PATTERN RECOGNITION (BRI623), SPRING 2016

ASSIGNMENT #1: BAYESIAN LINEAR REGRESSION

PROF. HEUNG-IL SUK

[Problem #1] With the likelihood function and the prior over the model parameter $\mathbf{w}$ defined as below, please derive the posterior probability distribution of $\mathbf{w}$ and the predictive distribution of t for a new sample x.

- Likelihood function $p(\mathbf{t}|\mathbf{w})$: exponential of a quadratic function of $\mathbf{w}$

$$p(\mathbf{t}|\mathbf{X},\mathbf{w},\beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n|\mathbf{w}^\top \phi(\mathbf{x}_n), \beta^{-1}\right)$$

- (Conjugate prior) Gaussian distribution

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

- Posterior probability distribution of $\mathbf{w}$

$$\begin{aligned} p(\mathbf{w}|\mathbf{x},\mathbf{t},\alpha,\beta) &\propto p(\mathbf{t}|\mathbf{w},\mathbf{x},\beta)\, p(\mathbf{w}) \\ &= \mathcal{N}\left(\mathbf{t}|\Phi^\top \mathbf{w}, \beta^{-1}\mathbf{I}\right) \mathcal{N}(\mathbf{w}|\mathbf{m}_0,\mathbf{S}_0) \\ p(\mathbf{w}|\mathbf{x},\mathbf{t},\alpha,\beta) &= \mathcal{N}(\mathbf{w}|\mathbf{m}_N,\mathbf{S}_N) \\ \text{where} &\begin{cases} \mathbf{m}_N = \mathbf{S}_N\left(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\Phi^\top\mathbf{t}\right) \\ \mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta\Phi^\top\Phi \end{cases} \end{aligned}$$

- Predictive distribution of t for a new sample x

$$\begin{aligned} p(t|x,\mathbf{x},\mathbf{t}) &= \int p(t|x,\mathbf{w})\, p(\mathbf{w}|\mathbf{x},\mathbf{t})\, d\mathbf{w} \\ &= \int \mathcal{N}\left(t\,\middle|\,\phi(x)^\top \mathbf{w}, \beta^{-1}\right) \mathcal{N}(\mathbf{w}|\mathbf{m}_N,\mathbf{S}_N) \\ &= \mathcal{N}\left(t\,\middle|\,\underbrace{\phi(x)^\top \mathbf{m}_N}_{m(x)}, \underbrace{\beta^{-1} + \phi(x)^\top \mathbf{S}_N \phi(x)}_{s^2(x)}\right) \\ \text{where} &\begin{cases} \mathbf{m}_N = \mathbf{S}_N\left(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\Phi^\top\mathbf{t}\right) \\ \mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta\Phi^\top\Phi \end{cases} \end{aligned}$$

[Problem #2] Programming

- Objective: To understand the differences among maximum likelihood estimation (MLE), maximum a posterior (MAP), and fully Bayesian approach in regression

- Goal: To write computer programs, no restriction on programming languages, of MLE, MAP, and Bayesian methods for predicting a continuous target variable t for a test sample x

- Detailed descriptions
  - Let us limit the input space x in the range of $[-1, 1]$.
  - Generate 100 *i.i.d.* sample points, $\{(x_n, t_n)\}_{n=1}^{100}$, from a function $\sin(2\pi x)$ with a random noise $\epsilon \sim \mathcal{N}\left(0, \beta^{-1}\right)$, where $\beta = 11.1$ fixed:
  $$t_n = \sin(2\pi x_n) + \epsilon.$$
  - Use a 9-th order polynomial function defined as follows:
  $$y(x, \mathbf{w}) = \sum_{j=1}^{9} w_j x^j + w_0.$$
  - For both MAP and Bayesian, use a prior distribution over $\mathbf{w}$ as $p(\mathbf{w}|\alpha) \sim \mathcal{N}\left(0, \alpha^{-1}\mathbf{I}\right)$, where $\alpha = 5 \times 10^{-3}$ fixed.

- Requirements
  - Submission: a zipped file including program source codes and a short description on how to run your program
  - The main function should take an input value of x for testing.
  - For the Bayesian method, it should display the predicted target value along with ± 1 standard deviation in a graph.

- Due date: April 30, 2017 (No Extension)