

# CSCU9YE - Artificial Intelligence

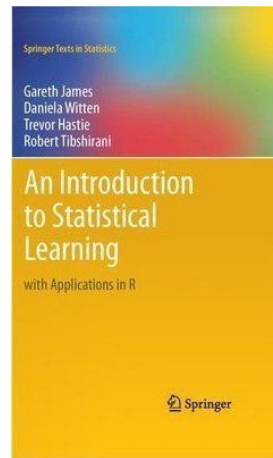


## Lecture 7: Introduction to Machine Learning

Prof. Gabriela Ochoa, University of Stirling

# Main sources

- Chapter 18 Learning from Examples from **Artificial Intelligence: A Modern Approach**,
- **A Course in Machine Learning** by Hal Daumé III (<http://ciml.info/>)
- **The Hundred-Page Machine Learning** Book by Andriy Burkov (<http://themlbook.com/wiki/doku.php>)
- **An Introduction to Statistical Learning** by Gareth James, Daniela Witten, Trevor Hastie and Rob Tibshirani (<http://faculty.marshall.usc.edu/gareth-james/ISL/>)
- Online courses: Udacity (<https://www.udacity.com/>)



# What is machine learning?

**Machine learning** is the science of getting computers to act without being explicitly programmed.

A system is **learning** if it improves its performance on future tasks after making observations about the world

**Machine learning** is the process of solving a practical problem by:

1. Gathering a dataset
2. Algorithmically building a statistical model based on that dataset

# Why do we need learning algorithms?

## 3 main reasons:

1. The designers cannot anticipate all possible situations that software will encounter (e.g. different texts for translation, different faces)
2. The designers cannot anticipate all changes over time. Conditions may change. (e.g. a program for predicting stock market prices)
3. Sometimes, human programmers have no idea how to program a solution themselves (e.g. language translations, face recognition)

# WHAT IS MACHINE LEARNING

THE  
ROYAL  
SOCIETY



What is Machine  
Learning? The  
Royal Society

# Types of machine learning algorithms

According to the *feedback* available to learn from:

## Supervised

Given example input-output pairs, learn a function that maps from input to outputs

