Machine Learning - MT 2017
1. Introduction

Christoph Haase & Varun Kanade

University of Oxford
October 9, 2017

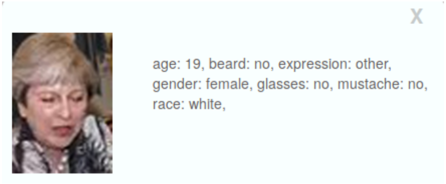
Machine Learning in Action

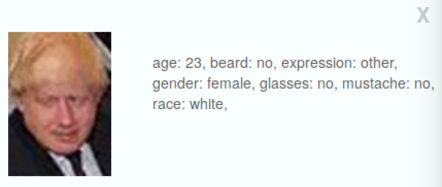
(Using https://www.betafaceapi.com/demo.html)

Machine Learning in Action

(Using https://www.betafaceapi.com/demo.html)

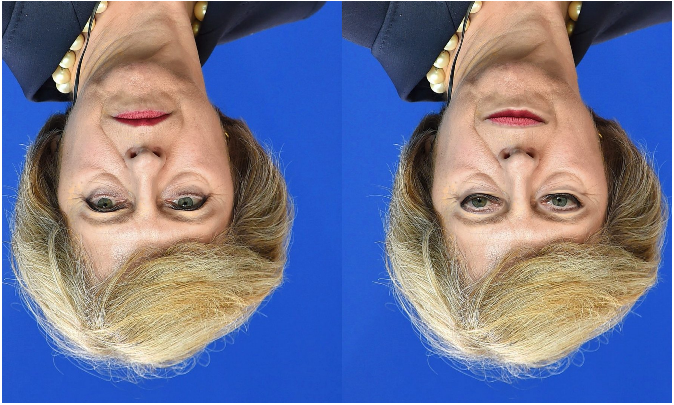
Machine Learning in Action





(Using https://www.betafaceapi.com/demo.html)

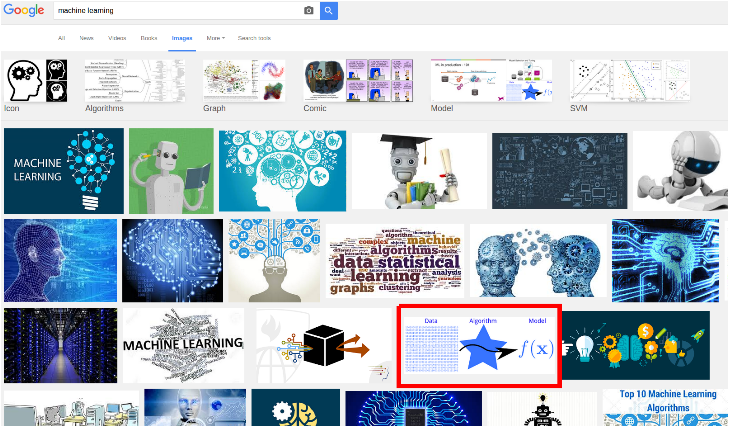
Is anything wrong?



Is anything wrong?

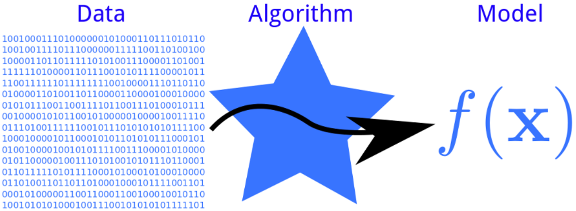
(See Guardian article)

What is machine learning?



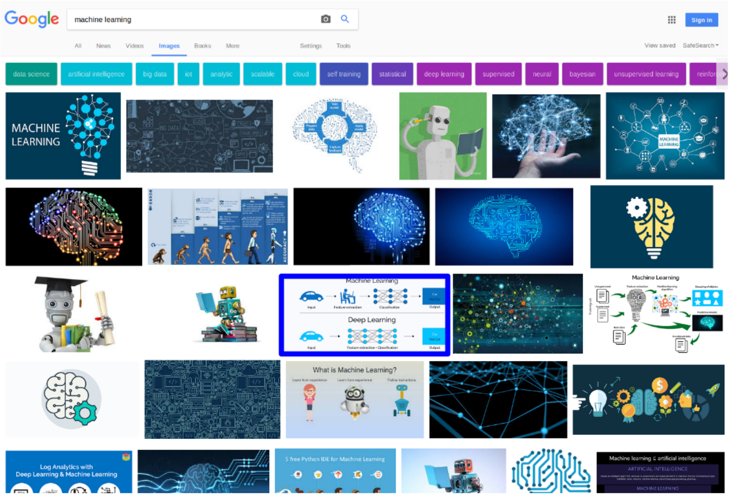
circa October 2016

What is machine learning?



