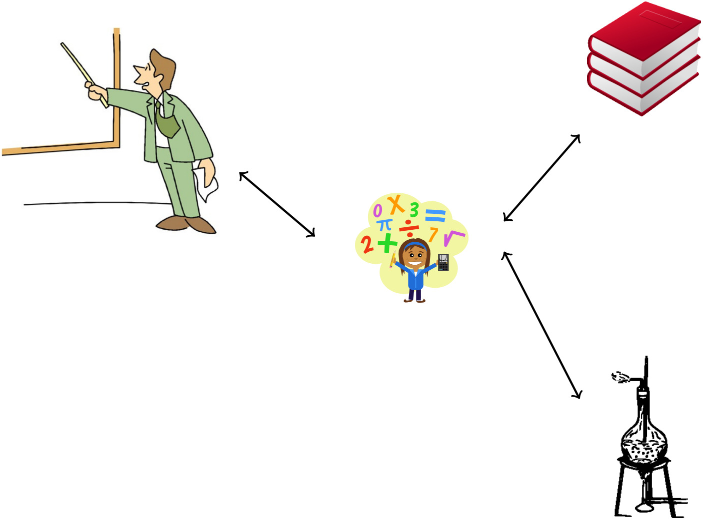
Machine learning - HT 2016
1. Introduction

Varun Kanade

University of Oxford
January 20, 2016

What is machine learning?



Machine learning and Artificial intelligence

What does intelligence entail?

- Reasoning, planning, representation, learning
- Courses: Intelligent Systems (MT), Knowledge Representation & Reasoning (HT)
- Open AI Initiative: `http://openai.com/`

Outline

History of Machine Learning

This Class

Some Machine Learning Applications

Some Practical Concerns

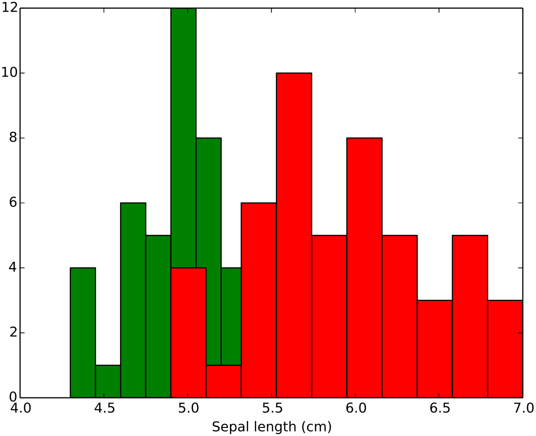
History of Machine Learning

Statistics: Ronald Fisher

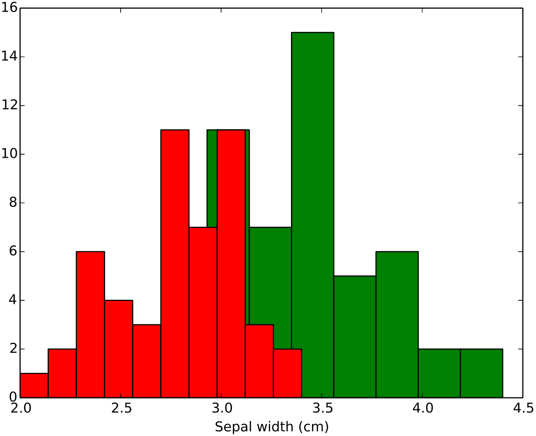
- Three types of iris: setosa, versicolour, virginica (1936)
- For each flower: sepal width (x_1), sepal length (x_2), petal width (x_3), petal length (x_4)



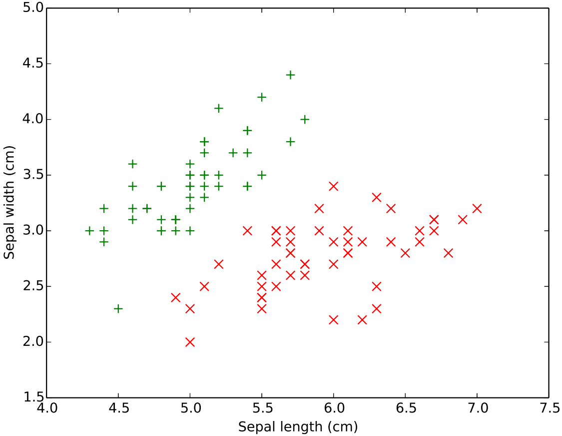
Visualize Iris Data: Setosa vs Versicolor