**Tasks:** classification, regression

## Unsupervised

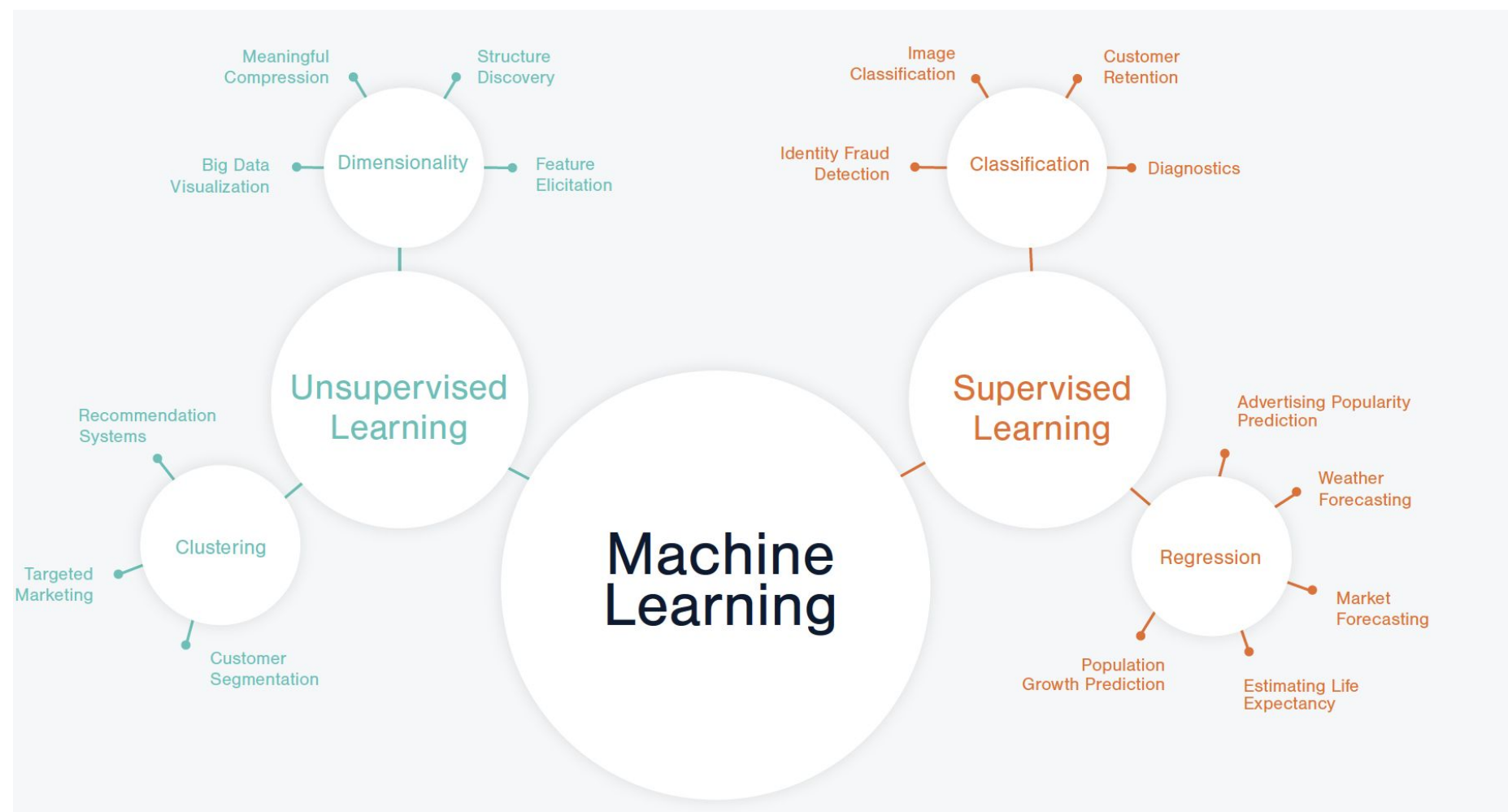
Given inputs, learn patterns even though no explicit feedback is supplied.

**Tasks:** clustering, dimensionality reduction

## Reinforcement

Given feedback from the environment (rewards, punishments), determine the ideal behaviour

**Tasks:** game playing, online control



# Typical machine learning tasks (supervised)

## Binary Classification

Predict a simple yes/no response, only two classes.

**Applications:** diagnostics, spam detection,

## Multiclass classification

Identify an object into one of a number of classes.

**Applications:** a story is about politics, science, entertainment, etc

## Regression

Predicting a continuous-valued attribute associated with an object

**Applications:** drug response, stock prices, etc.

## Ranking

Put a set of objects in order of relevance

**Applications:** ranking preferences of user for movies or books



# Typical machine learning tasks (unsupervised)

## Clustering

- Organise similar items into groups.
- **Applications:** customer segmentation, grouping experiment outcomes

## Dimensionality reduction

- Taking high dimensional data and embedding it in a lower dimension space
- Low-dimensional data representations that remove noise but retain the signal of interest can be instrumental in understanding hidden structures and patterns
- **Applications:** feature extraction, data compression, exploration, and visualisation.

# Terminology

Input variables				Target variable
Date	Temperature	Humidity	Wind Speed	Weather
01/10/17	22	48	2.7	Sunny
02/10/17	15	80	3.8	Rainy
03/10/17	12	45	17.9	Windy
04/10/17	14	77	4.2	Cloudy

## Other names for variable

- Column
- Feature
- Attribute
- Field
- Dimension

## Other names for sample

- Row
- Record
- Observation
- Example
- Instance

# Variable types

## Most common

- **Numeric**: discrete, continuous
- **Categorical**: binary, nominal, ordinal

## Others

- String
- Date ...

