Machine Learning (Homework #1)

Due date: October 28, 2016

1. Bayesian Linear Regression

For a given input value x, the corresponding target value t is assumed as a Gaussian distribution $p(t|x, \mathbf{w}, \beta) = \mathcal{N}(t|y(x, \mathbf{w}), \beta^{-1})$ and the prior distribution of $\mathbf{w}$ is also assumed as a Gaussian distribution $p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|0, \alpha^{-1}\mathbf{I})$. A linear regression function is expressed by $y(x, \mathbf{w}) = \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(x)$. We are not only interested in the optimal value of $\mathbf{w}$ but also in making prediction of t for new test data x. To do so, we multiply the likelihood function of new data $p(t|x, \mathbf{w}, \beta)$ and the posterior distribution of the training data $p(\mathbf{w}|\mathbf{x}, \mathbf{t})$ and take the integral over $\mathbf{w}$ as $\int_{-\infty}^{\infty} p(t|x, \mathbf{w}, \beta)p(\mathbf{w}|\mathbf{x}, \mathbf{t})d\mathbf{w}$ where $\mathbf{x} = \{x_1, \dots, x_N\}$ and $\mathbf{t} = \{t_1, \dots, t_N\}$. Please show the details of your derivation for the predictive distribution which is a Gaussian distribution of the form $p(t|x, \mathbf{x}, \mathbf{t}) = \mathcal{N}(t|m(x), s^2(x))$ where

$$m(x) = \beta\boldsymbol{\phi}(x)^{\mathrm{T}}\mathbf{S}\sum_{n=1}^{N}\boldsymbol{\phi}(x_n)t_n, \;\; s^2(x) = \beta^{-1} + \boldsymbol{\phi}(x)^{\mathrm{T}}\mathbf{S}\boldsymbol{\phi}(x)$$

and

$$\mathbf{S}^{-1} = \alpha\mathbf{I} + \beta\sum_{n=1}^{N}\boldsymbol{\phi}(x_n)\boldsymbol{\phi}(x_n)^{\mathrm{T}}.$$

Hint: you may use these formulas in Page 93:

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$$
$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \boldsymbol{b}, \boldsymbol{L}^{-1})$$
$$p(\mathbf{y}) = \mathcal{N}(\boldsymbol{y}|\mathbf{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\top})$$
$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}\{\mathbf{A}^{\top}\boldsymbol{L}(\mathbf{y} - \boldsymbol{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\}, \boldsymbol{\Sigma})$$

2. Maximum Entropy

Gaussian is known as the optimal distribution which achieves the maximum entropy given by the constraints of the mean and variance of the distribution. Please illustrate this property by showing the details of finding this optimal distribution or equivalently deriving the Equations (1.108) and (1.109) in text book.

3. Application for Polynomial Regression

In this exercise, you are facing the problem in practical applications when dealing with high-dimensional data by using general regression techniques. Here, the Combined Cycle Power Plant Data Set is given in **data.xlsx**. Please write a regression program for the **net hourly electrical energy output** estimation by minimizing the error function

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(\mathbf{x}_n, \mathbf{w}) - \boldsymbol{t}_n\}^2.$$

A general polynomial with coefficients, for example, up to order 2 is formed by

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{i=1}^{D} w_i x_i + \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij} x_i x_j$$

where $\mathbf{x} = [x_1, \ldots, x_D]^\top$ and $\mathbf{w} = \{w_0, w_i, w_{ij}\}$. The data set contains 500 instances. Each instance has 5 attributes and the EP attribute is the target. In this exercise, the first 400 samples are used as the training set and the last 100 samples are used as the test set. The attribute information is described as

T: Temperature (T) in the range 1.81°C and 37.11°C
V: Exhaust Vacuum (V) in teh range 25.36-81.56 cm Hg
AP: Ambient Pressure (AP) in the range 992.89-1033.30 milibar
RH: Relative Humidity (RH) in the range 25.56% to 100.16%
EP: Net hourly electrical energy output (EP) 420.26-495.76 MW

(1) In the training stage, please apply the polynomials of order $M = 2$ and $M = 3$ over input data. Please evaluate the corresponding root-mean-square RMS error on the training set and test set.
(2) Please apply polynomials of order $M = 3$ and select the most contributive attribute with the lowest RMS error both on the training set and test set.

4. Cross Validation

In this exercise, you are facing the problem on dealing with high-order model by using cross-validation. Here, the data set is given (**x3.mat and t3.mat**). Please write a regression program for the regression estimation by minimizing the error function:

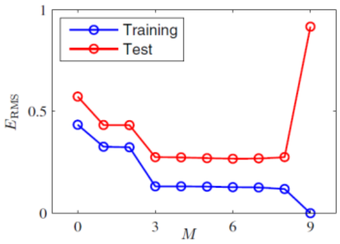
$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2$$

where a polynomial of order M is formed by

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M.$$

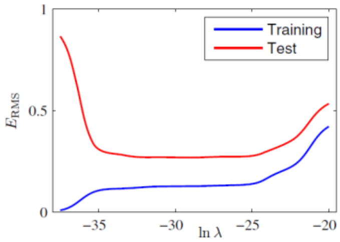
(1) Using cross-validation to select the best model. Please draw the training and

testing error using the polynomial with different orders. Explain which order will cause the overfitting problem. Below is an example in the textbook.



(2) Adding a regularization term into the model, and replying again (1). Below is the example in the textbook.

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}||\mathbf{w}||^2$$



Note:

a. Equally divide the training set into 3 validation sets, and report the validation error.

b. In (2), using cross validation to choose regularization parameter λ.