Visualize Iris Data: Setosa vs Versicolor



Visualize Iris Data: Setosa vs Versicolor



History of Machine Learning

Statistics: Ronald Fisher

- Three types of iris: setosa, versicolour, virginica (1936)
- For each flower: sepal width (x_1), sepal length (x_2), petal width (x_3), petal length (x_4)
- Find $X = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4$ that maximizes D^2/S
- $D = \sum_{i=1}^{4} w_i(\mathbb{E}[x_i \mid \text{setosa}] - \mathbb{E}[x_i \mid \text{versicolour}])$
- $\mu_i = \mathbb{E}[x_i]$
- $S = \sum_{i=1}^{4} \sum_{j=1}^{4} w_i w_j \mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)]$
- We will see that this is basically linear regression



History of Machine Learning

Computer Science: Alan Turing

Turing Test: The Imitation Game

Learning Machines (Computing Machinery and Intelligence. *Mind* (1950))

"Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain."

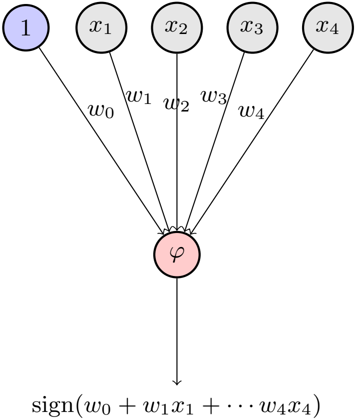
Quantitative computational considerations

- How much memory would be required?
- How much computational power would be required?

History of Machine Learning

Neuroscience: Frank Rosenblatt

- Perceptron - neurally-inspired
- Simple training (learning) algorithm
- Built using specialized hardware



Perceptron Training Algorithm

Setting

- Get a sequence of points $(\mathbf{x}_t, y_t)$ (where only $\mathbf{x}_t$ is observed at first)
- After prediction is made y_t is revealed
- Start with $\mathbf{w}_0$ some arbitrary starting weights for the perceptron

Perceptron Training Algorithm

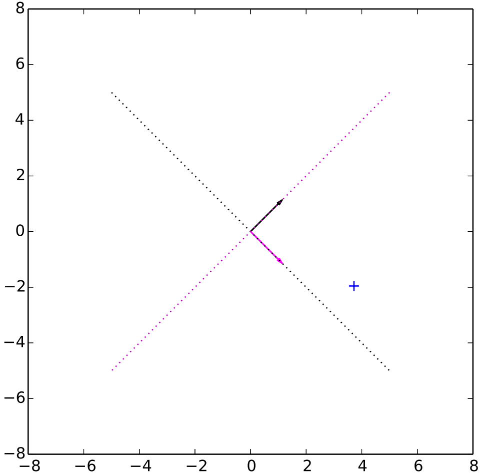
Setting

- Get a sequence of points $(\mathbf{x}_t, y_t)$ (where only $\mathbf{x}_t$ is observed at first)
- After prediction is made y_t is revealed
- Start with $\mathbf{w}_0$ some arbitrary starting weights for the perceptron