## Numeric Variables

Values are numbers

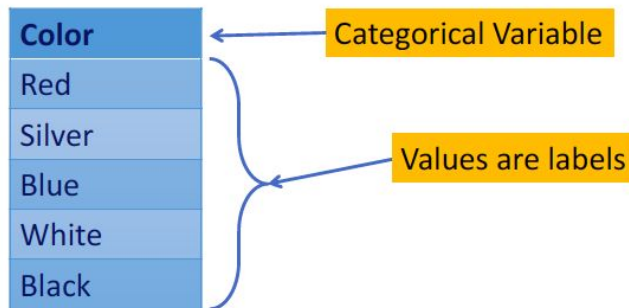
Also called 'quantitative'

1, 4.450, -0.8767,  $7 \times 10^5$

## Categorical Variables

Values are labels, names, or categories

Also called 'qualitative' or 'nominal'



# Classification: weather forecast

Given input variables, predict a a category or class

Data:

Input variables				Target variable
Date	Temperature	Humidity	Wind Speed	Weather
01/10/17	22	48	2.7	Sunny
02/10/17	15	80	3.8	Rainy
03/10/17	12	45	17.9	Windy
04/10/17	14	77	4.2	Cloudy



What would be the wheatear in the future?

Weather in 05/10/17?

# Binary classification: spam filter

Input: an email

Output: spam/no spam

Setup:

- Get a large collection of example emails, each labeled “spam” or “not spam”
- **Note:** someone has to hand label all this data!
- Want to learn to predict labels of new, future emails

Features: Attributes used to decision

- Words: FREE!
- Text Patterns: \$dd, CAPS
- Non-text: SenderInContacts

...



Dear Sir.

First, I must solicit your confidence in this transaction, this is by virtue of its nature as being utterly confidential and top secret. ...



TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.

99 MILLION EMAIL ADDRESSES FOR ONLY \$99



Ok I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I ....

# Classification: digit recognition

- **Input:** images / pixel grids
- **Output:** a digit 0-9
- **Setup**
  - Get a large collection of example images, each labeled with a digit
  - **Note:** someone has to hand label all this data!
  - Want to learn to predict labels of new, future digit images
- **Features:** The attributes used to make the digit decision
  - Pixels: (6,8)=ON
  - Shape Patterns: NumComponents, AspectRatio, NumLoops
  - ...



0



1



2



1



??

# Demo

Microsoft Research ML

- Facial detection and recognition
- Gender prediction
- Age prediction

<https://how-old.net/>

# Regression: stock price forecast

Predicting the value of a target continuous numerical value from input variables



Microsoft's stock price (blue line) and forecast (purple line) made in 22-Oct-2001.

The yellow points show stock prices since the forecast was made.



# Clustering: grouping google news

Google News Search and browse 25,000 news sources updated continuously.

World » U.S. »

**Heavy Fighting Continues As Pakistan Army Battles Taliban**  
Voice of America - 10 hours ago  
By Barry Newhouse Pakistan's military said its forces have killed 55 to 60 Taliban militants in the last 24 hours in heavy fighting in Taliban-held areas of the northwest. [Pakistani troops battle Taliban militants for fourth day](#) guardian.co.uk  
[Army: 55 militants killed in Pakistan fighting](#) The Associated Press  
[Christian Science Monitor](#) - [CNN International](#) - [Bloomberg](#) - [New York Times](#)  
[all 3,824 news articles »](#)

**Sri Lanka admits bombing safe haven**  
guardian.co.uk - 3 hours ago  
Sri Lanka has admitted bombing a "safe haven" created for up to 150,000 civilians fleeing fighting between Tamil Tiger fighters and the army.  
[Chinese billions in Sri Lanka fund battle against Tamil Tigers](#) Times Online  
[Huge Humanitarian Operation Under Way in Sri Lanka](#) Voice of America  
[BBC News](#) - [Reuters](#) - [AFP](#) - [Xinhua](#)  
[all 2,492 news articles »](#)

