CS3390 Assignment 1
Problem 3

Gautam Singh (CS21BTECH11018)
Jaswanth Beere (BM21BTECH11007)

Contents

This document describes a solution for weighted least squares in a heteroscedastic setting.

1 Likelihood and Prior

We assume that the response variable t_n is approximated by a Gaussian distribution with mean $\mathbf{w}^\top \mathbf{x_n}$ and variance $\sigma^2(\mathbf{x}_n)$ in a heteroscedastic setting. Thus, the likelihood for a single datapoint $(\mathbf{x}_n, t_n)$ is

$$\begin{aligned}\Pr\left(t_n|\mathbf{x}_n; \mathbf{w}, \sigma^2(\mathbf{x}_n)\right) &= \mathcal{N}\left(t_n|\mathbf{w}^\top\mathbf{x}_n, \sigma^2(\mathbf{x}_n)\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2(\mathbf{x}_n)}}\exp\left(-\frac{(t_n - \mathbf{w}^\top\mathbf{x}_n)}{2\sigma^2(\mathbf{x}_n)}\right) \quad (1)\end{aligned}$$

For the prior we assume that $\mathbf{w}$ is drawn from a multivariate Gaussian distribution, that is,

$$\begin{aligned}&\Pr(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\right). \quad (2)\end{aligned}$$

where D is the dimension of $\mathbf{w}$.

2 ML and MAP Objective Functions

Considering a dataset $\mathcal{D}$ of N independent and identically distributed samples $\left(\mathbf{x}_i, \sigma^2(\mathbf{x}_i), t_i\right)$, $1 \le i \le N$, the ML objective using (1) is

$$\begin{aligned}\hat{\mathbf{w}}_{ML} &= \underset{\mathbf{w}}{\operatorname{argmax}} \log\Pr(\mathcal{D}|\mathbf{w}) &(3)\\ &= \underset{\mathbf{w}}{\operatorname{argmax}} \log\prod_{i=1}^{N}\Pr\left(t_i|\mathbf{x}_i; \mathbf{w}, \sigma^2(\mathbf{x}_i)\right) &(4)\\ &= \underset{\mathbf{w}}{\operatorname{argmax}} \sum_{i=1}^{N}\log\Pr\left(t_i|\mathbf{x}_i; \mathbf{w}, \sigma^2(\mathbf{x}_i)\right) &(5)\\ &= \underset{\mathbf{w}}{\operatorname{argmax}} \sum_{i=1}^{N} -\frac{1}{2}\log\left(2\pi\sigma^2(\mathbf{x}_i)\right) - \frac{(t_i - \mathbf{w}^\top\mathbf{x}_i)^2}{2\sigma^2(\mathbf{x}_i)} &(6)\\ &= \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{2}\sum_{i=1}^{N}\frac{(t_i - \mathbf{w}^\top\mathbf{x}_i)^2}{\sigma^2(\mathbf{x}_i)}. &(7)\end{aligned}$$

On the other hand, using (1), (2), and (7), the MAP objective becomes

$$\begin{aligned}\hat{\mathbf{w}}_{MAP} &= \underset{\mathbf{w}}{\operatorname{argmax}} \log\Pr(\mathbf{w}|\mathcal{D}) &(8)\\ &= \underset{\mathbf{w}}{\operatorname{argmax}} \log\left(\frac{\Pr(\mathcal{D}|\mathbf{w})\Pr(\mathbf{w})}{\Pr(\mathcal{D})}\right) &(9)\\ &= \underset{\mathbf{w}}{\operatorname{argmax}} \log\Pr(\mathcal{D}|\mathbf{w})\Pr(\mathbf{w}) &(10)\\ &= \underset{\mathbf{w}}{\operatorname{argmax}} \log\Pr(\mathcal{D}|\mathbf{w}) + \log\Pr(\mathbf{w}) &(11)\\ &= \underset{\mathbf{w}}{\operatorname{argmin}} \left(\frac{1}{2}\sum_{i=1}^{N}\frac{(t_i - \mathbf{w}^\top\mathbf{x}_i)^2}{\sigma^2(\mathbf{x}_i)} + \frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\right). &(12)\end{aligned}$$

Note that taking $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \lambda^{-1}\mathbf{I}$ gives

$$\hat{\mathbf{w}}_{\mathbf{MAP}} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{2}\sum_{i=1}^{N}\frac{(t_i - \mathbf{w}^\top\mathbf{x}_i)^2}{\sigma^2(\mathbf{x}_i)} + \frac{\lambda}{2}\|\mathbf{w}\|^2 \quad (13)$$

which is the objective function for *regularized weighted least squares* method.

3 Maximum Likelihood Solution

Defining

$$r_n = \frac{1}{\sigma^2(\mathbf{x}_n)} \tag{14}$$

$$\boldsymbol{\phi}(\mathbf{x}_n) = \mathbf{x}_n \tag{15}$$

then from (7), the sum of squares error function becomes

$$E_{\mathcal{D}}(\mathbf{w}) \triangleq \frac{1}{2}\sum_{i=1}^{N} r_n \left\{t_n - \mathbf{w}^\top \boldsymbol{\phi}(\mathbf{x}_n)\right\}^2 . \tag{16}$$

where $r_n > 0 \ \forall \ 1 \leq n \leq N$.
Notice that (16) can be written as

$$E_{\mathcal{D}}(\mathbf{w}) = \frac{1}{2}\left(\mathbf{y} - \mathbf{X}^\top \mathbf{w}\right)^\top \mathbf{R}\left(\mathbf{y} - \mathbf{X}^\top \mathbf{w}\right). \tag{17}$$

where

$$\mathbf{y} \triangleq \begin{pmatrix} y_1 & y_2 & \dots & y_N \end{pmatrix}^\top \tag{18}$$

$$\mathbf{X} \triangleq \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \dots & \mathbf{X}_n \end{pmatrix} \tag{19}$$

$$\mathbf{R} \triangleq \operatorname{diag}\begin{pmatrix} r_1 & r_2 & \dots & r_n \end{pmatrix}. \tag{20}$$

Taking the gradient of (17) and setting it to $\mathbf{0}$ gives (where we note that $\mathbf{R}$ is symmetric.)

$$-\left(y - \mathbf{X}^\top \mathbf{w}\right)^\top \mathbf{R}\mathbf{X}^\top = \mathbf{0} \tag{21}$$

$$\mathbf{X}\mathbf{R}\mathbf{X}^\top \mathbf{w} = \mathbf{X}\mathbf{R}\mathbf{y} \tag{22}$$

$$\implies \hat{\mathbf{w}}_{ML} = \left(\mathbf{X}\mathbf{R}\mathbf{X}^\top\right)^{-1} \mathbf{X}\mathbf{R}\mathbf{y} \tag{23}$$