Lecture 0
Introduction to Machine Learning

I. Tinnirello

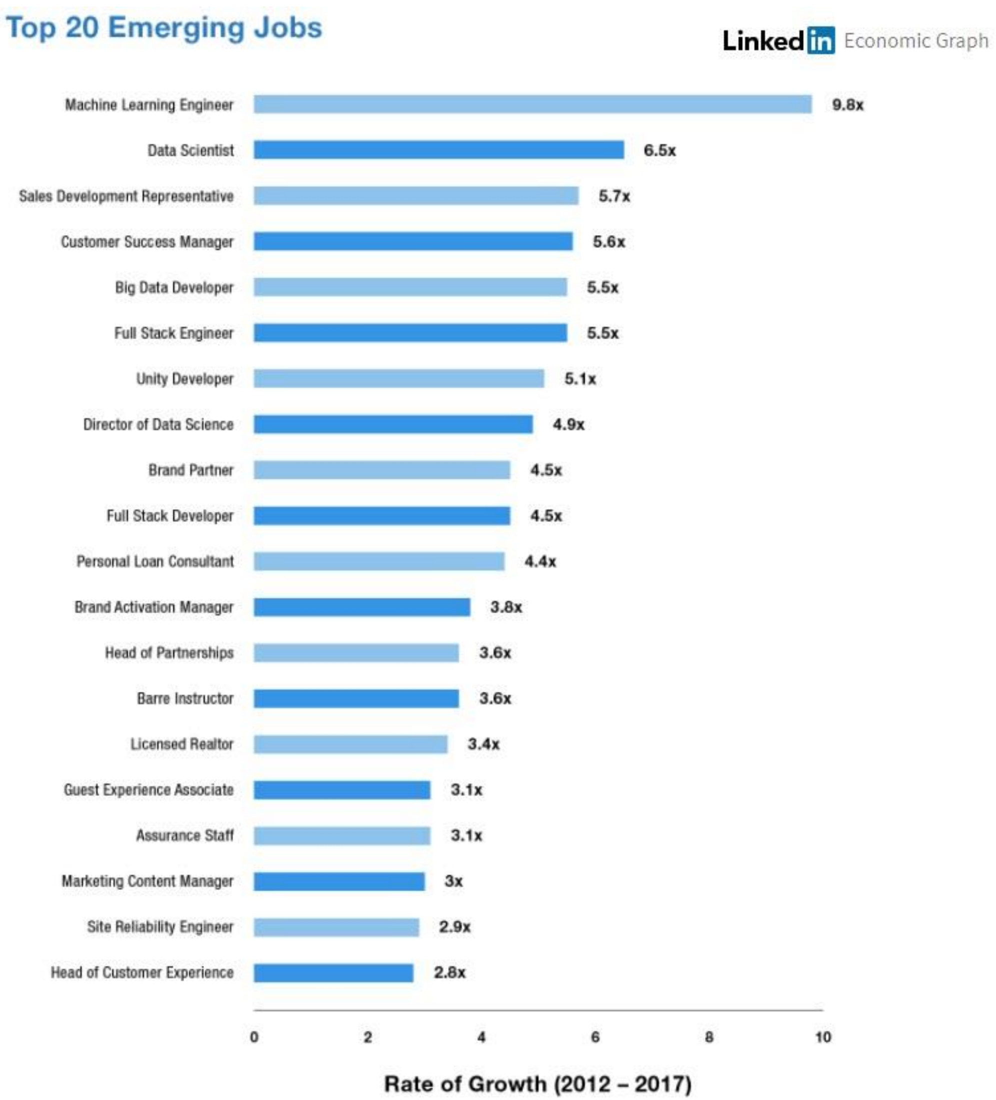
Machine Learning: why?

- ➔ We are entering the era of big data
  - ⇨ 1 trilion of web pages
  - ⇨ 1 hour of video on you tube every second
  - ⇨ thousands of DNA genomes
- ➔ Can we automatically detect patterns on data?
  - ⇨ Predict future data
  - ⇨ Take decisions
  - ⇨ Extract knowledge

Demand of machine learning experts

- Machine Learning Engineers, Data Scientists, and Big Data Engineers rank among the top emerging jobs on LinkedIn. (january 2018)
- Data scientist roles have grown over 650% since 2012, but currently, 35,000 people in the US have data science skills
- Job growth in the next decade is expected to outstrip growth during the previous decade, creating 11.5M jobs by 2026, according to the U.S. Bureau of Labor Statistics.

The job trend



A Few Quotes

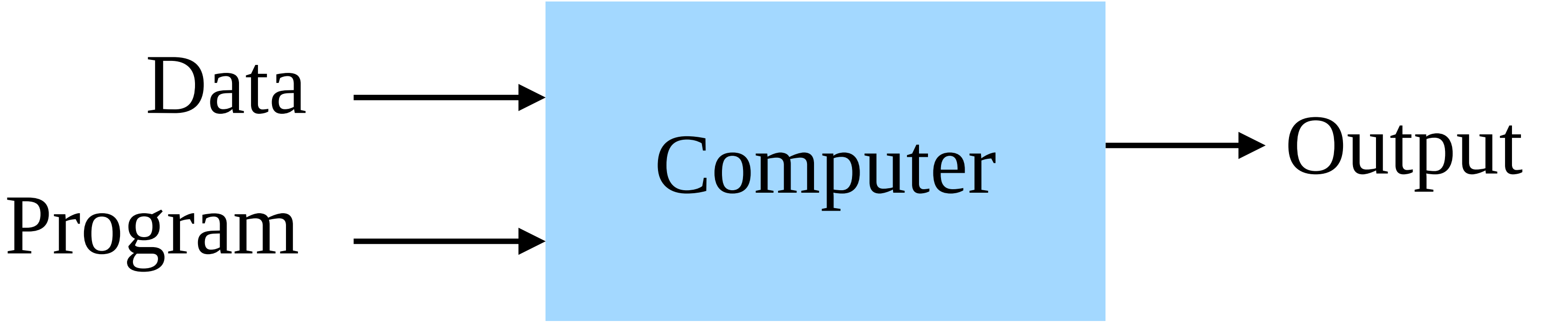
- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet” (Tony Tether, Director, DARPA)
- Machine learning is the hot new thing” (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, CTO, Sun)
- “Machine learning is today’s discontinuity” (Jerry Yang, CEO, Yahoo)

From Pedro Domingos, Washington Univ.