**Weekend Opinionator: Souter, Specter and the Future of the GOP**  
New York Times - 48 minutes ago  
By Tobin Harshaw An odd week. While Barack Obama celebrated his 100th day in office, the headlines were pretty much dominated by the opposition party, albeit not in the way many Republicans would have liked.  
[US Supreme Court Vacancy An Early Test For Sen Specter](#) Wall Street Journal  
[Letters: Arlen Specter, Notre Dame, Chrysler](#) Houston Chronicle  
[The Associated Press](#) - [Kansas City Star](#) - [Philadelphia Inquirer](#) - [Bangor Daily News](#)  
[all 401 news articles »](#)

**Joe Biden, the Flu and You**  
New York Times - 48 minutes ago  
By GAIL COLLINS The swine flu scare has made it clear why Barack Obama picked Joe Biden for vice president. David Brooks and Gail Collins talk between columns.  
[After his flu warning, Biden takes the train home](#) The Associated Press  
[Biden to visit Balkan states in mid-May](#) Washington Post  
[AFP](#) - [Christian Science Monitor](#) - [Bizjournals.com](#) - [Voice of America](#)  
[all 1,506 news articles »](#)

**Business »**

**Buffett Culls Investment Candidates' 2008 Performance Subpar**  
Bloomberg - 2 hours ago  
By Hugh Son, Erik Holm and Andrew Frye May 2 (Bloomberg) -- Billionaire Warren Buffett said all of the candidates to replace him as chief investment officer of Berkshire Hathaway Inc. failed to beat the 38 percent decline of the Standard & Poor's 500 ...  
[Buffett offers bleak outlook for US newspapers](#) Reuters  
[Buffett: Limit CEO pay through embarrassment](#) MarketWatch  
[CNBC](#) - [The Associated Press](#) - [guardian.co.uk](#)  
[all 1,454 news articles »](#) BRK.A

**Chrysler's Fall May Help Administration Reshape GM**  
New York Times - 5 hours ago  
Auto task force members, from left: Treasury's Ron Bloom and Gene Sperling, Labor's Edward Montgomery, and Steve Rattner. BY DAVID E. SANGER and BILL VLASIC  
WASHINGTON - Fresh from pushing Chrysler into bankruptcy, President Obama and his economic team ...  
[Comment by Gary Chaison](#) Prof. of Industrial Relations, Clark University  
[Bankruptcy reality sets in for Chrysler, workers](#) Detroit Free Press  
[Washington Post](#) - [Bloomberg](#) - [CNNMoney.com](#)  
[all 11,028 news articles »](#) OTC:FIAT - [BIL:FR](#) - [GM](#)

Top-level categories: supervised classification

Story groupings: unsupervised clustering

# Supervised learning

Data a list of observations  $\langle X, y \rangle$

Learn a function from examples

$f$  is the target function - unknown!

An **example** is a pair  $(x, y)$   $y=f(x)$

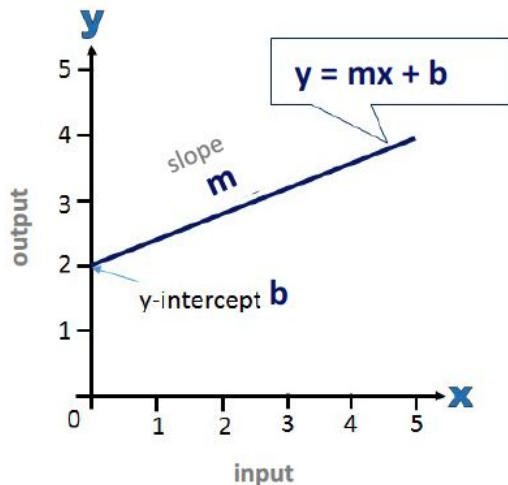
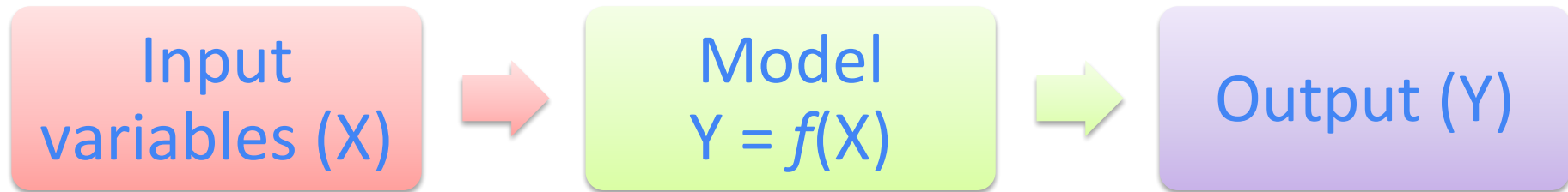
Problem: find a **hypothesis** or **estimate** of  $f$ , let us call it  $h$

