# Assignment - 1

Name :- Prince Makwana.

Subject :- Statistical Foundations For Machine Learning

Problem Statement : Find $m(x)$ and $s^2(x)$ of $p(t/x, x, t)$ for Gaussian Distribution.

Soln :- In Bayesian curve fitting, we want to estimate the distribution of a target variable (i.e., 't') given some input variable (i.e., 'x') and an observed data (i.e., 't/x').

Assume a linear model with Gaussian noise, we can represent the relationship b/w the target variable and the input variable as follows:

$$t = y(x, w) + \epsilon$$

where, ~~t ⇒ predicted value~~
$y(x,w)$ ⇒ predicted value.
$w$ ⇒ weight
$x$ ⇒ input variable.
$\epsilon$ ⇒ Gaussian noise.
$t$ ⇒ target value

To find the distribution of the parameters $(w)$ given the observed data, we use Bayes' theorem. It allows us ~~the~~ to calculate the posterior distribution, which tells us the most likely values of the parameters given the data.

The posterior distribution is given by:

$$p(w/t, x) = (p(t/w, x) * p(w/x)) / p(t/x)$$

In this equation, $p(\omega|t, x)$ is the posterior distribution we want to find, $p(t|\omega, x)$ is the likelihood function, $p(\omega|x)$ is the prior distribution representing our initial beliefs about the parameters, and $p(t|x)$ is the evidence.

For a Gaussian likelihood function, we assume that the noise follows a Gaussian distribution. This means that likelihood can be written as:

$$p(t|\omega, x) = N(t|y(x, \omega), \beta^{-1})$$

Here, $N(\mu, \sigma^2)$ denotes a Gaussian distribution with mean $\mu$ and variance $\sigma^2$. The precision $(\beta^{-1})$ represents the inverse of the noise variance.

Assuming a Gaussian prior distribution for the parameters the prior distribution can be written as:

$$p(\omega|x) = N(\omega|\mu_0, \Sigma_0).$$

Here, $\mu_0$ is mean of the prior distribution.
$\Sigma_0$ is covariance matrix.

To compute the posterior distribution, we need to compute the mean $(\mu_n)$ and covariance matrix $(\Sigma_n)$ of the parameters. These can be obtained using the following ~~parameters~~ formulas:

$$\cancel{\Sigma_n^{-1} = \Sigma_0^{-1} + \beta * \phi^T * \phi}.$$

$$\mu_n = \Sigma_n * (\Sigma_0^{-1} * \mu_0 + \beta * \phi^T * t).$$

where, ~~$\Sigma_n$/ represen~~

$\Sigma_n \Rightarrow$ covariance matrix of the posterior distribution. It captures the uncertainty or spread of the parameter estimates.

$\Sigma_0^{-1} \Rightarrow$ inverse of the covariance matrix of the prior distribution. It represents the ~~ancest~~ initial uncertainity or spread in our beliefs about the parameters.

$\mu_0 \Rightarrow$ mean of the prior distribution. It represents our initial expectations for the parameter values.

$\beta \Rightarrow$ precision parameter, which is the inverse of the noise variance. It quantifies the level of noise in the observed data.

$\phi \Rightarrow$ Design matrix, which is constructed by applying basis functions to the input variables $x$. Each row of $\phi$ corresponds to an input point and contains the evaluations of the basis functions at that point.

$t \Rightarrow$ Observed data.

The formula for the mean $\mu_n$ combines the prior mean $\mu_0$, the weighted sum of the prior mean and the contribution of the observed data $(\beta * \phi^T * t)$, and the spread of the posterior distribution $(\Sigma_n)$.

Now, let's move on ~~the~~ to calculating the variance of the posterior distribution, denoted as $s^2(x)$. It represents the uncertainty or ~~variable~~ variability of the predicted values for new input values ~~$x_{new}$~~ ~~or~~ $x_{new}$.

The variance can be computed using the formula:

DOMS Page No.
Date / /

$$s^2(x_{new}) = \frac{1}{\beta} + \phi(x_{new})^T * \Sigma_n * \phi(x_{new}).$$

~~Here,~~ Here's a breakdown of the components involved:

- $\frac{1}{\beta}$ ⇒ represents the noise variance. It quantifies the uncertainty or spread due to the noise in the observed data.
- $\phi(x_{new})$ represents the vector of basis function evaluations at the new input values $x_{new}$. It captures the relationship between the inputs and the parameters.
- $\Sigma_n$ is the covariance matrix of the posterior distribution, which reflects the uncertainity or spread in the parameter estimates.