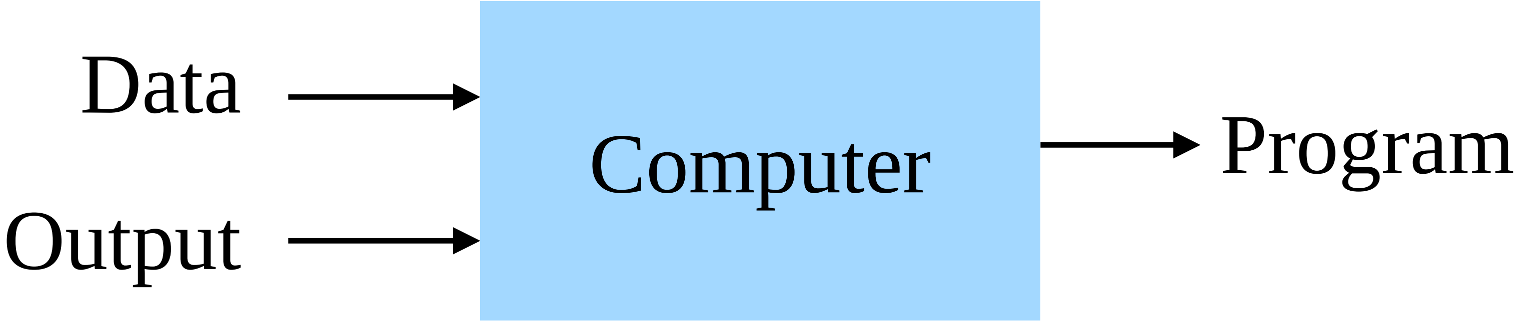
So What is Machine Learning?

- ➔Automating automation
- ➔Getting computers to program themselves
- ➔Writing software is the bottleneck
- ➔Let the data do the work instead!

Traditional Programming



Machine Learning



Sample Applications

- ➔ Web search
  - ⇨ Which page is more important?
- ➔ Computational biology
  - ⇨ Which genes are responsible of a given problem?
- ➔ E-commerce
  - ⇨ Which product is of interest for a customer?
- ➔ Robotics
  - ⇨ How should the robot move the arm?
- ➔ Social networks
  - ⇨ Which candidate friends?
- ➔ Speech recognition, computer vision
  - ⇨ Who is this face?
- ➔ [Your favorite area]

ML in a Nutshell

- ➔ From algorithms and data, build a model/program automatically
  - ⇨ Tens of thousands of machine learning algorithms
  - ⇨ Hundreds new every year
- ➔ Every machine learning algorithm has three components:
  - ⇨ Representation
  - ⇨ Evaluation
  - ⇨ Optimization

Representation

- ➔Decision trees
- ➔Sets of rules / Logic programs
- ➔Instances
- ➔Graphical models (Bayes/Markov nets)
- ➔Neural networks
- ➔Support vector machines
- ➔Etc.

Evaluation

- ➔Accuracy
- ➔Squared error
- ➔Likelihood
- ➔Posterior probability
- ➔Cost / Utility
- ➔Margin
- ➔Entropy
- ➔Etc.

Optimization

- ➔ **Combinatorial optimization**
  - ⇨ E.g.: Greedy search
- ➔ **Convex optimization**
  - ⇨ E.g.: Gradient descent
- ➔ **Constrained optimization**
  - ⇨ E.g.: Linear programming

Types of Learning

- **Supervised (inductive) learning**
  - Training data includes desired outputs
- **Unsupervised learning**
  - Training data does not include desired outputs
- **Semi-supervised learning**
  - Training data includes a few desired outputs
- **Reinforcement learning**
  - Rewards from sequence of actions

Inductive Learning

- ➔ **Given examples of a function *(X, F(X))***
- ➔ **Predict function *F(X)* for new examples *X***
  - ⇨ Discrete *F(X)*: Classification
  - ⇨ Continuous *F(X)*: Regression
  - ⇨ *F(X)* = Probability(*X*): Probability estimation

Text Categorization
(multi category)

Categorize text documents into predefined categories.

For example, ‘sports’, ‘politics’, ‘science’, etc.

Soft tissue found in T-rex fossil
Find may reveal details about cells and blood vessels of dinosaurs
Th
WA
dir
res
inc
Tyran

