

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 $\beta(\mathbf{x})$. 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 distribution i.e. the likelihood function given by:

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

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

$$(1) \quad p(\mathbf{t} | \mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^N \sqrt{\frac{\beta_n}{2\pi}} \exp \left\{ -\frac{\beta_n}{2} (t_n - y(\mathbf{x}_n, \mathbf{w}))^2 \right\}$$

$$(2) \quad = \mathcal{N}(\mathbf{t} | \Phi^T \mathbf{w}, \mathbf{R})$$

$$(3) \quad \equiv \ln p = \text{constant} + \frac{1}{2} \sum_{n=1}^N \beta_n (t_n - y(\mathbf{x}_n, \mathbf{w}))^2$$

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

Given

•

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w} | \mathbf{m}_0, \mathbf{S}_0)$$

•

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

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

• $\boldsymbol{\Sigma} = (\mathbf{S}_0^{-1} + \Phi^T \mathbf{R} \Phi)^{-1}$

• $\boldsymbol{\mu} = \boldsymbol{\Sigma}(\mathbf{S}_0^{-1} \mathbf{m}_0 + \Phi^T \mathbf{R} \mathbf{t})$

Thus, we have

$$p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w} | (\mathbf{S}_0^{-1} + \Phi^T \mathbf{R} \Phi)^{-1} (\mathbf{S}_0^{-1} \mathbf{m}_0 + \Phi^T \mathbf{R} \mathbf{t}), (\mathbf{S}_0^{-1} + \Phi^T \mathbf{R} \Phi)^{-1})$$

(III) Show

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

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

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

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:

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

Whence

$$\sum_{n=1}^N t_n \beta_n \phi_n^T = \mathbf{w}^T \sum_{n=1}^N \beta_n \phi(\mathbf{x}_n) \phi(\mathbf{x}_n)^T$$

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

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

$$(5) \quad \implies \mathbf{w}_{\text{ML}} = (\Phi^T \mathbf{R} \Phi)^{-1} \Phi^T \mathbf{R} \mathbf{t}$$