such that  $h \approx f$

given a **training set** of examples

X	y
$\langle x_{11}, x_{12}, \dots, x_{1p} \rangle$	$y_1$
$\langle x_{21}, x_{22}, \dots, x_{2p} \rangle$	$y_2$
$\langle x_{31}, x_{32}, \dots, x_{3p} \rangle$	$y_3$
...	...
...	...
...	...
$\langle x_{n1}, x_{n2}, \dots, x_{np} \rangle$	$y_n$

# What is a machine learning model?



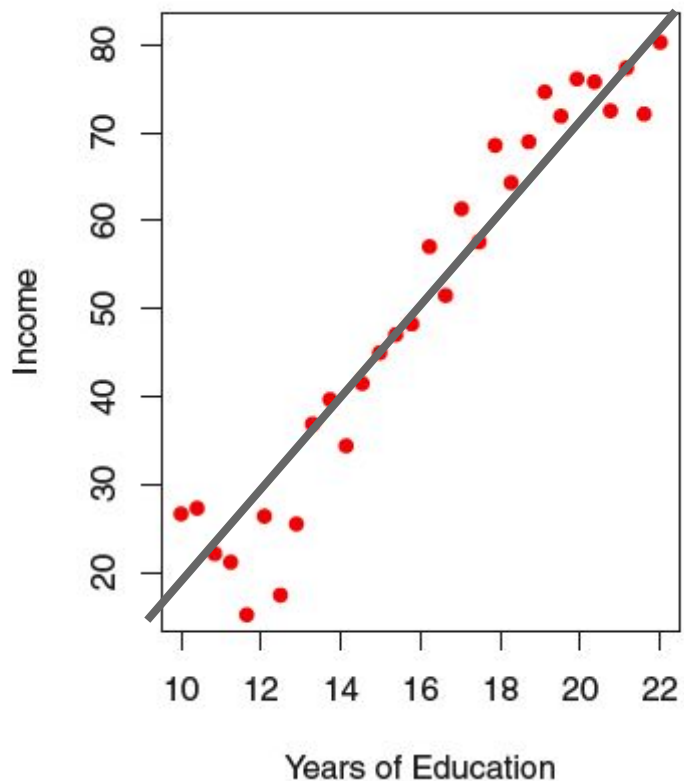
A *mathematical* expression or algorithmic representation that maps input to output

Example of the simplest possible model, a line

$y = mx + b$ ,  $m$  and  $b$  are the parameters

Model parameters are adjusted during model training to change input-output mapping.

