathbf{w}\right)p\left(\mathbf{w}\right)=\prod_{n=1}^{N}\sqrt{\frac{\beta_n}{2\pi}}\exp\left\{-\frac{\beta_n}{2}\left(t_n-y(x_n,\mathbf{w})\right)^2\right\}\mathcal{N}(\mathbf{w}|\mathbf{m_0},\mathbf{S_0})$$
(4)
$$\equiv-\ln p(\mathcal{D}|\mathbf{w})-\ln p(\mathbf{w})=\frac{N}{2}\ln\beta_n-\frac{N}{2}\ln(2\pi)-\frac{1}{2}\sum_{n=1}^{N}\beta_n\left(t_n-y(x_n,\mathbf{w})\right)^2+\frac{1}{2}(\mathbf{w}-\mathbf{m_0})^T\mathbf{S_0}^{-1}(\mathbf{w}-\mathbf{m_0})$$

(3) 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 2 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 3 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
$$\sum_{n=1}^{N}t_n\beta_n\phi_n^T=\mathbf{w^T}\sum_{n=1}^{N}\beta_n\phi\left(\mathbf{x_n}\right)\phi\left(\mathbf{x_n}\right)^{\mathbf{T}}$$

and thus,
$$\mathbf{w_{ML}}=\mathcal{A}^{\dagger}\mathbf{t}$$
where $\mathcal{A}$ is the design matrix with $\mathcal{A}_{ij}=\beta_i\phi_j(\mathbf{x_i})$, $\mathbf{t}$ is the vector of targets, and $\mathcal{A}^{\dagger}$ is the Moore-Pensore Pseudoinverse.