Machine Learning (Homework #1)

Due date: October20, 2017

1. Jensen's inequality (10%)

Convexity implies

$$f(\lambda a + (1-\lambda)b) \le \lambda f(a) + (1-\lambda) f(b) \tag{1}$$

A convex function $f(x)$ satisfies

$$f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \le \sum_{i=1}^{M} \lambda_i f(x_i) \tag{2}$$

where

$$\lambda_i \ge 0 \text{ and } \sum_i \lambda_i = 1 \tag{3}$$

Using the technique of proof by **induction**, please derive the equation (2) from equation (1) and property (3).

Hint : $\sum_{i=1}^{M} \lambda_i x_i = a \sum_{i=1}^{M} b_i x_i; a = g(\lambda_{M+1}); b_i = h(\lambda_{M+1}, \lambda_i)$

2. Entropy of the univariate Gaussian (15%)

Derive the entropy of the univariate Gaussian

$$\mathrm{H}[x] = \frac{1}{2}\{1 + \ln(2\pi\sigma^2)\}$$

from

$$p(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\int_{-\infty}^{\infty} x p(x) dx = \mu$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx = \sigma^2$$

Hint : $\mathrm{E}[X^2] = \sigma^2 + \mu^2$

3. KL divergence between two Gaussians (15%)

$$\mathrm{KL}(p\|q) = -\int p(\mathbf{x}) \ln q(\mathbf{x}) d\mathbf{x} - \left(-\int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}\right)$$

$$= -\int p(\mathbf{x}) \ln \left\{\frac{q(\mathbf{x})}{p(\mathbf{x})}\right\}$$

Evaluate the KL divergence between two Gaussian $p(x) = N(x|\mu, \sigma^2)$ and $q(x) = N(x|m, s^2)$.

Hint : $\mathrm{E}[X^2] = \sigma^2 + \mu^2$

4. Polynomial Curve Fitting (30%)

Please write a program to implement linear regression. You are given a data set (**4_train.csv and 4_test.csv**), which contains two arrays:

$\mathbf{x} : \{x_1, x_2, \ldots, x_{20}\}$ represents the input values and

$\mathbf{t} : \{t_1, t_2, \ldots, t_{10}\}$ represents the target values.

In the training stage, please fit the data by applying a polynomial function of the form

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

and minimizing the error function:

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

(1) Directly use the training set for regression and show the Root-Mean-Square (RMS) error ($E_{RMS} = \sqrt{2E(\mathbf{w})/N}$) evaluated on the training set and test set for various values of M from 1 to 9.

(2) Please plot the training and test error for various order M from 1 to 9 similar to Figure 1.5 in textbook. Explain which order will cause the overfitting problem.

(3) Considering the regularized error function

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

where $||\mathbf{w}||^2 \equiv \mathbf{w}^\top \mathbf{w} = w_0^2 + w_1^2 + w_2^2 + \cdots + w_M^2$. Please draw the training and test error similar to Figure 1.8 in textbook by using the polynomial function with order $M = 9$ and setting regularization parameter $\ln \lambda$ from -20 to 0.

5. Polynomial Regression (30%)

In real-world applications, the dimension of data is usually more than one. Here, the Iris data set is given (**5_X.mat and 5_T.mat**). Please write a regression program for the iris class estimation by minimizing the error function:

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

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

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

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.

<u>Data Description</u>

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=2 and select the most contributive attribute which has the lowest RMS error on the Training Set.