1 Intro and Declaration

This note is all based on a series of blogs from Chen Haoran, whose personnal blog is chrer.com. I may delete this blog if I break any lisence. Best regrads to Chen and his girlfriend, and may they be together forever.

2 Concept, history and future of ML

2.1 concept

Machine learning is about designing an algorithm which learnshow to design a algorithm, kind of a recurrence definition, especially when such an algorithm cannot/is hard to be programmed explicitly.

What does a machine learn from? Dataset! A typical dataset consists of several attributes, which may be regarded as a attribute space. And out of these data, we want to predict a label, and we prefer some models that generlize well to empericial data, the real world, not only to the dataset, which is merely a sample/subset of empreical data.

Ml can be categorized into four types: supervised, unsupevised, semisupervised and reinforcement.

- Supervised: each entry has its own label given, can be further split into two classes: categorization and regression, depends on whether the label is discrete or continous
- Unsupervised: no labels are given
- Semisupervised: simply a mix of above
- RL: no label....but we have a reward function

2.2 history and application

from inference to knowledge to ML, boring.

google news, auto drive, ads

3 Linear regression and Gradient descent

Modeling first, how to encode these features? A real number or one-hot encoding? And better normalize the features to $[-1, 1]$

Well, linear regerssion has a formular, and is actually easy to obtain.

$$y = (A^T A)^{-1} A^T x \tag{1}$$

Gradient descent is like go down a hill, we simply look around, find a direction which goes down fastest, and just go that way, repeat until we get to the minimum. How large a step should be? Learning rate

4 Logistics and Problems in Linear Models

How to map a continous value into $\{0, 1\}$, suppose we are dealing with a binary classification problem? Introduce sigmoid function

$$y = \frac{1}{1 + e^{-x}} \tag{2}$$

, and set a threhold of 0.5. Notice that this function is not convex (and not convave, neither), introduce new cost functions:

$$Cost(h_w(x), y) = -\log(h_w(x)), y = 1 \quad Cost(h_w(x), y) = -\log(1 - h_w(x)), y = 0 \tag{3}$$

, and a regularization term like $\sum_{i=1}^{m} \theta_i^2$, may reduce overfit

5 Decision Tree

refer to *ISLR*, boring. Information Entropy, sounds interesting, search later.

6 Intro to Neural Network

Lots of basic elements, which connect and interact with each other. What is so-called connect and interact?

$$y = f(\sum_{1}^{n} w_i x_i - \theta) \tag{4}$$

, where the f here is called activate function.

Perceptrons and deeper perceptrons, inference and back propogation. Some theoritical statements.

7 Naive and Semi-naive Bayesian

Not directly estimate the probility of x in class c, but estimate the probility of one element in class c that turns out to be like x. MLE, and laplacian smoothing

If we allow some of attributes depends on other attributes, we obtain a semi-naive Bayesian model.

8 Integrate Learning

Keep a set of week learners(?), each one is good enough and different. Refer to *ISLR*, I remember its in Chap. Desition Trees, and review this!

omit Theory of IL, Abnormal Detection

9 Cluster Analysis

refer to KNN, hope I still remember this...and can be generlize to Gausian mixed model

10 A Paper

momomo...doushidalao....yijingyoupaperle

11 Support Vector Machine

Please refer to *ISLR* from max margin classifier, to SVM, soft margin

12 Intro to Tensorflow

Well, eventually, we step into engineering and coding now...some common mistakes noobs make. Some lectures by Mr. Andrew Ng.

And some notes about RNN