4105931 機器學習

Assignment #1 – Regression

Due Date : **04/07 23:59pm**

本次作業主要是實作 Unit-5、p.14 頁的迴歸公式，就可以完成(a) - (d)題；實作 Unit-7、p.11 頁的正規化公式，可以完成題(e)。本次作業需要自己實作上述公式，請勿使用任何現成的 regression 函式。

Use the linear model $y = 2x + \varepsilon$ with zero-mean Gaussian noise $\varepsilon \sim N(0, 1)$ to generate 20 data points with (equal spacing) $x \in [-3, 3]$.

(a) Perform *linear regression*. 20 data points are split into 15 training samples and 5 testing samples (75% for training and 25% for testing). Show the **fitting plots** of the **training error**, **cross-validation errors** for both leave-one-out and five-fold, and **testing errors**.

(b) Perform *polynomial regression* with degree 5, 10 and 14, respectively. For each case, show the fitting plots of the training error, cross-validation errors (both leave-one-out and five-fold) and testing errors.

(c) Generate data using $\mathbf{y = \sin(2\pi x) + \varepsilon}$ with the noise $\varepsilon \sim N(0, 0.04)$ and (equal spacing) $x \in [0, 1]$. Show the fitting plots of the training error, cross-validation errors for both leave-one-out and five-fold, and testing errors via polynomial regression with degree 5, 10 and 14.

(d) Consider the model in (b) with degree 14 via varying the number training data points m, say, $m = 60, 160, 320$. Show the five-fold cross-validation errors, testing error and the fitting plots with 75% for training and 25% for testing.

(e) Consider again the model in (b) with degree 14 via **regularization**:

$$J_m(\mathbf{w}) = \frac{1}{m}\sum_{i=1}^{m}(y_i - f(\mathbf{x}_i; \mathbf{w}))^2 + \lambda\|\mathbf{w}\|^2$$

Compare the results derived by setting $\lambda = 0, 0.001/m, 1/m, 1000/m$, where $m = 20$ is the number of data points (with $x = 0, 1/(m-1), 2/(m-1), \ldots, 1$). Show the five-fold cross-validation errors, testing errors and the fitting plots with regularization using the following

equation:

$$\mathbf{w} = \left(\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I}\right)^{-1} \mathbf{X}^\top \mathbf{y}$$

Reference Answer

Degree	Training Error	Leave-One-Out	Five-Fold
1	0.6275	0.6583	0.3736
5	0.4093	2.1785	20.0600
10	0.1165	3.3287e+003	1.3077e+005
14	1.0564e-013	7.3137e+005	2.7471e+008