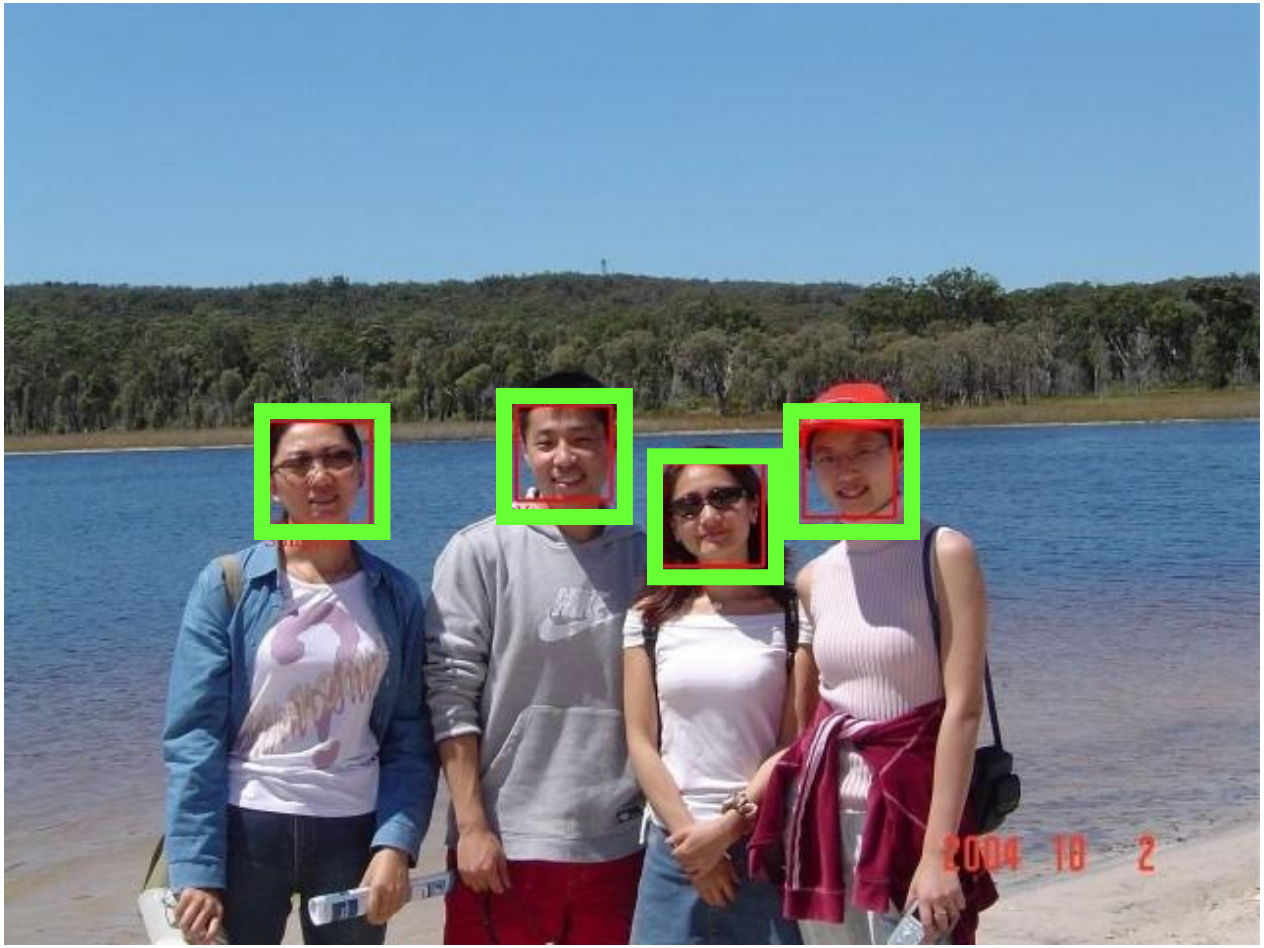
Health may be concern when giving kids cell phones
Wednesday, March 23, 2005 Posted: 11:14 AM EST
SEA
bef
pho
risk

Wall Street gears up for jobs
Saturday, March 26, 2005: 11:41 AM EST

NE
200
dep

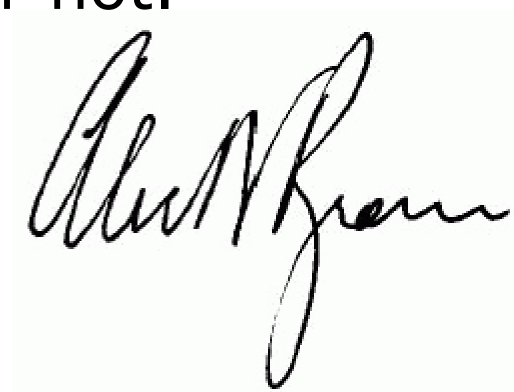
Probe finds atmosphere on Saturn moon
Thursday, March 17, 2005 Posted: 11:17 AM EST
LOS ANGELES, California (Reuters) -- The space probe Cassini discovered a significant atmosphere around Saturn's moon Enceladus during two recent passes close by, the Jet Propulsion Laboratory said on Wednesday