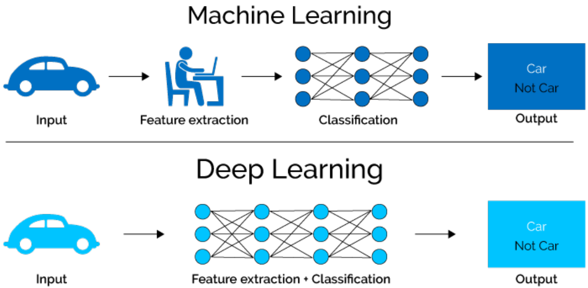
circa October 2016

What is machine learning?



circa October 2017

What is machine learning?

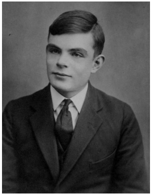


circa October 2017

What is machine learning?

What is artificial intelligence?

''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.''



Turing, A.M. (1950). Computing machinery and intelligence. Mind, 59, 433-460.

What is machine learning?

Definition by 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.

Face Detection

- E : images (with bounding boxes) around faces
- T : given an image without boxes, put boxes around faces
- P : number of faces correctly identified

An early (first?) example of automatic classification

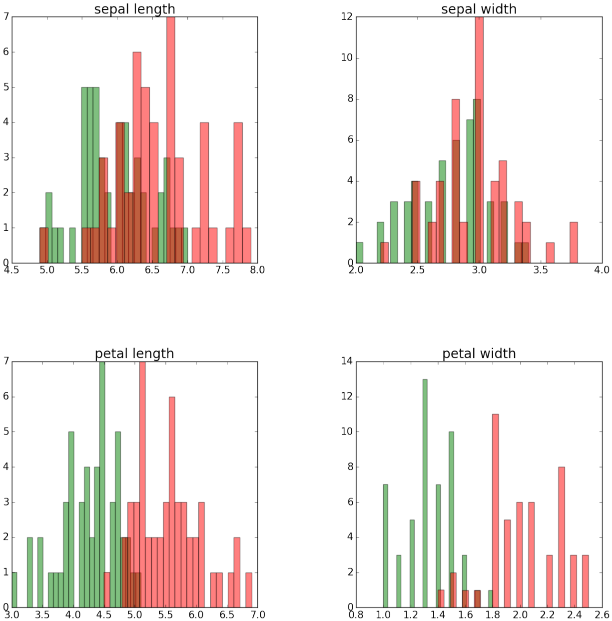
Ronald Fisher: Iris Flowers (1936)

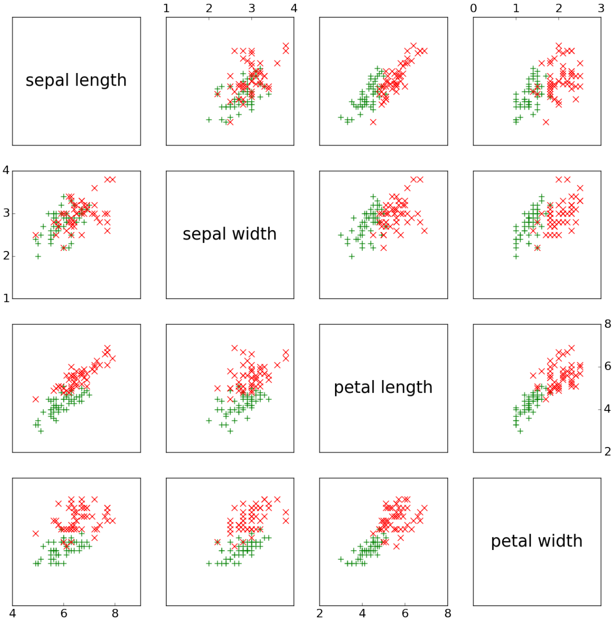
- Three types: setosa, versicolour, virginica
- Data: sepal width, sepal length, petal width, petal length

setosa

versicolour

virginica


sepal length
0
1
2
3
4
5
6
7
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
sepal width
0
2
4
6
8
10
12
2.0
2.5
3.0
3.5
4.0
petal length
0
1
2
3
4
5
6
7
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
petal width
0
2
4
6
8
10
12
14
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6


sepal length
sepal width
petal length
petal width
1 2 3 4
0 1 2 3
4 3 2 1
8 6 4 2
4 6 8
2 4 6 8

