Learning From Data

Philip Tracton

May 18, 2018

Contents

1 Lecture 01 The Learning Problem

- These are my notes based on the online course Learning from Data at Cal Tech by Professor Yaser Abu-Mostafa
- The first lecture is at You Tube
- The audio does kind of fade in and out

1.1 Story Line

There is a theme through the course. Each lecture is not wholely independent from each other. The key steps in our story are

- What is Learning?
- Can we learn?
- How to learn?
- How to learn well?
- Take home lessons

Lecture 3 is a practical topic not really part of the story. Gives you tools to actually work on this material. Avoids the course from being too theoretical at the start.

1.2 Examples of Machine Learning

- Predicting how a viewer will rate a movie
- The essesnce of machine learning, need these 3 components to do machine learning
  - A pattern exists → without there is nothing to look for
  - We can not pin it down mathematically → can not write down a single equation for this
  - We have to have data → no data no learning

1.2.1 Learning Approach

- Each viewer's vector will be different
- Each movie's vector will be different
- combine these 2 to see if a user will like a particular movie
- Machine Learning will reverse engineer this process
  - start with rating and find consistent factors
  - Nudge factors of the vectors to get back to rating ever so slowly
  - Do this not for a single rating but millions. Do it over and over again and eventually the factors become mearningful for the ratings.

1.3 Components of Learning

- Credit Card example
  - bank wants to make money on new cards
  - based on historical data predict how a new customer will do
- Formalization
  - Input X (customer application)
  - Output Y (give credit or no)
  - Target function f: x $\rightarrow$ y (ideal credit approval formula)
  - Data (x_1, y_1), (x_2, y_2) $\ldots$ (x_N, y_N) (historical record)
  - Target function in machine learning is the unknown, solve with data
  - Hypothesis is the formual that approximates the target function g: x $\rightarrow$ y (forumla to be used)
    - g approximates f. g is known f is not!
  - Data used to train the learning algorithm to make g approximate f
  - Learning Algorithm based on data and Hypothesis Set of forumlas (where do these come from?) Guesses at g, learning algorithm will pick the winner.
  - Why have hypothesis set?

        - There is no downside to it you decide how you are learning (linear, neural net, etc...)
        - There is an upside not obvious now. Plays a pivotal role. Lets us know if we can learn.
        - You can do a set of all possible hypothesis

- Solution Components
    - 2 solutions components to learning
        - No control over target function
        - No control over data
        - Final hyposthesis is dictated
        - Learning Algorithm and Hypothesis set are your solution tools!
    - Hypothesis Set. The small h is the function the large one is all of the possible options. g is the selected one.

$$\mathscr{H} = h$$

$$g \in \mathscr{H}$$

- The Learning Algorithm and Hypothesis together are the Learning Model. Many options.

1.4 A simple model

- Simple hypothesis the perception, very simple and not very useful in reality
- For input $\mathrm{x} = (\mathrm{x}_1, \mathrm{x}_2, \ldots \ \mathrm{x}_\mathrm{d})$ are attributes of the customer
- The w vector is weighing which of the inputs x are important.
- approve credit if $\sum_{\mathrm{i}=1}^{\mathrm{d}} \mathrm{w}_\mathrm{i}{*}\mathrm{x}_\mathrm{i} > \mathrm{threshold}$
- deny if below threshold
- this is sort of a credit score
- do not know the w vector or threshold
- The linear formula $\mathrm{h} \in \mathrm{H}$

$$h(x) = sign((\sum_{i=1}^{d} w_i * x_i) - threshold)$$

- Change notation and consider threshold as a weight (w_0 = - threshold)

$$h(x) = sign((\sum_{i=1}^{d} w_i * x_i) + w_0)$$

- Introduce artificial coordinate $x_0 = 1$ to simplify the equation to

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

- In vector form, inner product of column w and vector x

$$h(x) = sign(\mathbf{w}^T\mathbf{x})$$

1.4.1 Perceptron Learning Algorithm

- Implements

$$h(x) = sign(\mathbf{w}^T\mathbf{x})$$

- uses historical data in attempt to make w correct
- pick a misclassified point

$$sign(\mathbf{w}^T\mathbf{x}) \neq y_n$$

- Update the weight (w) vector to be better for this point, y_n is +1 or -1

$$\mathbf{w} \leftarrow \mathbf{w} + y_n * x_n$$

1. Iterations of PLA
   - One iteration, where (x,y) is misclassified

   $$\mathbf{w} \leftarrow \mathbf{w} + y * x$$

   - At iteration t=1,2,3… pick a misclassified point from (x_1, y_1), (x_2, y_2) … (x_N, y_N) and run a PLA iteration on it
   - That's it!
   - Here is an implementation from machine learning master

1.5 Types of Learning

1.5.1 Basic Premise of Learning

- using a set of observations to uncover and underlying process
- very broad, leads to many variations
- Types
  - Supervised Learning – concentation of this course
    - Any time the data and output are explicitly given, like a supervisor is helping you out
  - Unsupervised learning
    - we get input data and no outputs. Like listening to another language on radio in an effort to learning it
  - Reinforcement Learning
    - get the input data and *some* of the output and grade of your output. Great for playing games

1.6 Puzzle

- Superivised Learning puzzle
  - I guess -1
  - Doesn't matter, wants to get both answers and impossible to answer this particular problem

1.7 Q & A

1.7.1 How to determine if linear seperable?

- Rarely true, good for examples
- Techniques to make it true
- Assume this is false
- pocket algoritm?

1.7.2 How do you know if there is a pattern?

- You don't
- Covered in a future lecture
- Take data, apply algorithm and you can detect if you learn or not and knowing this
- avoid looking at data

1.7.3 Global optimization or local optimization?

- Whichever works for us

1.7.4 Hypothesis continuous or discrete

- Can be either

1.7.5 How much data for a particular problem?

- Theory: this is the crux of theory
- Practical: not under your control

2 Lecture 01

3 Lecture 01

4 Lecture 01

5 Lecture 01

6 Lecture 01

7 Lecture 01

8 Lecture 01

9 Lecture 01

10 Lecture 01

11 Lecture 01

12 Lecture 01

13 Lecture 01

14 Lecture 01

15 Lecture 01

16 Lecture 01

17 Lecture 01

18 Lecture 01