Week1 Report

Zhengyi Wang
wangzhen17@mails.tsinghua.edu.cn

Abstract

This week, we talked about the fundemental of machine learning.

1 Basic terminology

Supervised learning, commonly used for classification and regression, means that the training process includes labeled data. Unsupervised learning, commonly used for clustering, is trained without labeled data.
The hypothesis space, is the set of all functions that probably describe the mapping between training set and test set.

2 Model evaluation

2.1 Empirical error

The error on training set is called empirical error. The $Error\ rate = \frac{Error\ cases}{All\ cases}$. And $The\ Accuracy = 1 - Error\ rate$.

2.2 Hold out

This method means to set aside some data for testing while the remaining for training. There is a dilemma–more data for training means less accuracy for testing while more data for testing means it might not be able to make the most of the model. In practice, training set usually takes up 2/3 - 4/5 of all data.

2.3 Cross validation

The main idea of cross validation is to devide the data into k subsets (mutually exclusive). And choose one from the k subsets each time as test set while others for training. Do that k times repeatedly.

2.4 Bootstrapping

The idea of bootstrapping is to generate a training set from the original set by randomly choosing one item from the data for D times (D=the size of data). Notice that the items in training set might be duplicated. The not-chosen items will be in the test set. There might be around 36.8% items don't appear in the training set. Which is shown by the following formula

$$\lim_{m \to +\infty} (1 - \frac{1}{m})^m = \frac{1}{e} = 0.368...$$

Reading Group on Machine Learning (Apr 7th, 2019).

3 Performance measure

There are many ways to measure the performance of our model.

3.1 Mean squared error

In regression tasks, mean squared error is used from time to time. It has different format in different cases. In discrete

$$E(f; D) = \frac{1}{m}\sum_{i=1}^{m}(f(x_i) - y_i)^2$$

In concrete

$$E(f; D) = \int_{x \sim D}(f(x) - y)^2 p(x)\underset{\cdot}{\mathrm{x}}$$

3.2 Bias-variance decomposition

Expectation of the learning algorithm is computed by the following formula

$$\bar{f}(x) = \mathbb{E}_D[f(x; D)]$$

Variance

$$var(x) = \mathbb{E}_D[(f(x; D) - \bar{f}(x)^2]$$

Noise

$$\mathcal{E}^2 = \mathbb{E}_D[(y_D - y)^2]$$

Bias

$$bias^2(x) = (\bar{f}(x) - y)^2)$$

And we can draw from the listed formula

$$E(f; D) = bias^2(x) + var(x) + \mathcal{E}^2$$