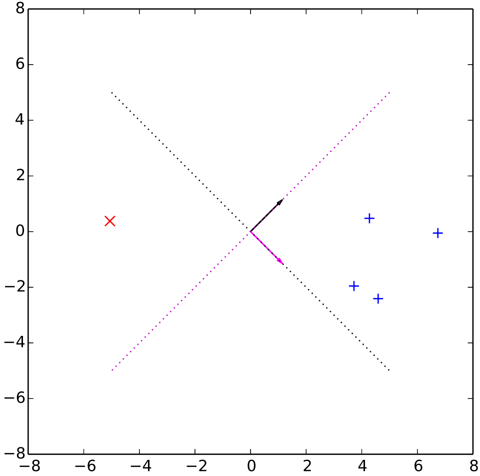
Algorithm

1. Suppose $\mathbf{w}_{t-1}$ are the weights after $t-1$ steps
2. Predict $\hat{y}_t = \mathrm{sign}(\mathbf{w}_{t-1} \cdot \mathbf{x}_t)$
3. Update:
    - If $\hat{y}_t = y_t$; do nothing
    - Else set $\mathbf{w}_t = \mathbf{w}_{t-1} - \eta(1 - 2y_t)\mathbf{x}_t$

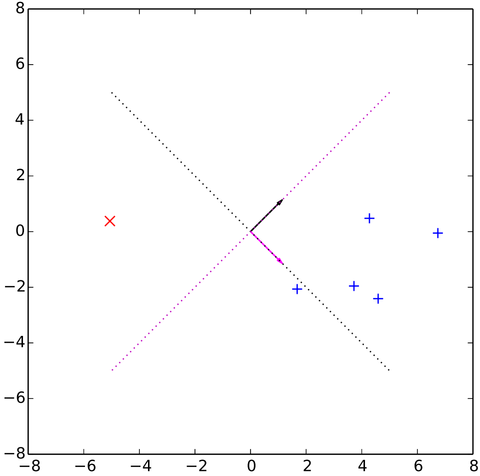
Perceptron Training Algorithm in Action



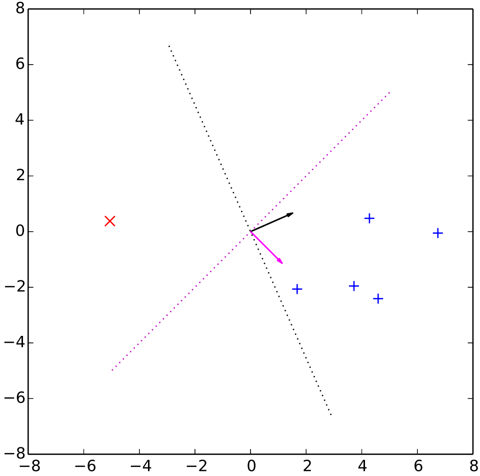
Perceptron Training Algorithm in Action



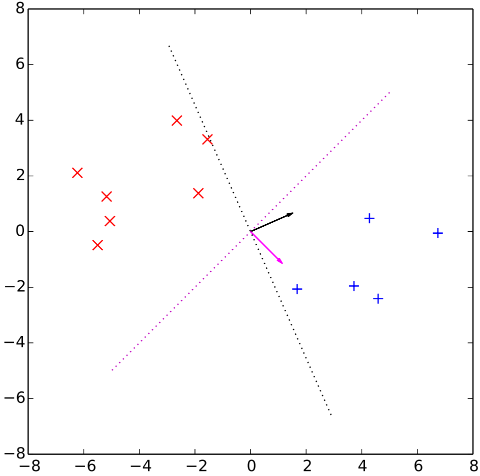
Perceptron Training Algorithm in Action



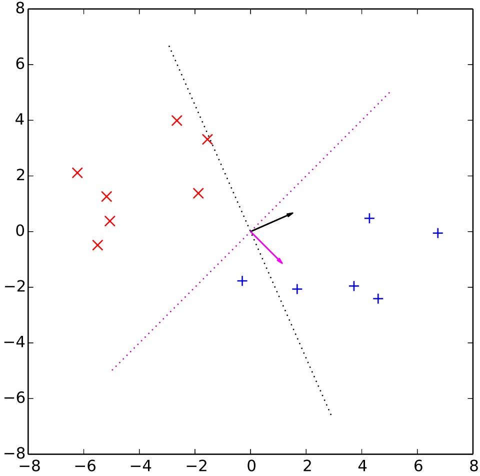
Perceptron Training Algorithm in Action



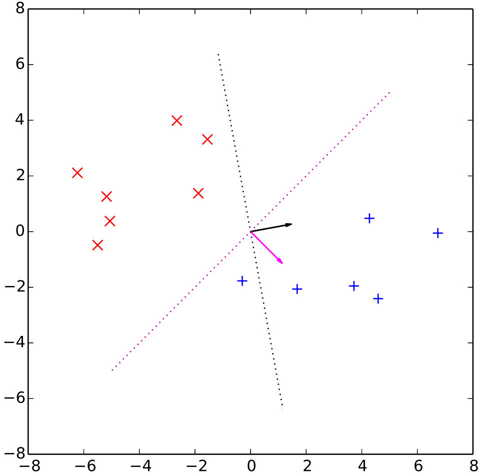
Perceptron Training Algorithm in Action



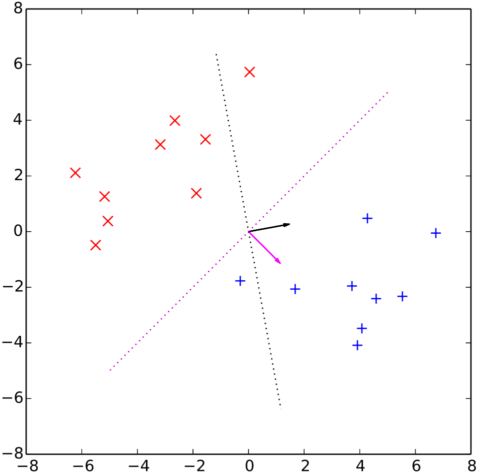
Perceptron Training Algorithm in Action



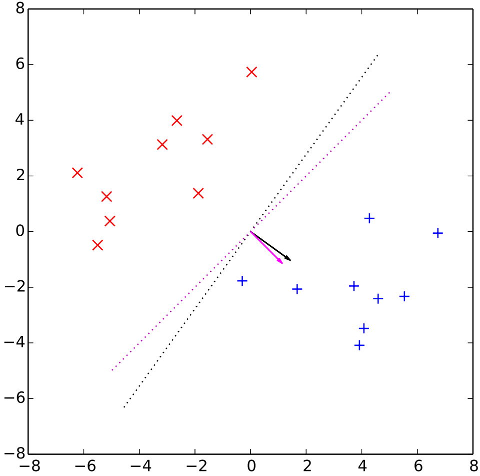