An early (first?) example of automatic classification

Ronald Fisher: Iris Flowers (1936)

- Three types: setosa, versicolour, virginica
- Data: sepal width, sepal length, petal width, petal length
- Method: Find linear combinations of features that maximally differentiates the classes

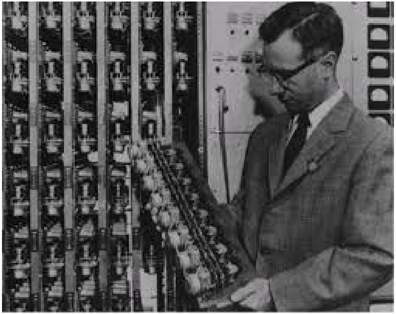
setos

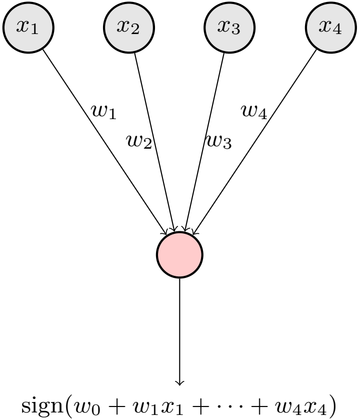
versicolour

virginica

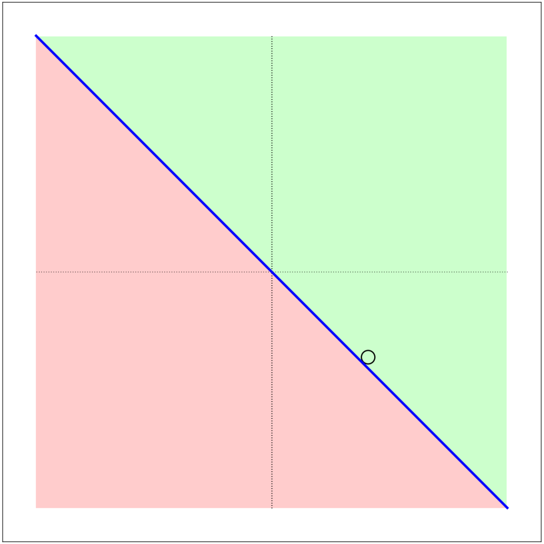
Frank Rosenblatt and the Perceptron

- Perceptron - inspired by neurons
- Simple learning algorithm
- Built using specialised hardware

