Machine Learning HW2

Shang-Ying He
Communication Engineering
sunnyhe1122@gmail.com

About this homework we need to implement the Bayesian linear regression, and for the weight (w) I choose Gaussian distribution as initial just like the function blow, mean and standard deviation called m_0 and S_0.

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

After training number of N data distribution of weight would be the function blow,

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

and for each iteration we need to update m_N and S_N, just follow the function. If data

$$\begin{aligned} \mathbf{m}_N &= \mathbf{S}_N \left(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\mathbf{\Phi}^{\mathrm{T}}\mathbf{t}\right) \\ \mathbf{S}_N^{-1} &= \mathbf{S}_0^{-1} + \beta\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}. \end{aligned}$$

points arrive sequentially, then the posterior distribution at any stage acts as the prior distribution for the subsequent data point. And last use the following function to predict the testing data.

$$p(t|\mathbf{x}, \mathbf{t}, \alpha, \beta) = \mathcal{N}(t|\mathbf{m}_N^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}), \sigma_N^2(\mathbf{x}))$$

For Q1 and Q2, it means that if the value much larger, the representatives of the feature for a data point would be larger, too. For example, if Q1=3 Q2=5 length of feature vector would be Q1*Q2+2=17, concept about these two value is to estimate the feature as Gaussian, and its smallest value is 2, so I set as this first, I got the mean square in testing data is 0.013672525222852791, next I try larger number as Q1=2 Q2=10, Q1=10 Q2=2, Q1=10 Q2=10 or larger but I got the same mean square error as 0.013672525222875893, compare with prior value it almost didn't change and even with no improve. Another thing is that I tuned the value alpha to improve the MSE in testing data, finally I set O1=2 O2=2 alpha=25 to get the MSE=0.01327.