Perceptron Training Algorithm in Action



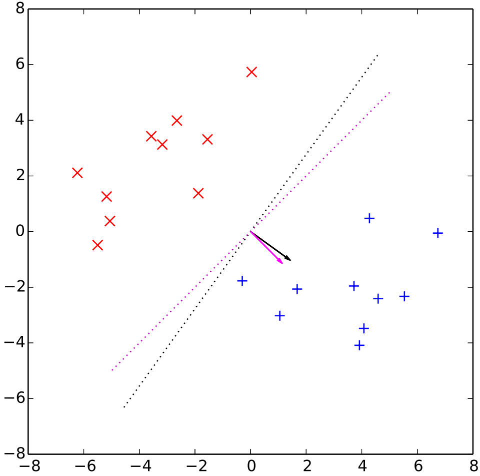
Perceptron Training Algorithm in Action



Perceptron Training Algorithm in Action



Perceptron Training Algorithm in Action



What is machine learning?

Some Definitions

- Kevin Murphy: *". . . , we define machine learning as a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty .."*
- Tom Mitchell: *"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."*
- Same (or similar) programs work across a range of learning tasks. (though not universally)

Machine Learning

- Intersection of computer science, statistics, neuroscience/biology, engineering, optimization *etc.*
- Statistics
  - How much data is needed?
  - When can we be confident in our predictions?
- Computer Science
  - Design algorithms for automated pattern discovery. How fast do these run?
  - How much computational power is needed?