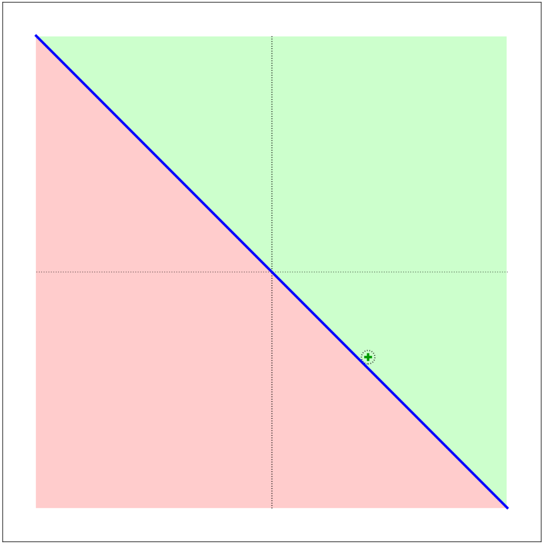




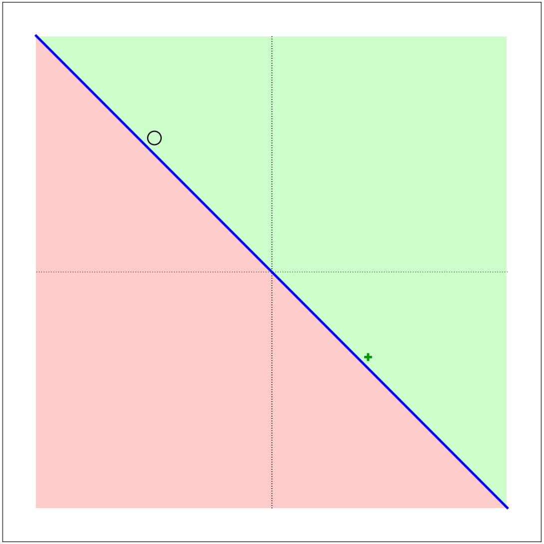
Perceptron Training Algorithm



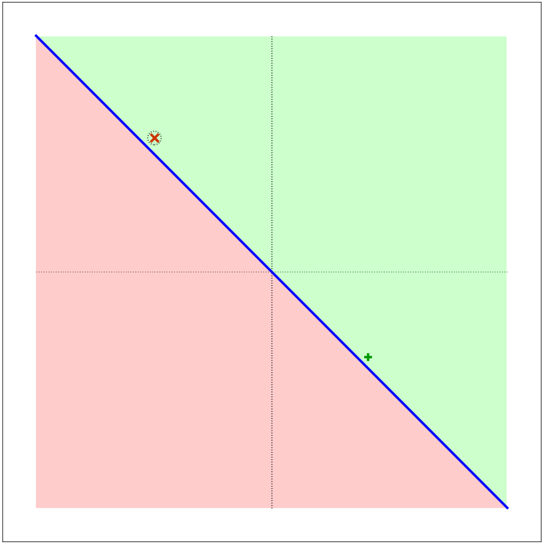
Perceptron Training Algorithm



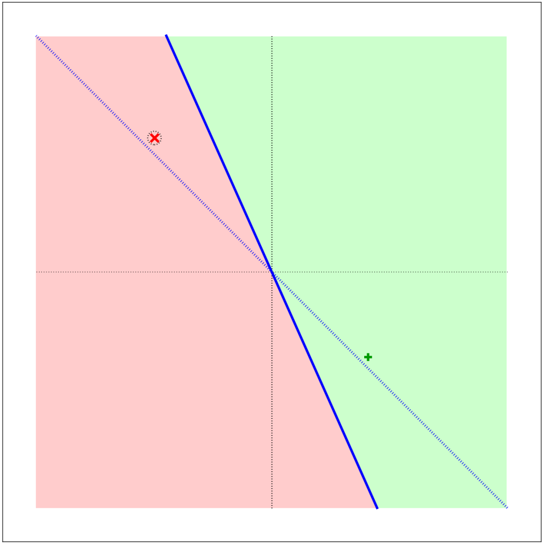
Perceptron Training Algorithm



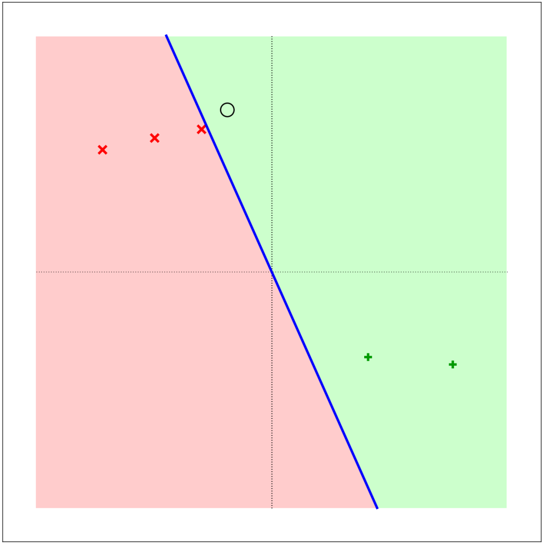
Perceptron Training Algorithm



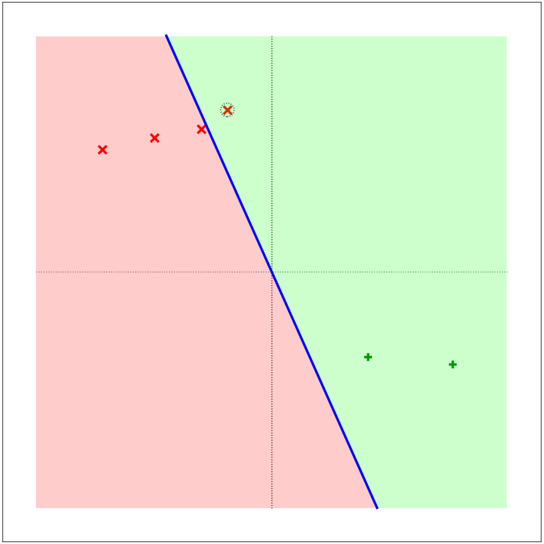
Perceptron Training Algorithm



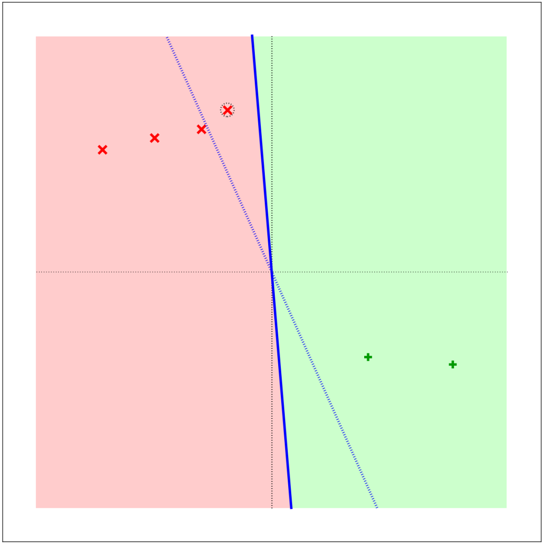
Perceptron Training Algorithm



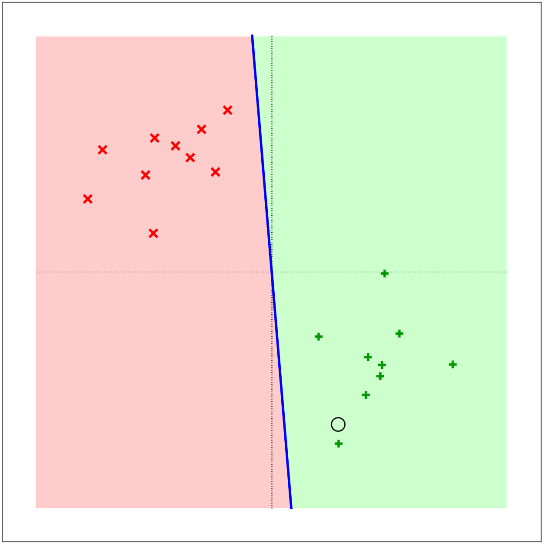
Perceptron Training Algorithm



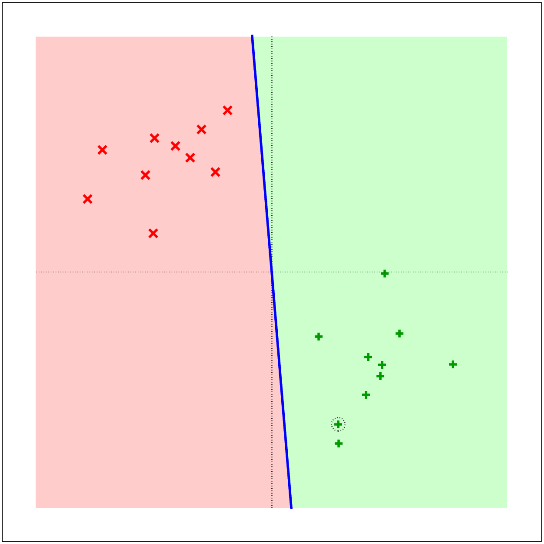
Perceptron Training Algorithm



Perceptron Training Algorithm



Perceptron Training Algorithm



Course Information

Website

`www.cs.ox.ac.uk/teaching/materials17-18/ml/`

Lectures

Mon, Wed (16h-17h), Fri (10h-11h only Weeks 1-4) in Lecture Theatre A

Classes

Weeks 2^*, 3, 5, 6, 8.

Instructors: Yarin Gal, Christoph Haase, Peter Minary

Practicals

Weeks 4, 6, 7, 8.

Demonstrators: Francisco Marmolejo, Javier Morales, Wenjie Ruan, Bo Yang

Office Hours

Wed 11h-12h in #417 (CH)

Mon 17h-18h in #449 (VK)

Course Information

Textbooks

Kevin Murphy - Machine Learning: A Probabilistic Perspective

- Online access through Bodleian library

Ian Goodfellow, Yoshua Bengio, Aaron Courville - Deep Learning

- Almost finished draft at www.deeplearningbook.org

Assessment

Sit-down exams. Different times for M.Sc. and UG

Piazza

Use for course-related queries

Sign-up at `piazza.com/ox.ac.uk/other/mlmt2017`

Is this course right for you?