# Example linear model income dataset



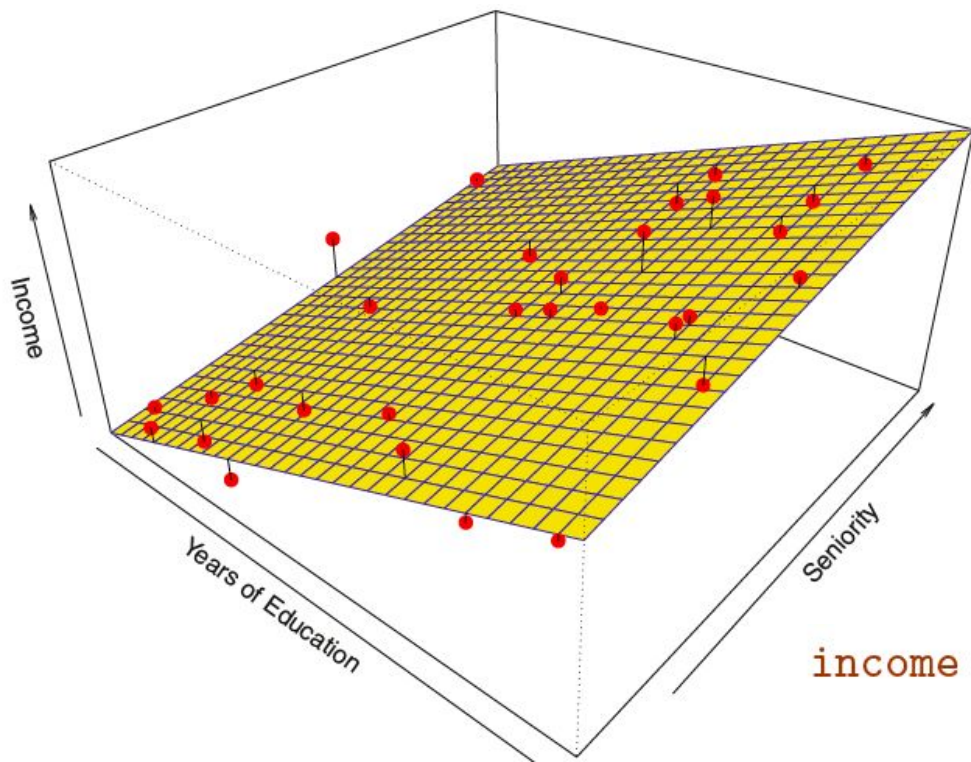
$Y$  = income (in thousands of £)

$X$  = single variable years of education

It seems that we can estimate income with a linear function.

Of course there is always some error in the estimation

# Example linear model income dataset



$Y$  = income (in thousands of £)

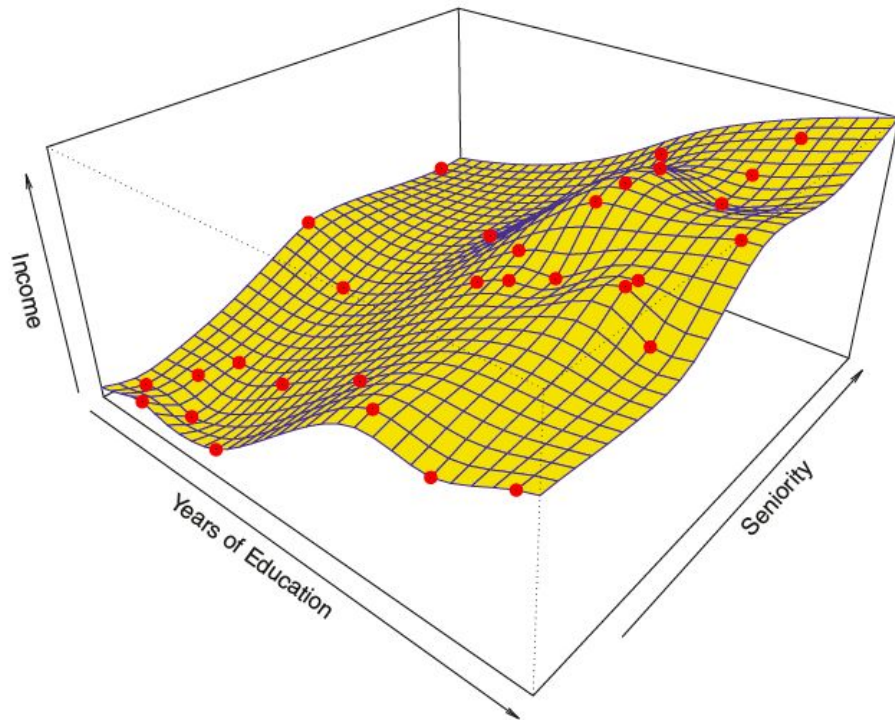
$X$ : 2 variables: Years of Education and Seniority

The observations are shown in red

The yellow plane indicates a **linear model** (parametric model) fitted with a method known as *least squares*.

$$\text{income} \approx \beta_0 + \beta_1 \times \text{education} + \beta_2 \times \text{seniority}.$$

# Example non-linear model income dataset



Non-parametric approach, *thin-plate spline*, is used to estimate  $f$ .

