CMPUT 466 Assignment 5

Arun Woosaree

October 20, 2021

Problem 1

	Larger C	Smaller C
Model capacity	large	small
Overfitting/Underfitting	overfitting	underfitting
bias variance	low bias / high variance	high bias / low variance

Problem 2, 3

$t^{(m)} \sim \mathcal{N}(w x^{(m)}, \sigma_\varepsilon^2) \Rightarrow \mathbf{t} \sim \mathcal{N}(\mathbf{X}w, \Lambda)$

$w \sim \mathcal{N}(0, \sigma_w^2)$

$$\Lambda = \begin{pmatrix} \sigma_\varepsilon^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_\varepsilon^2 \end{pmatrix}$$

$$S_0 = \begin{pmatrix} \sigma_w^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_w^2 \end{pmatrix}$$

$m_0 = \mathbf{0}$

$$p(w|D) \underset{w}{\propto} p(w)\, p(D|w)$$

$$\propto \exp\left\{-\frac{1}{2}(w - m_0)^T \mathbf{S}_0^{-1} (w - m_0)\right\} \times \exp\left\{-\frac{1}{2}(\mathbf{t} - \mathbf{X}w)^T \Lambda^{-1} (\mathbf{t} - \mathbf{X}w)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(w - m_0)^T S_0^{-1} (w - m_0) - \frac{1}{2}(t - Xw)^T \Lambda^{-1} (t - Xw)\right\}$$

complete the square: $\exp\left(-\frac{1}{2} w^T\left(S_0^{-1} + \frac{1}{\sigma_\varepsilon^2} X^T X\right) w + (\cdots) w + \text{const}\right)$

$$\propto \exp\left\{-\frac{1}{2}(w - m_N)^T S_N^{-1} (w - m_N)\right\}$$

matrix form of

$$\boxed{\begin{aligned} \text{where } m_N &= S_N\left(S_0^{-1} m_0 + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{t}\right) \quad \mu_{post} \\ S_N &= \left(\mathbf{S}_0^{-1} + \frac{1}{\sigma^2} X^T \mathbf{X}\right)^{-1} \quad \sigma_{post}^2 \end{aligned}}$$

this posterior is a Gaussian distribution of the form:

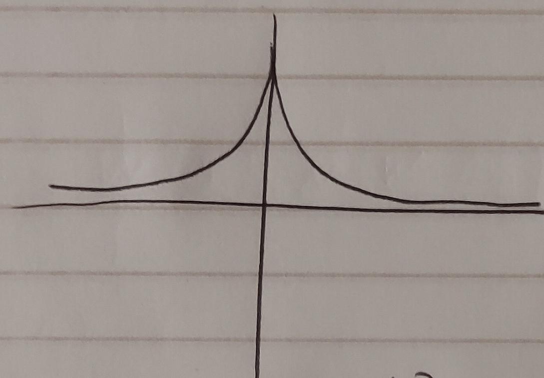
$$\mathcal{N}(\mathbf{w}; m_N, S_N)$$

$$J_{\ell_1 MSE}(w) = \frac{1}{2M} \| Xw - t \|_2^2 + \frac{\lambda}{2J} \|w\|_1$$

$\Longleftrightarrow$ MAP Laplace prior

$\Longleftrightarrow$ $\underset{w}{\text{minimize}} \ \frac{1}{2n} \|Xw - t\|_2^2$ subject to $\|w\|_1 \leq c$

The laplace prior distribution looks like this:



$\sim \text{Laplace}(0, b)$

$$p(\theta) = \frac{1}{2b} \exp\left\{ \frac{-|\theta|}{b} \right\}$$