Week One

Zhang Yichi, 2018011390
zhangyic18@mails.tsinghua.edu.cn

April 13, 2019

Abstract

In this week's group discussion, Xiao Yijia delivered the lecture naming *Introduction to the Machine Learning* systematically, through which we got to learn the basic concept and methods of machine learning, decision tree and neural network. In the article below, I will summarize some of the knowledge that I find interesting and impressive.

1 Machine Learning

Machine learning is a field part of AI, which is specifically relevant to algorithms and statistical models. By building a model according to the training data, the computer can make decisions and predictions following this model.

1.1 Concept

In the field of machine learning, the learning tasks are divided into supervised learning, semi-supervised learning and unsupervised learning according to whether the training data has labels.
Lable is the expected output that is given to the samples by people.
Hypothesis space is a set of functions that can map the input, which is usually a eigenvector, to the output.
In machine learning, there is a key procedure called **reduction bias**. It is an essential method which can make sure the output is confirmable by giving different priority to different attributes. What's worth attention is that different bias will fit different models because there is no model that can perform well on all problems.**(No Free Lunch Theorem)**

1.2 Model Evaluation

A trained model need to be evaluated. Below are some typical methods:
• Hold-out: Dividing the dataset into training set and testing set, usually with the ratio from 2/3-4/5.

• Cross Validation: Dividing the dataset into k mutually exclusive subsets and using each one as testing set and the rest as training sets in turn. Usually, k is 10.
• Bootstrapping: Randomly picking some samples from the dataset and therefore generating a new set which will be used as training set. In this way, there will always about 36.8% samples that will not used to train the model.

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

1.3 Performance Measure

• Error rate:

$$E(f;D) = \frac{1}{m}\sum_{i=1}^{n} I(f(x_i) \neq y_i)$$

• Bias-Variance: To evaluate the generalization error of a model, Bias-Variance Decomposition is needed. In the following formulas, we fix the sample x and consider its different output in different dataset.
· Expectation:

$$\overline{f}(x) = E_D[f(x;D)]$$

· Variance: describe the fluctuation of the model's learning ablitiy when trained by different dataset.

$$var(x) = E_D[(f(x;D) - \overline{f}(x))^2]$$

· Noise: describe the lower bound of generalization error expectation of all possible algorithms, caused by the wrong labels.

$$\epsilon^2 = E_D[(y_D - y)^2]$$

· Bias: describe the learning ability of the algorithm itself, by measuring the expectation of output minus the actual lable.

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

· **B**ias-Variance Decomposition: assuming

$$E_D[y_D - y] = 0$$

then

$$E[f;D] = E_D[(f(x;D) - y_D)^2] = bias^2(x) + var(x) + \epsilon^2,$$

2 Decision Tree

2.1 Procedure

Setting each node with an attribute and keep dividing unless it meets one of the following demands:

· all samples remained in the dataset are of the same kind;
· attribute set is empty/ all samples are the same in all attributes left;
· the initial dataset is empty.

2.2 Information Entropy

Information entropy is used to measure the purity of the samples. Assuming the proportion of type k is p_k, then

$$Ent(D) = -\sum_{k=1}^{|y|} p_k log_2 p_k$$

The lower Ent(D) is, the higher the purity is.
• Information Gain: this is a basis for the priority among the attributes, according to the principle that the attribute with the highest priority should show the most potential of improvement in purity.
Assuming attribute a has multiple possible values and the dataset D can be divided according to different values into Dv

$$Gain(D, a) = Ent(D) - \sum_{v=1}^{V} \frac{|D^v|}{D} Ent(D^v)$$

2.3 Pruning

The goal of pruning is to improve the generalization ability of the decision tree, avoiding the possible outcome like overfitting which takes the unique features of the samples into consideration when training the model. Prepruning and Postpruning.

3 Neural Network

3.1 Perceptron

Perceptron is made up of two layers of neurons, which can only deal with problems that are linearly separable.
By regarding the thresholds as dummy nodes whose input is constantly -1.0, we can unify the learning into the learning of weight only.