Machine learning is mathematically rigorous making use of probability, linear algebra, multivariate calculus, optimisation *etc.*

Lots of equations, derivations, not "proofs"

Try Sheet 0 (optional class in Week 2)

For M.Sc. students:

- Computational Learning Theory

Practicals

You will have to be an efficient programmer

Implement learning algorithms discussed in the lectures

We will use `python v3.6` (`anaconda`, `tensorflow`)

Familiarise yourself with `python` and `numpy` by Week 4

A few last remarks about this course

As ML developed through various disciplines - CS, Stats, Neuroscience, Engineering, *etc.*, there is no consistent usage of notation or even names among the textbooks. At times you may find inconsistencies even within a single textbook.

You will be required to read, both before and after the lectures. We will post suggested reading on the website.

Resources:

- Wikipedia has many great articles about ML and background material
- Online videos: Andrew Ng on coursera, Nando de Freitas on youtube, *etc.*
- Many interesting blogs, podcasts, *etc.*

Learning Outcomes

On completion of the course students should be able to

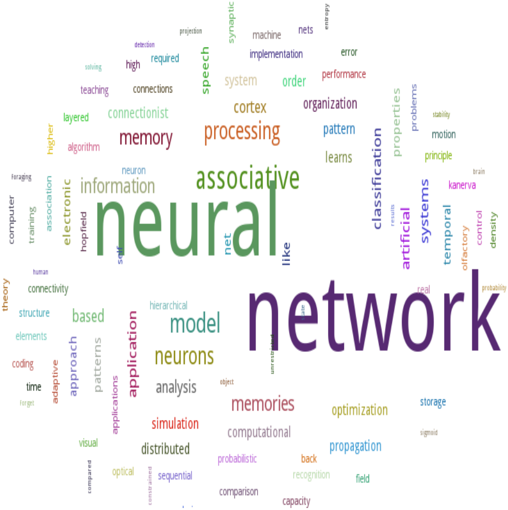
- Describe and distinguish between various different paradigms of machine learning, particularly supervised and unsupervised learning
- Distinguish between task, model and algorithm and explain advantages and shortcomings of machine learning approaches
- Explain the underlying mathematical principles behind machine learning algorithms and paradigms
- Design and implement machine learning algorithms in a wide range of real-world applications (not to scale)

Machine Learning Models and Methods

k-Nearest Neighbours
Linear Regression
Logistic Regression
Ridge Regression
Hidden Markov Models
Mixtures of Gaussian
Principle Component Analysis
Independent Component Analysis
Kernel Methods
Decision Trees
Boosting and Bagging
Belief Propagation
Variational Inference
EM Algorithm
Monte Carlo Methods
Spectral Clustering
Hierarchical Clustering
Recurrent Neural Networks

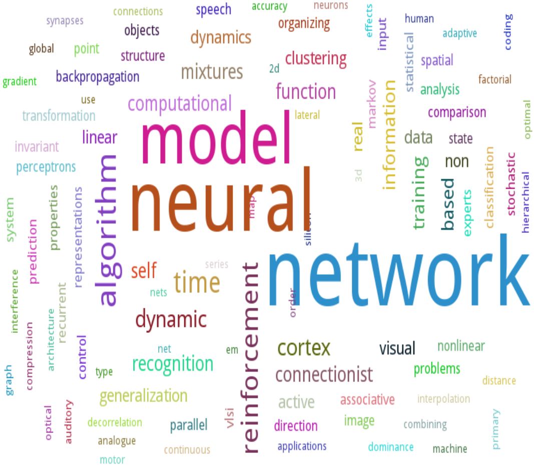
Linear Discriminant Analysis
Quadratic Discriminant Analysis
Perceptron Algorithm
Naïve Bayes Classifier
Hierarchical Bayes
k-means Clustering
Support Vector Machines
Gaussian Processes
Deep Neural Networks
Convolutional Neural Networks
Markov Random Fields
Structural SVMs
Conditional Random Fields
Structure Learning
Restricted Boltzmann Machines
Multi-dimensional Scaling
Reinforcement Learning
$\cdots$

NIPS Papers!