Outline

History of Machine Learning

This Class

Some Machine Learning Applications

Some Practical Concerns

About this course

- Pre-requisites: Basic linear algebra, calculus, probability, algorithms, programming.
- Mathematical foundations of ML; not computational or statistical learning theory
- Regression, support vector machines, neural networks, deep learning, clustering [video]
- Conceptual programming assignments; not scaling to real-world systems

About this course

- Discussion forum on Piazza (link on webpage)
- Classes in Weeks 3-7 (Mon, Wed, Fri - 6 groups)
- Practicals in Weeks 2-8 (Tue, Thu - 2 groups)
- Final examination over easter break
- Office Hours: Tue 15:30-16:30 (449 Wolfson Building)

Outline

History of Machine Learning

This Class

Some Machine Learning Applications

Some Practical Concerns

Application: Boston Housing Dataset

Real attributes

- Crime rate per capita
- Non-retail business fraction
- Nitric Oxide concentration
- Age of house
- Floor area
- Distance to city centre

Integer attributes

- Number of rooms

Categorical attributes

- On the Charles river?
- Index of highway access (1-5)

Source: UCI repository

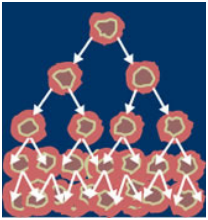
Predict house cost

Application: Breast Cancer

Integer attributes

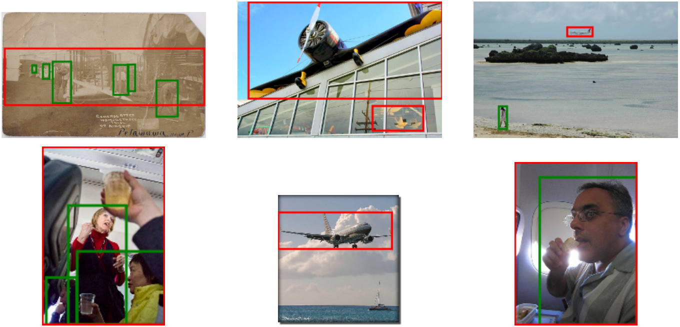
- Clump thickness
- Uniformity of cell size
- Uniformity of cell shape
- Marginal adhesion
- Single epithelial cell size
- Bare nuclei
- Bland Chromatin
- Normal nucleoli
- Mitoses

Predict: Benign vs Malignant



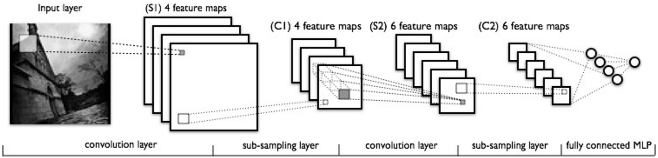
Source: UCI repository

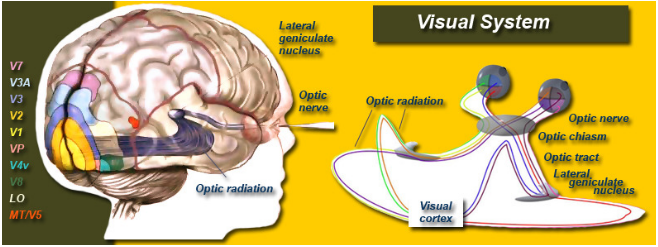
Application: Object Detection and Localization



- 200-basic level categories
- Dataset contains over 400,000 images
- Imagenet competition (2010--15)

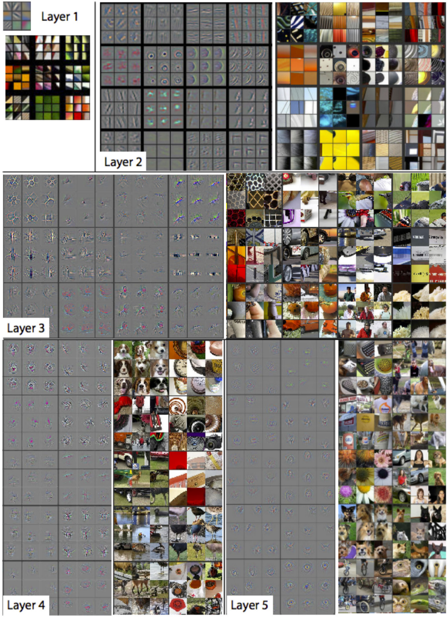
Application: Object Detection and Localization





Source: DeepLearning.net (top); Brain-Maps.com (bottom)

