BITS F464: Machine Learning

02 Basics of Machine Learning

Dr. Kamlesh Tiwari
Assistant Professor, Department of CSIS,
BITS Pilani, Pilani Campus, Rajasthan-333031 INDIA

Jan 20, 2021 ONLINE (Campus @ BITS-Pilani Jan-May 2021)

http://ktiwari.in/ml

Introduction

ML depends upon **Pattern Recognition** which corresponds to finding regularities in the data.

- There should be a pattern.
- No issues if we are unable to describe it mathematically.
- Sufficient number of examples or data is required.

Consider e-mail filtering SPAM/Not-SPAM

Assumption is that there are some words whose frequency is correlated to this filtering.

Netflix Prize (2009)

Open competition to predict user ratings for films.
Prize of USD 1 million was given to the BellKor's Pragmatic Chaos team which improved previous prediction by by ~10.06%
(used matrix factorization)

Building Blocks

- Input: x
- Output: y
- Training data: $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), ..., (x^{(m)}, y^{(m)})$
- $x^{(i)}$ could be a multivariate say $x^{(i)} = (x_1^{(i)}, x_2^{(i)}, ..., x_n^{(i)})$
- Concept, target function: true function

$$f : x \rightarrow y$$

- Hypothesis:

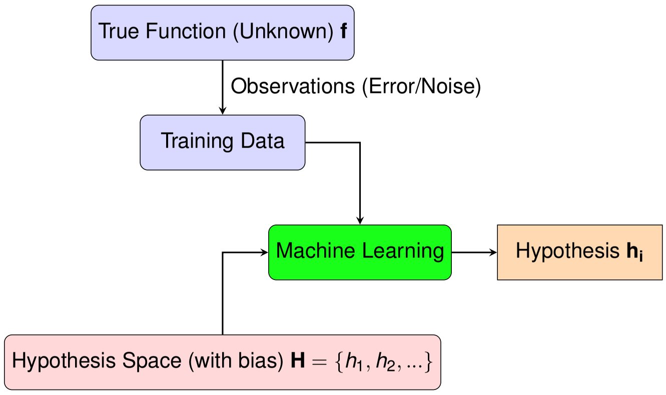
$$h : x \rightarrow y$$

- Accuracy: agreement b/w f and h

Issue is

The true function is NOT known.

The Flow of ML



A Toy model

- **The Problem:** credit approval.
- Input: $x = (x_1, x_2, ..., x_n)$
- Let x_1=accountBal, x_2=Salary, x_3=age ...
- What weights we should give w_1=0.6, w_2=0.3, w_3=-0.1 ...
- The Model

$$\sum_{i=1}^{n} w_i \times x_i = \begin{cases} > \textit{Threshold} & \text{Then APPROVE} \\ \textit{otherwise} & \text{DENY/REJECT} \end{cases}$$

- Simplified:

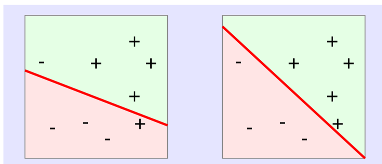
$$h(x) = sign(\sum_{i=1}^{n} w_i \times x_i - Threshold)$$

- Add an extra term x_0 (that is always 1), then

$$h(x) = sign(\sum_{i=0}^{n} w_i \times x_i)$$

A Toy model (Contd..)

- Can you recognize $h(x) = sign(\sum_{i=0}^{n} w_i \times x_i)$
- It is a linear equation (in two dimension) or hyper plane
- Sign could be positive or negative, so two classes are **+1** and **–1**



- Vector $w = (w_0, w_2, ..., w_n)$ would be normal to the plane of linear decision boundary. (why? because dot product is $\cos\theta$)
- What could change this plane? w_i's

Learning: Use misclassified examples to update $w_i = w_i + \alpha y_i x_i$

A Toy model (What yw^Tx tells)

- Classification of a point x can be obtained by w^Tx. If w^Tx if positive then x is positive, otherwise negative.
- ML assumes that data point is never on hyperplane so $w^Tx \neq 0$
- There could be two cases
  1. **When classification of the model is correct**:
     For $y = +1$ we have $w^Tx > 0$
     For $y = -1$ we have $w^Tx < 0$ | In both the cases $yw^Tx > 0$
  2. **When classification of the model is wrong**:
     For $y = +1$ we have $w^Tx < 0$
     For $y = -1$ we have $w^Tx > 0$ | In both the cases $yw^Tx < 0$

So we have a simple test

$$yw^Tx \begin{cases} > 0 & \text{Classification is correct} \\ < 0 & \text{Classification is wrong} \end{cases}$$

Loss Function

Performance is the closeness of hypothesis with target function

- For example
  - Classification

$$loss(y, h(x)) = \begin{cases} 1 & \text{if } h(x) \neq y \\ 0 & \text{otherwise} \end{cases}$$

  - Regression

$$loss(y, h(x)) = \begin{cases} (h(x) - y)^2 & \text{if } h(x) \neq y \\ 0 & \text{otherwise} \end{cases}$$

Issue is that, we can only do Empirical Risk Minimization.

Since only training data is available, on can use only this to be good (minimize risk on empirical data)

Thank You!

Thank you very much for your attention!

Queries ?