ici : Predicting the Future

Predicting the Future

MLaaS

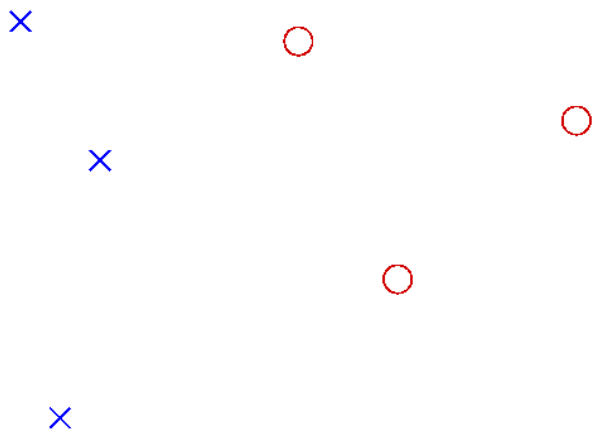
Jeff Abrahamson

Jellybooks

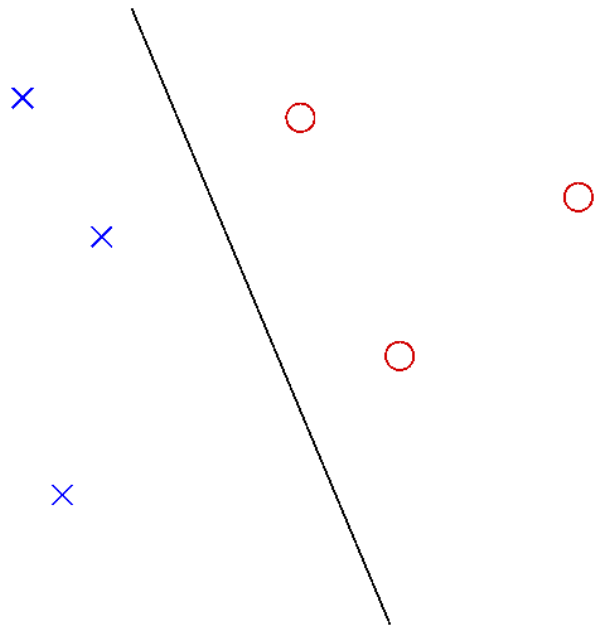
Prediction is very difficult, especially about the future.

Niels Bohr (1970) (or Yogi Bera or Piet Hein, or Steincke (1948)), but probably Markus M. Ronner (1918)

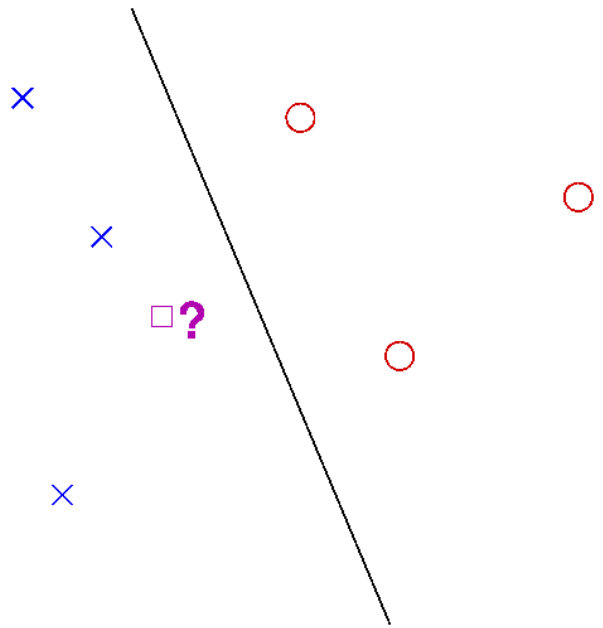
What's a simpler problem?



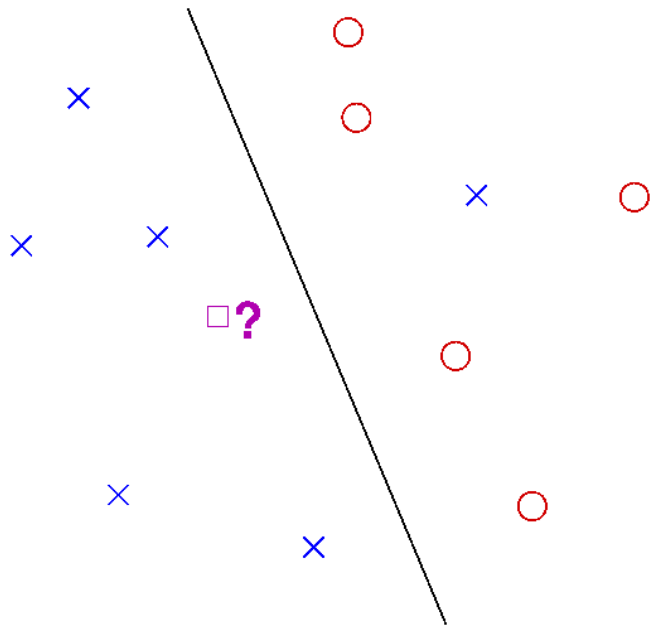
What's a simpler problem?