Application: Object Detection and Localization



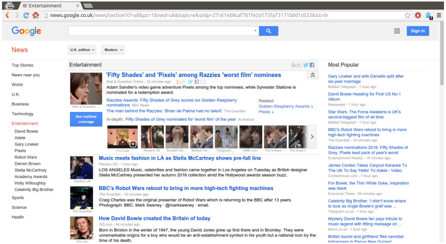
Source: Zeiler and Fergus (2013)

Supervised Learning

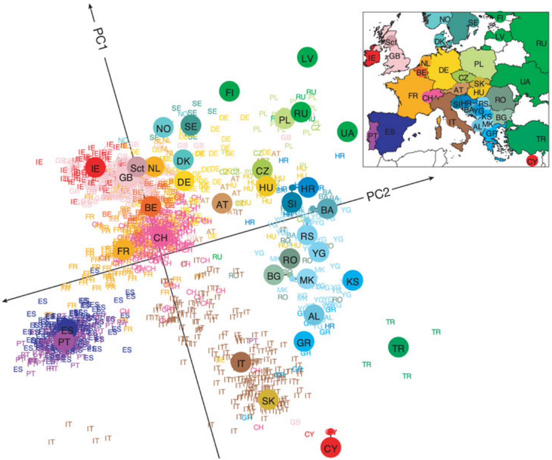
- Training data has inputs ($\mathbf{x}$) as well as outputs (y)
- Regression: When the output is real-valued, *e.g.,*Housing data
- Classification: Output is a category
  - Binary classification -- only two classes *e.g.,*Cancer, spam
  - Multi-class classification -- several classes *e.g.,*Object detection

Unsupervised Learning : Grouping News Articles

- Group items into categories: sports, music, business, *etc.*
- Labels are not known
- Algorithm cannot know "label names"



Unsupervised Learning : Genetic Data of European Populations



Source: Novembre *et al.*, Nature (2008)

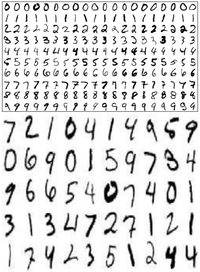
Active and Semi-Supervised Learning

Active Learning

- Data is unlabelled
- Learning algorithm can ask for a label (from a human)

Semi-supervised Learning

- Some data is labelled, a lot more unlabelled
- Can using the two together help?



Anomaly Detection or One-class Classification

Examples

- Detect possible malfunction at nuclear reactors
- Detect fraudulent transactions for credit cards

- Supervised learning vs anomaly detection
- Anomalous events much rarer, possibly not related to each other

Recommendation Systems

Movie / User	Alice	Bob	Charlie	Dean	Eve
The Shawshank Redemption	7	9	9	5	2
The Godfather	3	?	10	4	3
The Dark Knight	5	9	?	6	?
Pulp Fiction	?	5	?	?	10
Schindler's List	?	6	?	9	?

- Netflix competition to predict user-ratings (2008-09)
- Applications to all kinds of product recommendations
- No user will have used several products; take advantage of large number of users

Reinforcement Learning

- Automatic flying helicopter; self-driving cars
- Cannot program by hand
- Stochastic environment (hard to define precisely)
- Must take sequential decisions
- Can define reward functions
- Fun: Playing Atari breakout! [video]

Outline

History of Machine Learning

This Class

Some Machine Learning Applications

Some Practical Concerns

Cleaning up data

Spam Classification

- Look for words such as Nigeria, millions, Viagra, *etc.*
- Features such as the IP, other metadata
- If email addressed by name

Getting Features

- Often hand-crafted features by domain experts
- This class mainly assumes we already have features
- Feature learning using deep networks

Some pitfalls

Sample Email

''To build a spam classifier, we look for words such as Nigeria, millions, *etc.*''

Some pitfalls

Sample Email

"To build a spam classifier, we look for words such as Nigeria, millions, *etc.*"

Training vs Test Data

- Future data should look like past data
- Not true for spam classification

Cats vs Dogs

Next Class

Linear Regression

- Brush up your linear algebra and calculus!