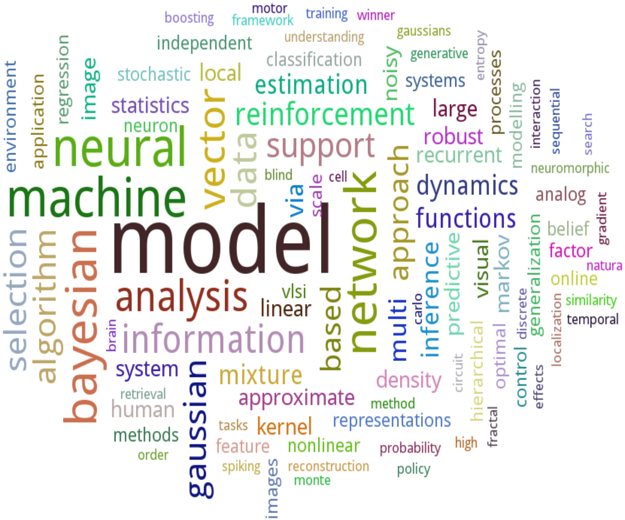
Advances in Neural Information Processing Systems 1988

NIPS Papers!



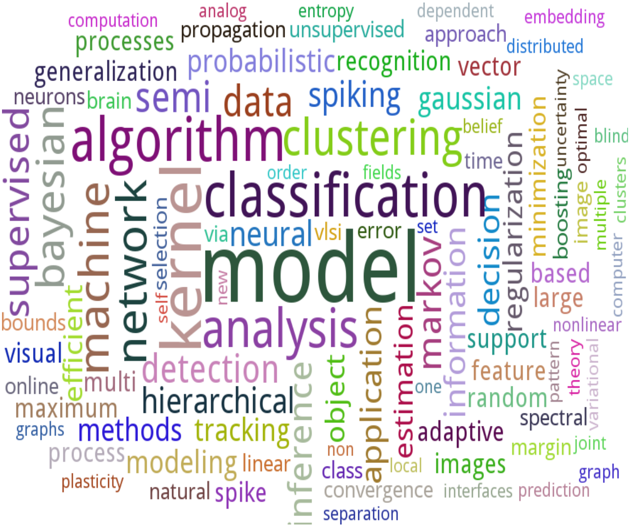
Advances in Neural Information Processing Systems 1995

NIPS Papers!



Advances in Neural Information Processing Systems 2000

NIPS Papers!

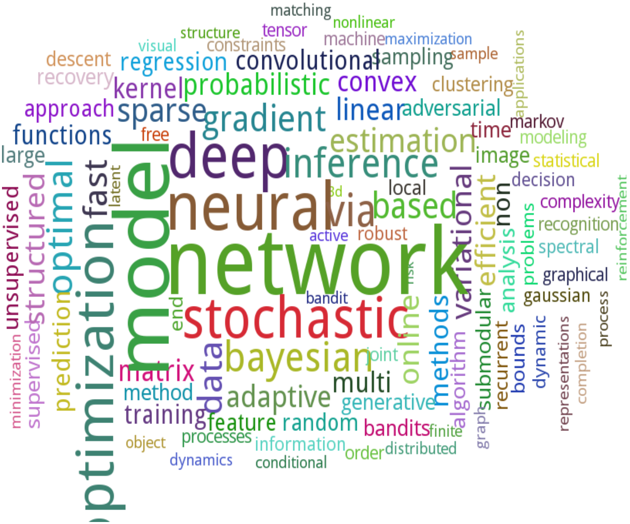


Advances in Neural Information Processing Systems 2005

NIPS Papers!

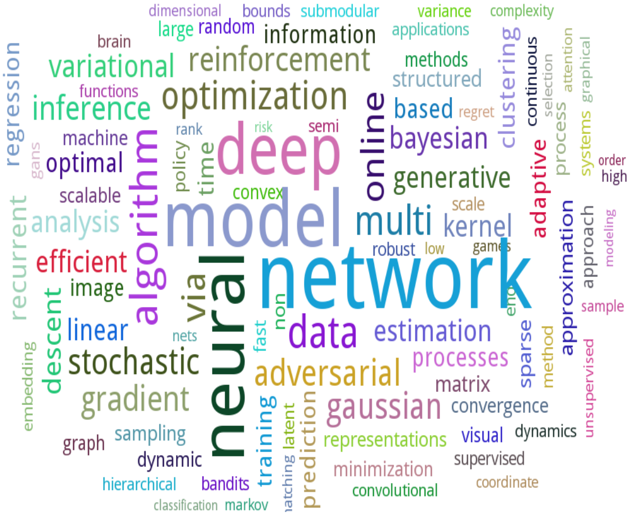
Advances in Neural Information Processing Systems 2009

NIPS Papers!