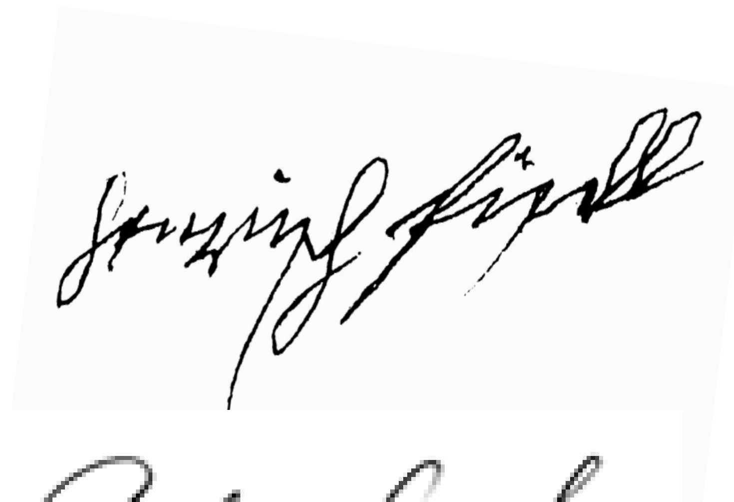
Face detection



Signature recognition

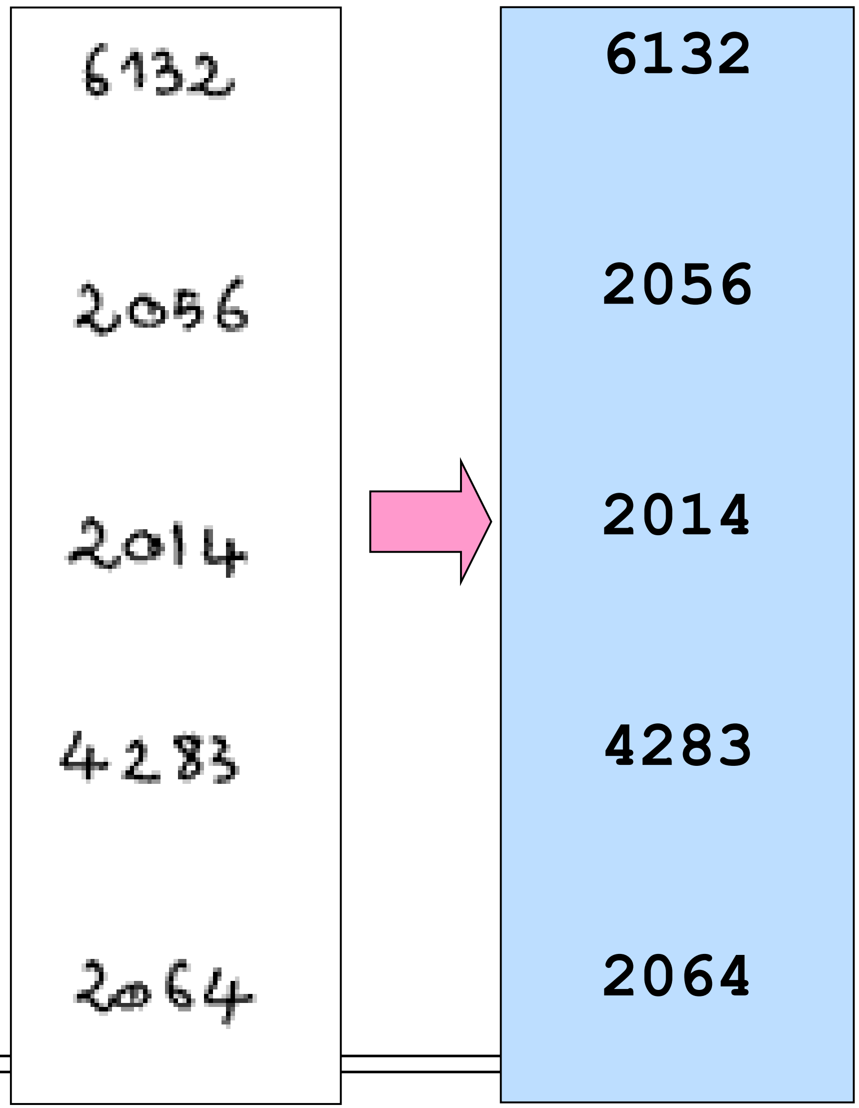
- ➔ Recognize signatures by structural similarities which are difficult to quantify.
- ➔ does a signature belongs to a specific person or not.





Character recognition
(multi category)

➔Identify handwritten characters: classify each image of character into one of 10 categories '0', '1', '2' …



What we will cover

- ➔ **<u>Supervised learning</u>**
  - ⇨ Bayesian learning
  - ⇨ Linear Classifiers
  - ⇨ Neural networks
  - ⇨ Support vector machines (just an introduction)
  - ⇨ Decision trees
  - ⇨ Markov Chains
    - → PageRank Google's core
- ➔ **Unsupervised learning**
  - ⇨ Clustering (just an introduction)
  - ⇨ Dimensionality reduction

Pre-requisites

- ➔ **Basics of calculus, algebra**
- ➔ **Basics of probability**
  - ⇨ Do we need a review?
- ➔ **Basics of programming!**

Teaching Materials

- ➔ **Text Books**
  - ⇨ Sergios Theodoridis. *Pattern Recognition*
  - ⇨ Kevin P. Murphy. *Machine Learning, a probabilistic perspective*
- ➔ **Lecture Notes**
  - ⇨ Slides/lesson notes
  - ⇨ <u>A lot of material available in internet!</u>
    - → e.g. real projects in python, such as Sebastian Raschka *Python Machine Learning* or Peter Harrington *Machine Learning in Action*

Final Exam

- ➔**Written test (mandatory)**
  - ⇨Simple exercises and theoretical questions
  - ⇨up to 30/30 with laude
- ➔**Practical project (optional)**
  - ⇨Designing a solution for a real problem and data set
  - ⇨up to 3 additional points

