Machine Learning Worksheet 05

Shang-Hsin Yu – 03681048 – shanghsin.yu@tum.de

Problem 1

$$-\frac{1}{2}\sum_{n=1}^{N} t_n[W^T\phi(x_n) - z_n]$$

$$= -\frac{1}{2}[T^{0.5}(\Phi W - Z)]^T[T^{0.5}(\Phi W - Z)]$$

let Φ become a new matrix $\tilde{\Phi}$ where every element in the ith row of Φ is multiplied by t_i.

$$= -\frac{1}{2}(\tilde{\Phi}W - Z)^T(\tilde{\Phi}W - Z)]$$

and we we know derivative with respect to W is

$$-\tilde{\Phi}^T(\tilde{\Phi}W - Z)$$

$$W_{optimal} = (\tilde{\Phi}^T\tilde{\Phi})^{-1}\tilde{\Phi}^T Z$$

The variance of the noise doesn't really change, and this kind of weighting is basically adding extra (partial) points or taking away (partial) points on the data set. A weighting of 2 will have the same effect on the result as if the specific data point is added twice.

Problem 2

Assume the original matrix Φ has n rows and m columns such that n is the size of data set and m is the number of features. Vector W obviously has m elements as well. Z is the target vector with n elements. Setting the number p to m, we stuck a m by m scalar matrix with $\sqrt{\lambda}$ at the diagonals and 0 everywhere else on top of the Φ matrix, and stucking m more 0s on top of vector Z. This way we get:

$$\frac{1}{2}(\sum_{n=1}^{N}(W^T\phi(x_n))^2 + \lambda W^T W)$$

which is exactly ridge regression.

Problem 3

$$p(Z \mid W, \beta) = \mathcal{N}(Z \mid \Phi W, \beta^-1)$$

$$p(W, \beta) = \mathcal{N}(W \mid M_0, \beta^{-1}S_0)Gam(\beta \mid a_0, b_0)$$

$$p(W, \beta \mid Z) \propto p(Z \mid W, \beta) \cdot p(W, \beta)$$

$$\propto e^{-\frac{\beta}{2}(Z-\Phi W)^T(Z-\Phi W)} \cdot e^{-\frac{\beta}{2}(W-M_0)^T S_0^{-1}(W-M_0)} \cdot Gam(\beta \mid a_0, b_0)$$

Where the two Gaussian seems to merge into one and the Gamma remain untouched. The Gaussian exponential part becomes

$$= -\frac{1}{2}W^T(\beta\Phi^T\Phi + \beta S_0^{-1})W - W^T(\beta\Phi^T Z + \beta S_0^{-1}M_0) - (\beta\Phi^T Z + \beta S_0^{-1}M_0)W + constant$$

Which is the new Gaussian

$$= -\frac{1}{2}(W - M_N)^T S_N^{-1}(W - M_N)$$

where

$$S_N = (\beta S_0^{-1} + \beta\Phi^T\Phi)^{-1}, M_N = S_N(\beta\Phi^T Z + \beta S_0^{-1}M_0)$$

and since Gamma remains the same

$$a_N = a_0, b_N = b_0$$

Problem 4

Since the distribution in this case is $\mathcal{N}(10, 4)$. The probability of this distribution greater than the mean is 50%.

Problem 5

In this case the distribution is $\mathcal{N}(15, 4)$. Expected value is just the mean so it is 15.

Problem 6