Advances in Neural Information Processing Systems 2016

NIPS Papers!



Advances in Neural Information Processing Systems 2017 [video]

Application: Boston Housing Dataset

Numerical attributes

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

Categorical attributes

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

Source: UCI repository

Predict house cost

Application: Object Detection and Localisation

- 200-basic level categories
- Here: Six pictures containing airplanes and people
- Dataset contains over 400,000 images
- Imagenet competition (2010-)
- All recent successes through very deep neural networks!

Supervised Learning

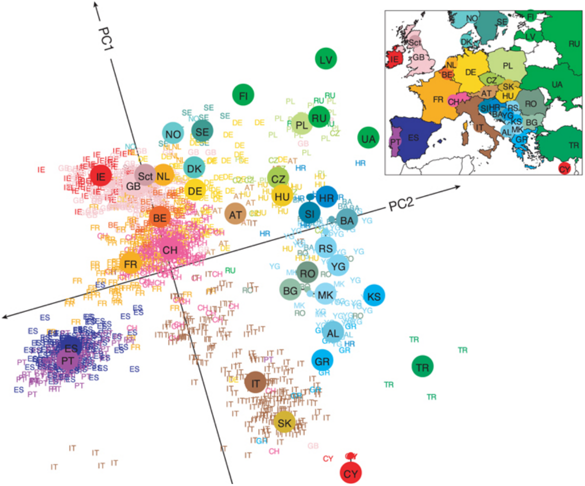
Training data has inputs $\mathbf{x}$ (numerical, categorical) as well as outputs y (target)

Regression: When the output is real-valued, *e.g.,* housing price

Classification: Output is a category

- Binary classification: only two classes *e.g.,* spam
- Multi-class classification: several classes *e.g.,* object detection

Unsupervised Learning : Genetic Data of European Populations



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

Experience (E)

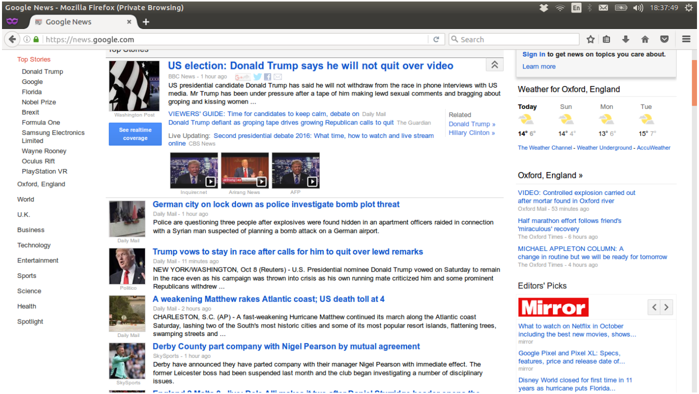
Task (T)

Performance (P)

Dimensionality reduction - Map high-dimensional data to low dimensions

Clustering - group together individuals with similar genomes

Unsupervised Learning : Group Similar News Articles



Group similar articles into categories such as politics, music, sport, *etc.*

In the dataset, there are no labels for the articles

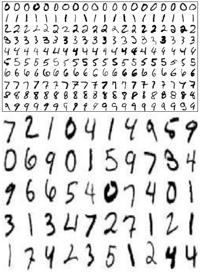
Active and Semi-Supervised Learning

Active Learning

- Initially all data is unlabelled
- Learning algorithm can ask a human to label some data

Semi-supervised Learning

- Limited labelled data, lots of unlabelled data
- How to use the two together to improve learning?



Collaborative Filtering : Recommender 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)

Any individual user will not have used most products

Most products will have been use by some individual

NETFLIX

Reinforcement Learning

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

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 to user personally

Getting Features

- Often hand-crafted features by domain experts
- In this course, we mainly assume that we already have features
- Feature learning using deep networks

Some pitfalls

Sample Email

"To build a spam classifier, we check if at least two words such as Nigeria, millions, *etc.* appear in the message. If that is the case, we mark the email as spam."

Training vs Test Data

- Future data should look like past data
- Not true for spam classification. Spammers will try adversarially to break the learning algorithm.

Cats vs Dogs

Next Time

Review of Maths for ML

- Brush up your linear algebra and calculus!