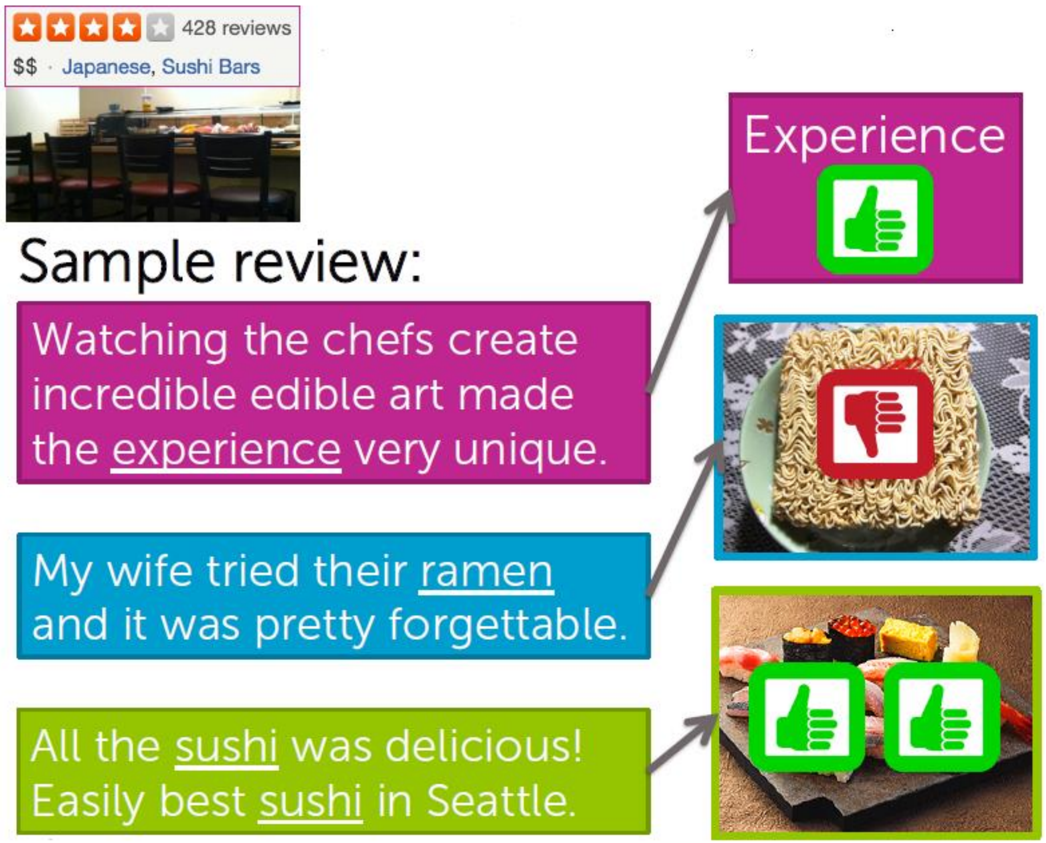
Introduction to classification

Classification Problems

- ➔Spam or not spam?
- ➔Is it a cat or a dog image?
- ➔Is it a male or a female speaker?
- ➔Is it a good or a bad restaurant?
- ➔..

An example of review systems

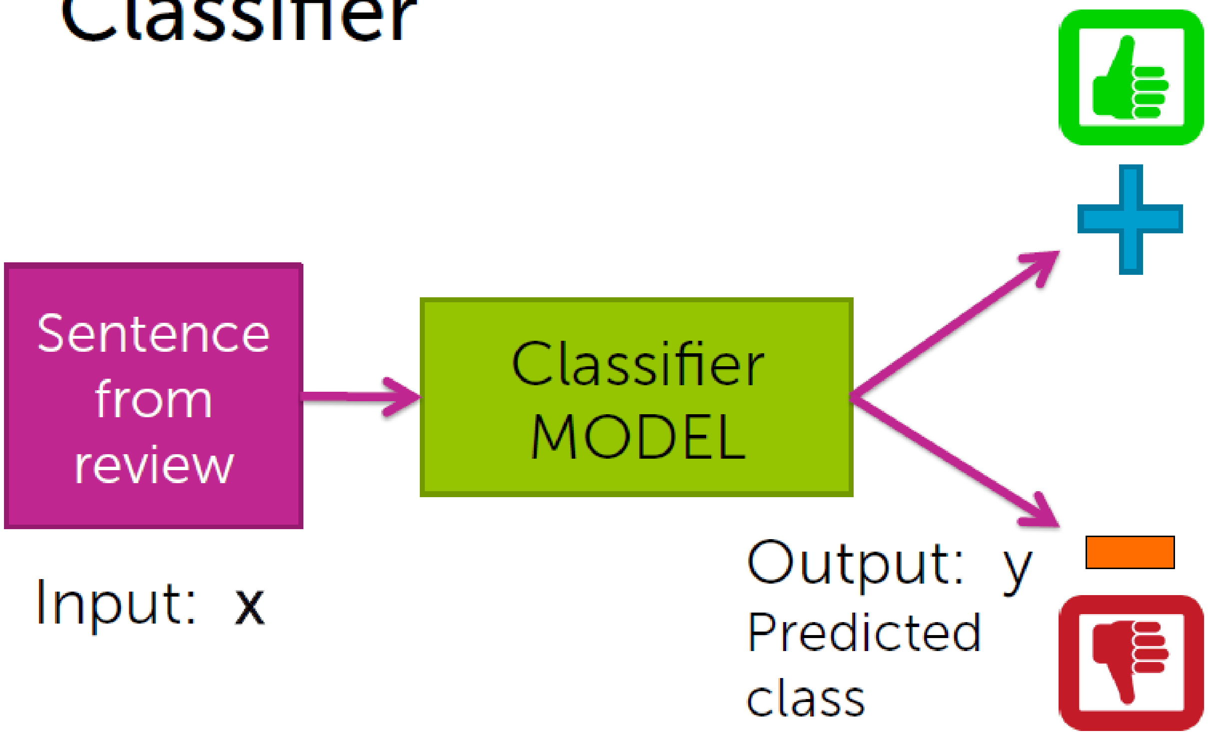
➔Reviews are not totally good or totaly bad



Example by
Emily Fox & Carlos Guestrin

An automatic review system

Classifier



Linear/Thresholds Classifiers

List of positive words	List of negative words
great, awesome, good, amazing,…	bad, terrible, disgusting, sucks,…

