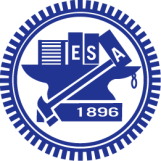
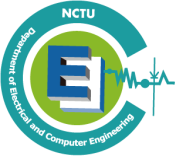
** Machine Learning (Homework #1) **

Due date: 10/23

1. **Bayesian Linear Regression**

For a given input value *x*, the corresponding target value *t* is assumed as a Gaussian distribution  and the prior distribution of  is also assumed as a Gaussian distribution . A linear regression function is expressed by . We are not only interested in the value  but also in making prediction of  for new test data . We multiply the likelihood function of new data  and the posterior distribution of the training data  and take the integral over  as . Please derive the resulting predictive distribution which is a Gaussian distribution of the form 

Hint: you can use these equations in page 93:



where

, 

and



1. **Polynomial Curve Fitting**

Please write a MATLAB program to implement linear regression. You are given a file **x2.mat** and **t2.mat**, which contains two arrays:

**x**: represents the input values and

**t**: represents the target values.

The number of data points is 100. Let us split them into training set (the first 80 points) and test set (the last 20 points). In the training stage, please fit the data by applying a polynomial function of the form

and minimizing the error function:

1. Only use the first 20 points of training set for regression and show the Root-Mean-Square (RMS) error () evaluated on the training set (only the first 20 points) and test set for various values of from 0 to 9.
2. Please plot the regression results on the training set (only the first 20 points) for various order from 0 to 9 similar to Figure 1.4 in textbook.
3. Please show the table of the estimated coefficients w\* for polynomials with various order similar to Figure 1.1 in textbook.
4. Considering the regularized error function

where . Please set two values for regularization parameter as and repeat ***Part [a]***.(note: is calculated using not )

1. Please plot the regularized regression result on the Training Set (only the first 20 points) for various order from 0 to 9.
2. Please show the table of the estimated coefficients w\* for polynomials with various order and .
3. Use the whole training set for regression and repeat ***Part [a] ~ Part [f]***.
4. **Application for Polynomial Regression**

In this exercise, you will experience the difficulty of practical applications when dealing with high-dimensional data by using general regression techniques. Here, the Iris data set is given (**x3.mat and t3.mat**). Please write a regression program for the iris class estimation by minimizing the error function:

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

The data set contains 3 classes of 50 instances each and the total number of samples is 150. In this exercise, the first 40 samples of each class are used as the Training set and the last 10 samples of each class are used as the test set.

Data Description

Number of Instances: 150

Number of Attributes: 5 (4-dimensional input + 1-dimensional target)

Attribute Information:

1. sepal length in cm   
2. sepal width in cm   
3. petal length in cm   
4. petal width in cm   
5. class (three class):   
-- Iris Setosa (class 1)  
-- Iris Versicolour (class 2)  
-- Iris Virginica (class 3)

http://archive.ics.uci.edu/ml/datasets/Iris

1. In the training stage, please apply polynomials of order *M*=1 and *M*=2 over the 4-dimensional input data. Please evaluate the corresponding RMS error on the training set and test set.
2. Please apply polynomials of order *M*=3 and select the most contributive attribute which has the lowest RMS error both on the training set and test set.