Bayesian Sequential Learning: Gaussian Models

Develop a program for determining the predictive probability distribution for predicting the target value \( t_n = w_0 + w_1 x + \epsilon \) where \( \epsilon \sim N(0, \beta^{-1}) \) using Bayesian models. Present the results first for the case (a) One unknown: \( w_0 = 0, w_1 \) unknown; and then for the two-parameter estimation: (b) Both \( w_0, w_1 \) are to be estimated.

A summary of the steps to be carried out is given below. This is given for the one parameter case. Let \( w_1 = w. \)

(a) Create a prior distribution for \( w : N(0, \frac{1}{\alpha}) \); You can select for example a range of \( w \) : \( -\hat{w}, \hat{w} \) and prescribe the zero-mean Gaussian distribution for the selected \( \alpha \) over this range of \( w \).

(b) Generate a set of target values \( t_n, = wx_n + \epsilon_n \ n = 1, 2, \ldots N \)

(c) Using the first target value (\( n=1 \)) compute the likelihood function : \( p(t_1 | x, \beta, w) : N(wx_1, \beta^{-1}) \). The likelihood function is the Gaussian probability for the observed value \( t_1 \) with a mean value of \( w x_1 \). Since \( t_1, x_1 \) known and \( w \) unknown, the likelihood function is to be represented as a function of \( w \), typically over the same range selected in (a).

(d) Multiply the prior in (a) with the likelihood in (c) and normalize it so that it represents a pdf. This is the posterior distribution. Plot this function again as a function of \( w \) and observe the change in the shape of the posterior.

(e) Repeat the step in (c) for second target value \( n = 2 \) to get the likelihood function for the second data point.

(f) Multiply the result of (e) with the posterior distribution computed in (d) and normalize the result, to obtain a new Bayesian posterior function of \( w \). Plot this as a function of \( w \) and observe the changes.

(g) Repeat steps of computing the likelihood function for each new target value \( t_n \) and multiplying it with the posterior function obtained from the previous iteration.

If done right, the posterior distribution of \( w \) should converge to approximately a delta function centered around the actual value of \( w \) as the number of training points increase.

Once determined, the posterior distribution can be used to determine the predictive distribution of \( y \) for test data. The predictive distribution is obtained as ,

\[
p(t|t, \alpha, \beta) = \int_{-\infty}^{\infty} p(t|w, \beta) p(w|t, \alpha, \beta) \ dw \quad (1)
\]