What's a simpler problem?



What's a simpler problem?



Training a Model

Known points, draw a line.
Model: the line

Predict

New point, which side of line?

Evaluation

Example :

- Rare illness, .001% of population each year
- Treatable early, but expensive and hurts
- Or. . . dead in a month

Evaluation

Example :

- Rare illness, .001% of population each year
- Treatable early, but expensive and hurts
- Or. . . dead in a month

Training: past cases (know outcomes)

Evaluation

Example :

- Rare illness, .001% of population each year
- Treatable early, but expensive and hurts
- Or. . . dead in a month

```
# First pass, reliability = 99.999%
def has_disease(person):  return false
```

Examples

- Credit scoring
- Fraud detection
- Personalization
- Spam on website
- Recommend a book, a movie, a holiday

Services

- Amazon Machine Learning (free: 12 months)
- Google Prediction API $\subset$ Cloud Platform (free: 6 months)
- PredicSis
- BigML (free: in development mode)
- Apache Mahout (free: beer and speech)
- Scikit-learn (free: beer and speech, pipeline sold separately)

Options

- Batch Prediction API
- Real-Time API

Amazon-speak

Comparison

- BigML and PredicSis faster
- Amazon and PredicSis slightly more accurate
- YMMV

http://www.louisdorard.com/blog/machine-learning-apis-comparison

Is it Magic?

Q: What algorithm does Amazon Machine Learning use to generate models?

Amazon Machine Learning currently uses an industry-standard logistic regression algorithm to generate models.

Is it Magic?

- Clean data
- Feature engineering
- (Sometimes) Choose a model
- Train model
- Send queries, receive predictions

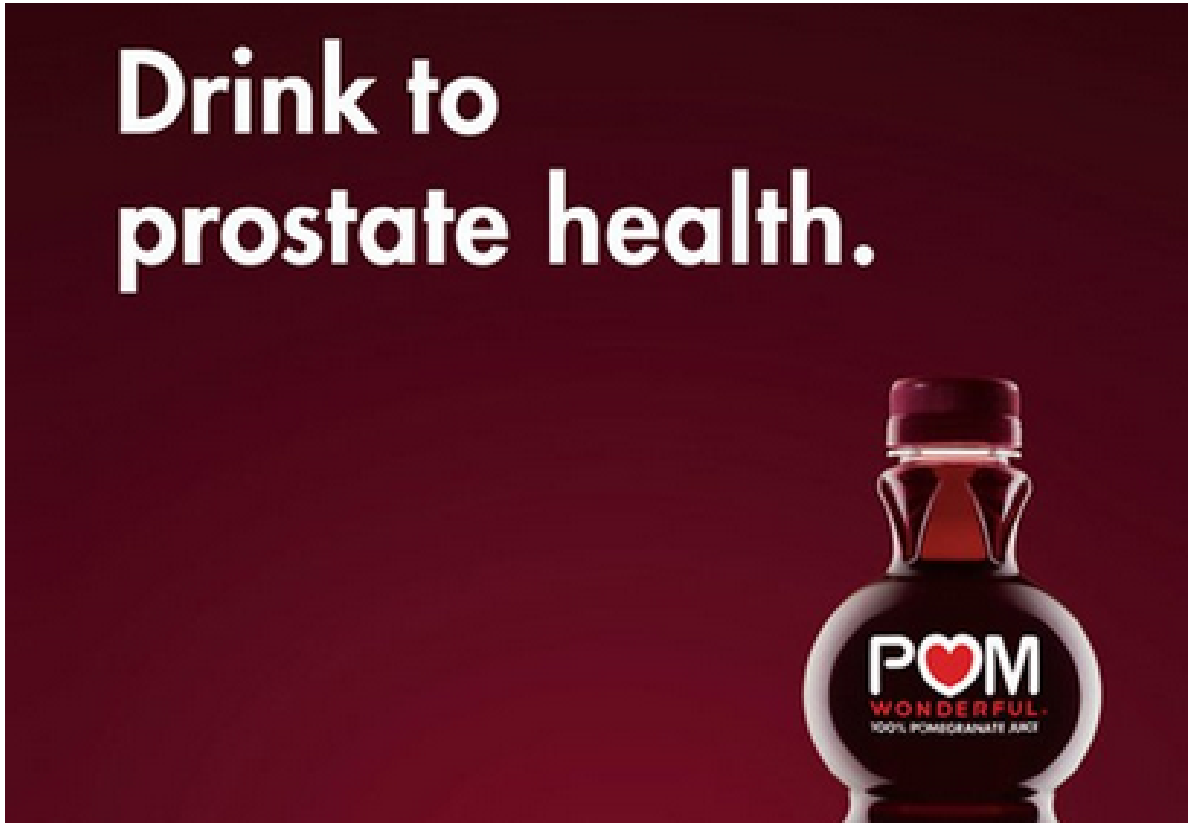
Is it Magic?

- Clean data
- Feature engineering
- (Sometimes) Choose a model
- Train model
- Send queries, receive predictions

Free pipeline (except scikit-learn)

Is it Real?