Sentence from review

Input: x

Simple threshold classifier

Count positive & negative words in sentence

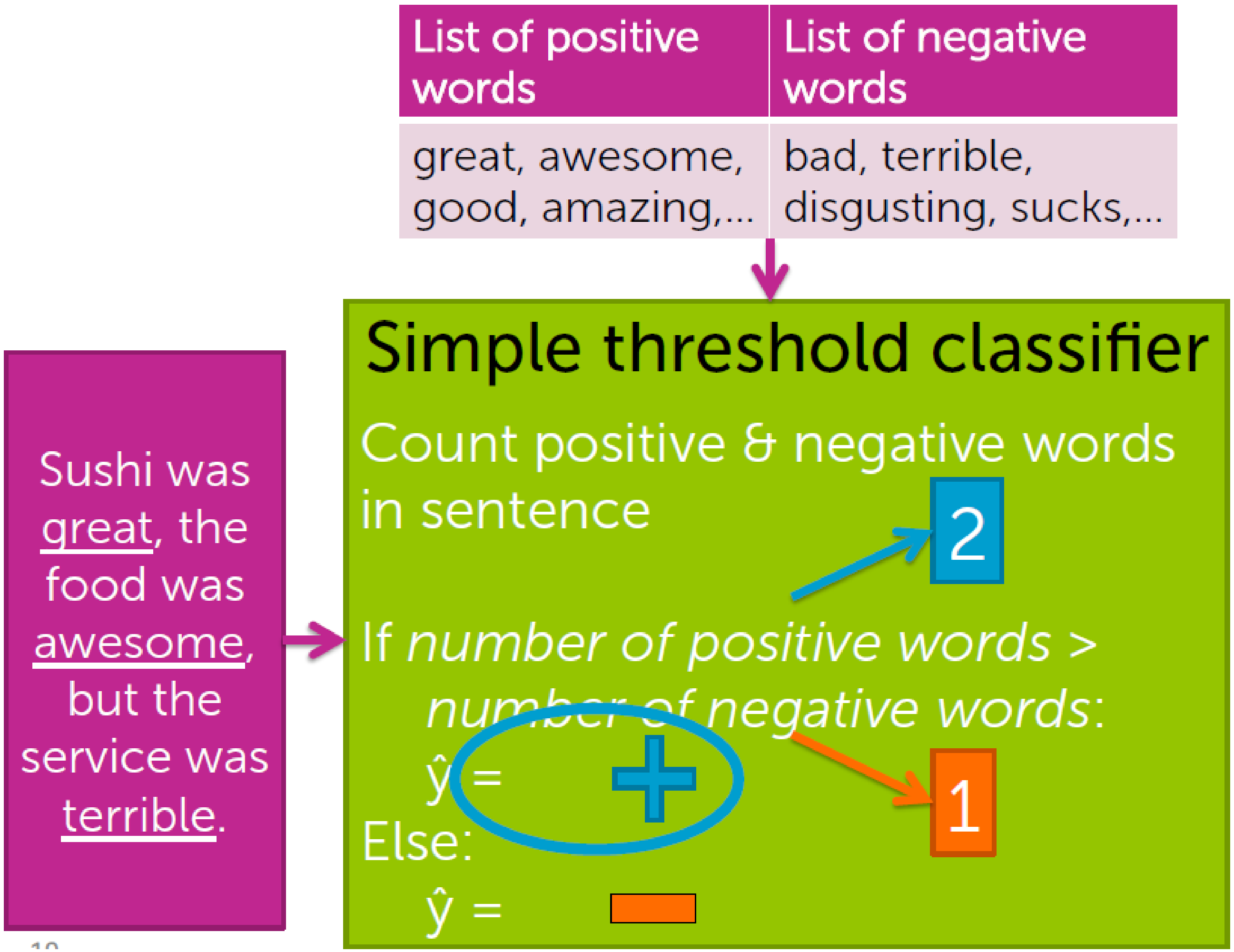
If *number of positive words > number of negative words*:

$\hat{y} = +$

Else:

$\hat{y} = -$

How does it work



Problems with thresholds classifiers

- ➔ **How do we get list of positive/negative words?**
- ➔ **Words have different degrees of sentiment**
  - ⇨ Great>good
  - ⇨ How do we weight different words?
- ➔ **Single words are not enough**
  - ⇨ Good -> positive
  - ⇨ Not good -> negative
- ➔ ***We need to train the classifier and work on complex features!***

A linear classifier

➔ **We use training data to learn the weight of each word**

Word	Weight
good	1.0
great	1.5
awesome	2.7
bad	-1.0
terrible	-2.1
awful	-3.3
restaurant, the, we, where, …	0.0
…	…

Scoring a sentence

Word	Weight
good	1.0
great	1.5
awesome	2.7
bad	-1.0
terrible	-2.1
awful	-3.3
restaurant, the, we, where, ...	0.0
...	...

- **Input x:**
  - Sushi was great, the food was awesome but the service was terrible
- **Score(x)= 1.5 + 2.7 – 2.1= 2.1**
- **Output y:**
  - If score(x)>0 +
  - If score(x)<0 -

Characterizing an audio signal

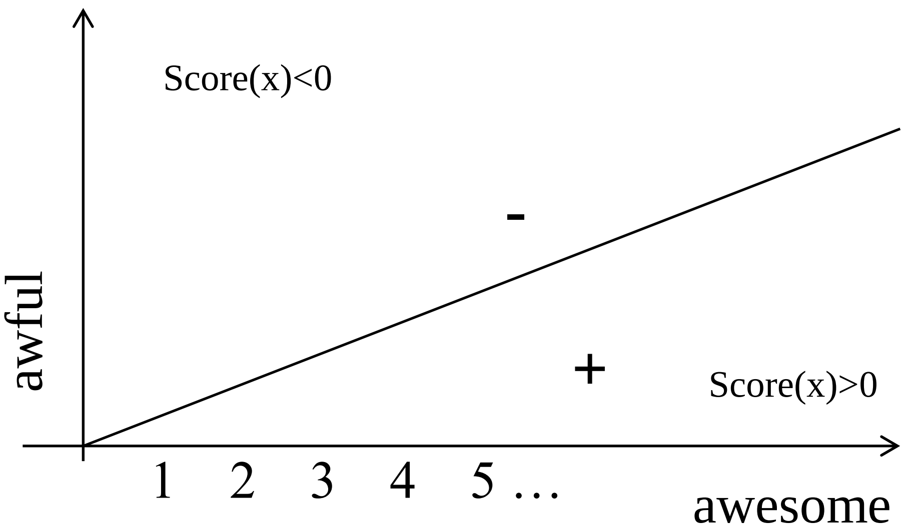
- ➔**r = audiorecorder(22050, 16, 1);**
- ➔**recordblocking(r, 5);**
- ➔**mySpeech = getaudiodata(r, 'int16');**
- ➔**mySpeech = getaudiodata(r);**
- ➔**[a g]=lpc(mySpeech, 3)**
  - ⇨Feature extraction
  - ⇨est_w=filter([0 -a(2:end)],1, w1);
    - ➔At each step n, coefficient n-th is estimated as a function of the previous n-1;

