How Machines Learn

a crash course

Bryan Healey (bryan@bryanhealey.com)

Who Am I?

Northeastern University, CS / Cognition

Recommendations at ShapeUp

Big Data, ASR, NLU at Amazon (Alexa)

AI and Data Science at Lola Travel

What is machine learning?

Making decisions without explicit programming

Learning by example

Data is everything

So many options, so little time...

What is this good for?

Big data analysis and classification

Natural language processing

Broad and complex decision making

Continuous learning and autonomous behavior

How is it done?

Very wide and diverse field

Two “buckets” standout:

- Neural Networks
- Statistical Modeling

Neural Networks

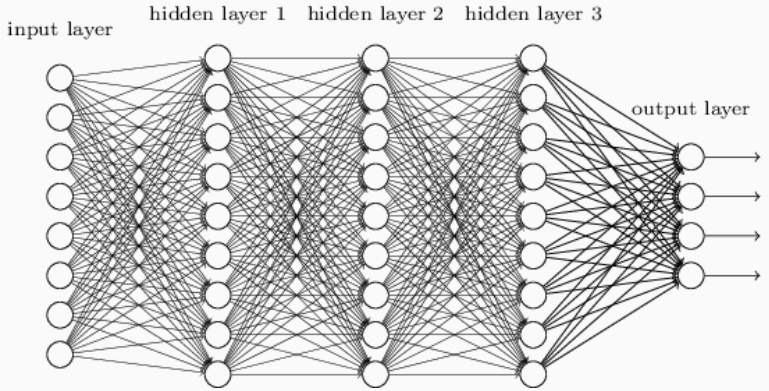
The hot new thing (but actually quite old)

Many different forms

ANN, DNN, RNN, CNN, etc

Simple decisions, aggregated

Collective decision making



Vacation?

Q: Should you vacation in Paris?

Consider important questions, and how answers interrelate

- Can I fly in the morning?
- Can I afford this trip without debt?
- Is there a nice hotel in my price range?
- Is it going to be warm when I arrive?

Vacation?

Q: Should you vacation in Paris?

Consider important questions, and how answers interrelate

- Can I fly in the morning? **4**
- Can I afford this trip without debt? **9**
- Is there a nice hotel in my price range? **6**
- Is it going to be warm when I arrive? **8**

Vacation?

Scenario 1:

- Can I fly in the morning? **YES (+4)**
- Can I afford this trip without debt? **NO**
- Is there a nice hotel in my price range? **YES (+6)**
- Is it going to be warm when I arrive? **NO**

Final score: 10

Vacation?

Scenario 2:

- Can I fly in the morning? **NO**
- Can I afford this trip without debt? **YES (+9)**
- Is there a nice hotel in my price range? **NO**
- Is it going to be warm when I arrive? **YES (+8)**

Final score: 17 *(better?)*

Vacation?

Large training set required

Largely free from feature engineering (a plus)

Tuning of the neurons, thresholds, and weights is the fun part

Very simplified….

Statistical Modeling

The boring standard (but still everywhere)

Many different forms

Linear Regression, Bayesian, Markov, etc

Supervised, semi-supervised

A lot more math….

$$P(class|word) = \frac{exp(\overrightarrow{weight} \cdot \overrightarrow{feature}(word,class))}{\sum_{c=All\ Classes} exp(\overrightarrow{weight} \cdot \overrightarrow{feature}(word,c))}$$

Names and Genders

“What's in a name? that which we call a rose
By any other name would smell as sweet!”

Names and Genders

- Michael (male)
- Stacey (female)
- Paul (male)
- David (male)
- Brenda (female)

Names and Genders

- Michael (male) L, E, **false**
- Stacey (female) Y, E, true
- Paul (male) L, U, **false**
- David (male) D, I, **false**
- Brenda (female) A, D, true

Change the world?

Greatly simplified, but the logic is sound

Data is everything

No problem is too big or too complex

Questions?

Ready for Code?

Let’s see some action!