Is it Real?



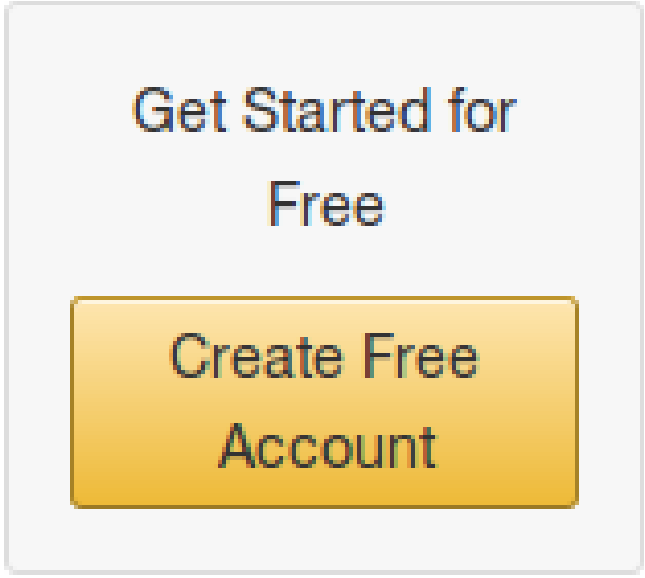
Is it Real?

Benefits

Easily Create Machine Learning Models

Amazon Machine Learning APIs and wizards make it easy for any developer to create and fine-tune ML models from data stored in Amazon Simple Storage Service (Amazon S3), Amazon Redshift, or MySQL databases in Amazon Relational Database Service (Amazon RDS), and query these models for predictions. The service's built-in data processors, scalable ML algorithms, interactive data and model visualization tools, and quality alerts help you build and refine your models quickly.

Free as in Beer



Free as in Speech

Q: Can I export my models out of Amazon Machine Learning?

No.

Q: Can I import existing models into Amazon Machine Learning?

No.

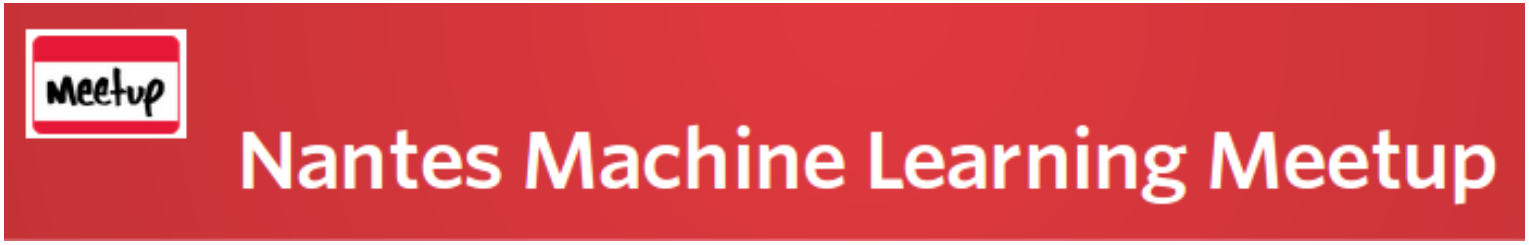
Reflections

- Data is most of the work: clean, select features, understand
- Choosing models requires knowledge
- Big data sets aren't always big
- If you need performance, *maybe*

More Reading

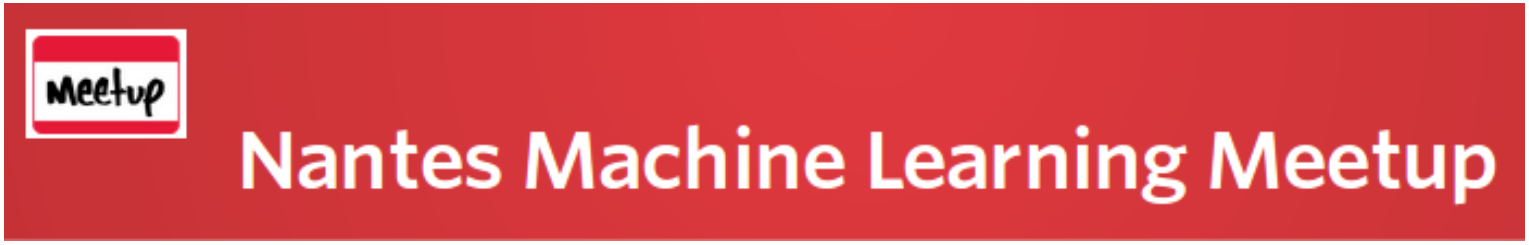
http://www.programmableweb.com/news/6-questions-you-should-ask-about-prediction-apis/analysis/2014/09/12

Sounds Easy?



http://www.meetup.com/Nantes-Machine-Learning-Meetup/

Sounds Easy?



http://www.meetup.com/Nantes-Machine-Learning-Meetup/

And there are jobs.

Sounds hard?

2–6 novembre 2015

http://ml-week.com/

Questions?

Feedback: `http://purple.com/1`

`https://github.com/JeffAbrahamson/talks/`