Machine learning Homework- Linear regression

Abinav Ravi Venkatakrishnan - 03694216 and Abhijeet Parida - 03679676

November 11, 2018

1 Least Square Regression

Problem 2

Given: The error function

$$E_{weighted}(w) = \frac{1}{2}\sum_{i=1}^{N} t_i[w^T\phi(x_i) - y_i]^2 \tag{1}$$

We know that taking the first derivative of the cost function gives the minimum of the error function.

$$\nabla E_{weighted}(w) = \nabla_w \frac{1}{2}\sum_{i=1}^{N} t_i[w^T\phi(x_i) - y_i]^2 = 0 \tag{2}$$

Now the RHS can be written as a matrix equation with T as the diagonal matrix.

$$\nabla_w E_{weighted} = \nabla_w \frac{1}{2}[\Phi^T \mathbf{T} w - y]^T[\Phi^T \mathbf{T} w - y] \tag{3}$$

If T = I(Identity Matrix) hence giving the solution

$$w_* = (\Phi^T\Phi)^{-1}\Phi^T y \tag{4}$$

t_i acts as a multiplication factor for variance in y_i due to x_i

when the data points for which there are exact copies in the dataset.The error contribution due to these repeated points are increased.

2 Ridge regression

Problem 3

Before Augmentation,

$$E_{ridge} = \frac{1}{2}\sum_{i=1}^{N}(w_N^T\phi(x_i) - y_i)^2 + \frac{\lambda}{2}w_N^T w_N \tag{5}$$

After Augmentation,

$$E_{ridge} = \frac{1}{2}\sum_{i=1}^{N+M}(w_{N+M}^T\phi(x_i) - y_i)^2 + \frac{\lambda}{2}w_{N+M}^T w_{N+M}$$

This can be split as,

$$\begin{aligned} E_{ridge} &= \tfrac{1}{2}\textstyle\sum_{i=1}^{N+M}(w_{N+M}^T\phi(x_i) - y_i)^2 + \tfrac{\lambda}{2}w_{N+M}^T w_{N+M} && (6)\\ &= \tfrac{1}{2}\textstyle\sum_{i=1}^{N}(w_N^T\phi(x_i) - y_i)^2 + \tfrac{\lambda}{2}w_N^T w_N + \tfrac{1}{2}\textstyle\sum_{i=N}^{N+M}(w_{N:N+M}^T\phi(x_i) - y_i)^2 + \tfrac{\lambda}{2}w_{N:N+M}^T w_{N:N+M} && (7)\\ & && (8)\end{aligned}$$

In 8, the term $w_{N:N+M} = 0$, because

$$w_{N:N+M} = (X_{N:N+M}^T X_{N:N+M})^{-1} X_{N:N+M}^T y_{N:N+M}$$

because

$$y_{N:N+M} = 0$$

Therefore 8 is equal to 5

3 Bayesian linear regression

Problem 4

We know that
posterior $\propto$likelyhood x prior
we will work in the log scale to easily deal with the $\prod$

$$\begin{aligned} log(\mathcal{N}(\mathbf{w}|\mathbf{m}_N, \beta^{-1}\mathbf{S}_N)) + log(Gamma(\beta|a_N, b_N)) = & \sum_{i=1}^{N} log(\mathcal{N}(y_i|\mathbf{w}^T\phi(x_i), \beta^{-1})) \\ & +log(\mathcal{N}(\mathbf{w}|\mathbf{m}_0, \beta^{-1}\mathbf{S}_0)) \\ & +log(Gamma(\beta|a_0, b_0)) \end{aligned}$$

By simple pattern matching after resolution we can easily find the unknowns and will hence prove the validity of the claim.