Feature extraction for audio

- **Linear predictive coding**
  - E.g. third order
  - Learning, machine
- **Threshold-based classification**
- **Other possible features**
  - MFCC
  - Wavelet transforms
  - FFT

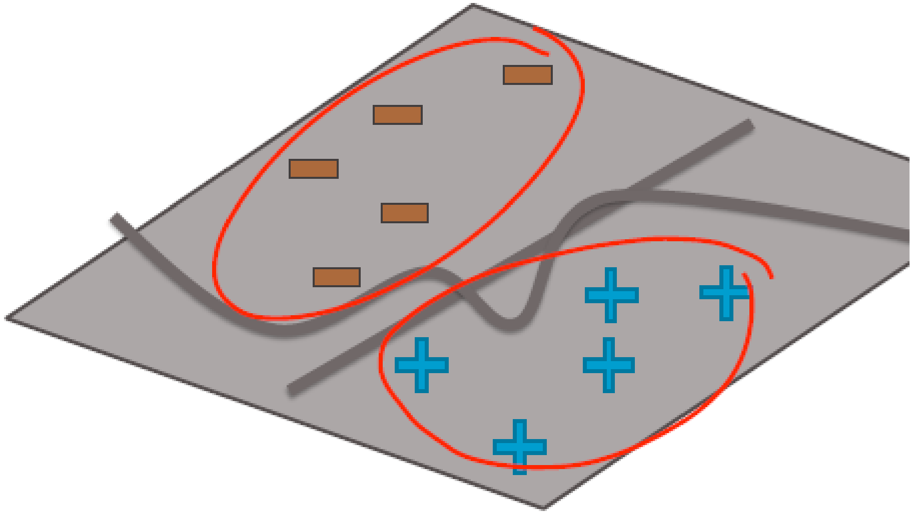
Decision boundary

- ➔ **Suppose only two words have non zero weight**
  - ⇨ Awesome = 1.0; Awful = -1.5
- ➔ **Score(x)=1.0 #awesome – 1.5#awful**
- ➔ **Score(x)>0 if #awful < 2/3 #awsome**



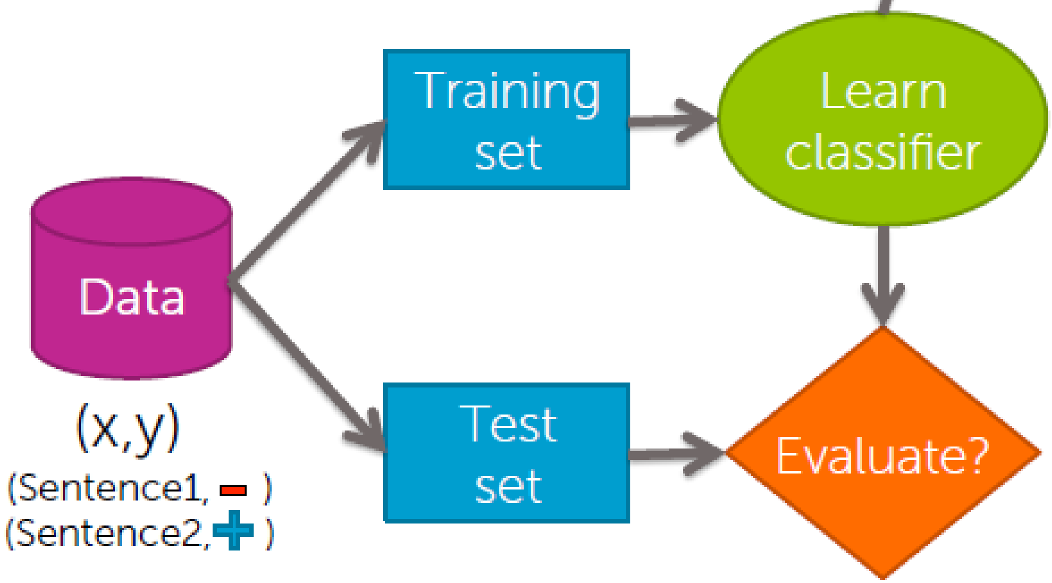
Decision Boundary (2)

- ➔For linear classifiers, they are lines, planes or hyperplanes
- ➔For more general classifiers, more complicated shapes



Training a Classifier

finding the weights



Word	Weight
good	1.0
awesome	1.7
bad	-1.0
awful	-3.3
...	...

Classification Errors

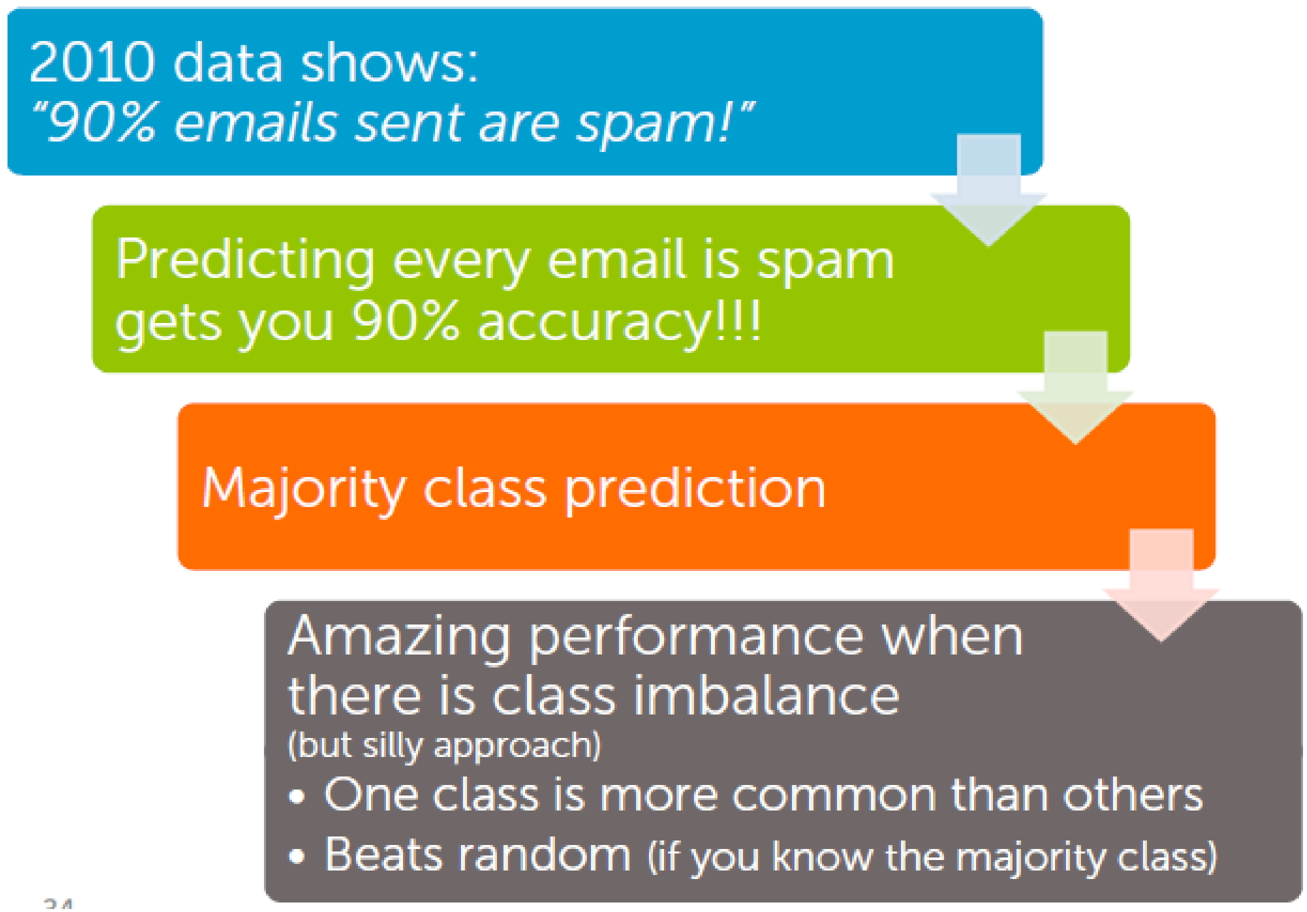
- ➔ **Errors measure fraction of mistakes (#mistakes/#sentences)**
- ➔ **Accuracy is the fraction of correct predictions (#correct/#sentences)**

- ➔ **What if you ignore the sentence and just guess?**
  - ⇨ For binary classification, half the time you will get it right!
  - ⇨ For k classes, accuracy =1/k
- ➔ **What is a good accuracy?**

At the very, very, very least,
you should healthily beat random…
Otherwise, it's (usually) pointless…

Is a classifier with 90% accuracy good?

Depends!!



Types of mistakes
(confusion matrix)

	Predicted Label: +	Predicted Label: −
Ture Label: +	True positive	False negative
Ture Label: −	False positive	True negative

Accuracy=0.85

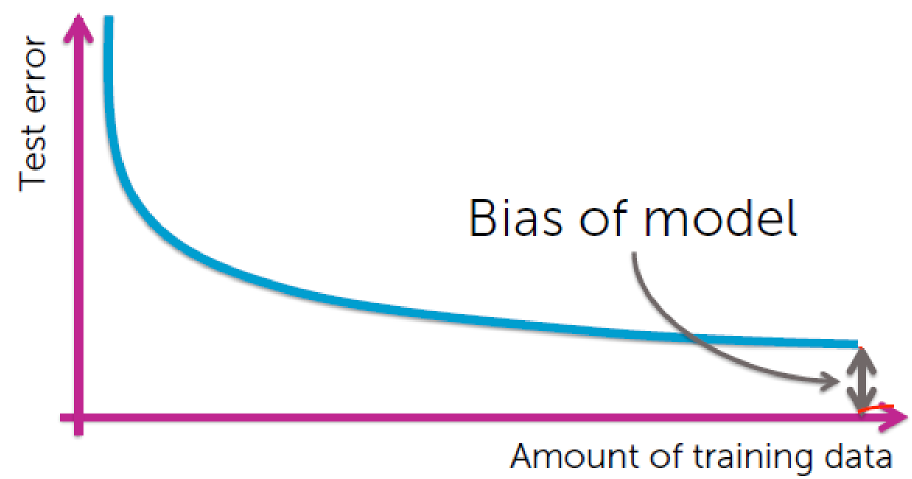
	+	−
+	50	10
−	5	35

Cost of errors

	Spam Filtering	Medical Diagnosis
False negative	Annoying	Disease not treated
False positive	Email lost	Wasteful treatment

How much data does a model need?

- ➔ **The more the merrier**
  - ⇨ But data quality is the most important factor!
- ➔ **Theoretical techniques sometimes can bound how much data is needed**
- ➔ **In practice:**
  - ⇨ More complex models require more data
  - ⇨ Empirical analysis



More complex models, less bias

Sentiment classifier using single words can do OK, but...

Never classifies correctly:
"The sushi was not good."

More complex model:
consider pairs of words (bigrams)

Word	Weight
good	+1.5
not good	-2.1

Less bias →
potentially more accurate,
needs more data to learn