Does not impose any pre-specified model on  $f$ .

Attempts an estimate for  $f$  that is as close as possible to the observed data (smooth function in yellow)

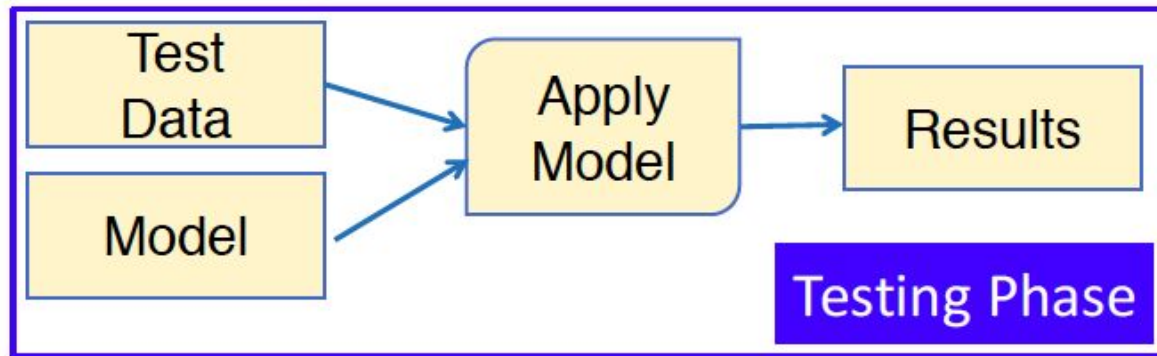
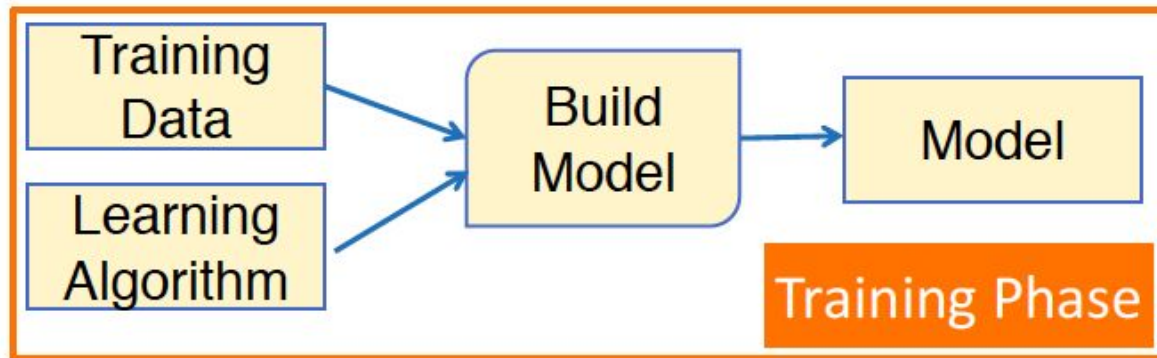
This model fits the data better, but may suffer from *overfitting*.

# What is overfitting?

- Occurs when a statistical model or machine learning algorithm captures the *noise* of the data (as opposed to the *signal*).
- Intuitively occurs when the model or the algorithm fits the data too well.
- Is often a result of an excessively complicated model
- Can be prevented by fitting multiple models and using validation or cross-validation to compare their predictive accuracies on test data.



# Training set and testing set



ML is about learning some properties of a data set and then testing those properties against another data set

ML algorithms are evaluated by splitting a data set into two sets.

1. **Training set:** learning the model;
2. **Testing set:** test the quality (accuracy) of the model.



# Python library: scikit-learn

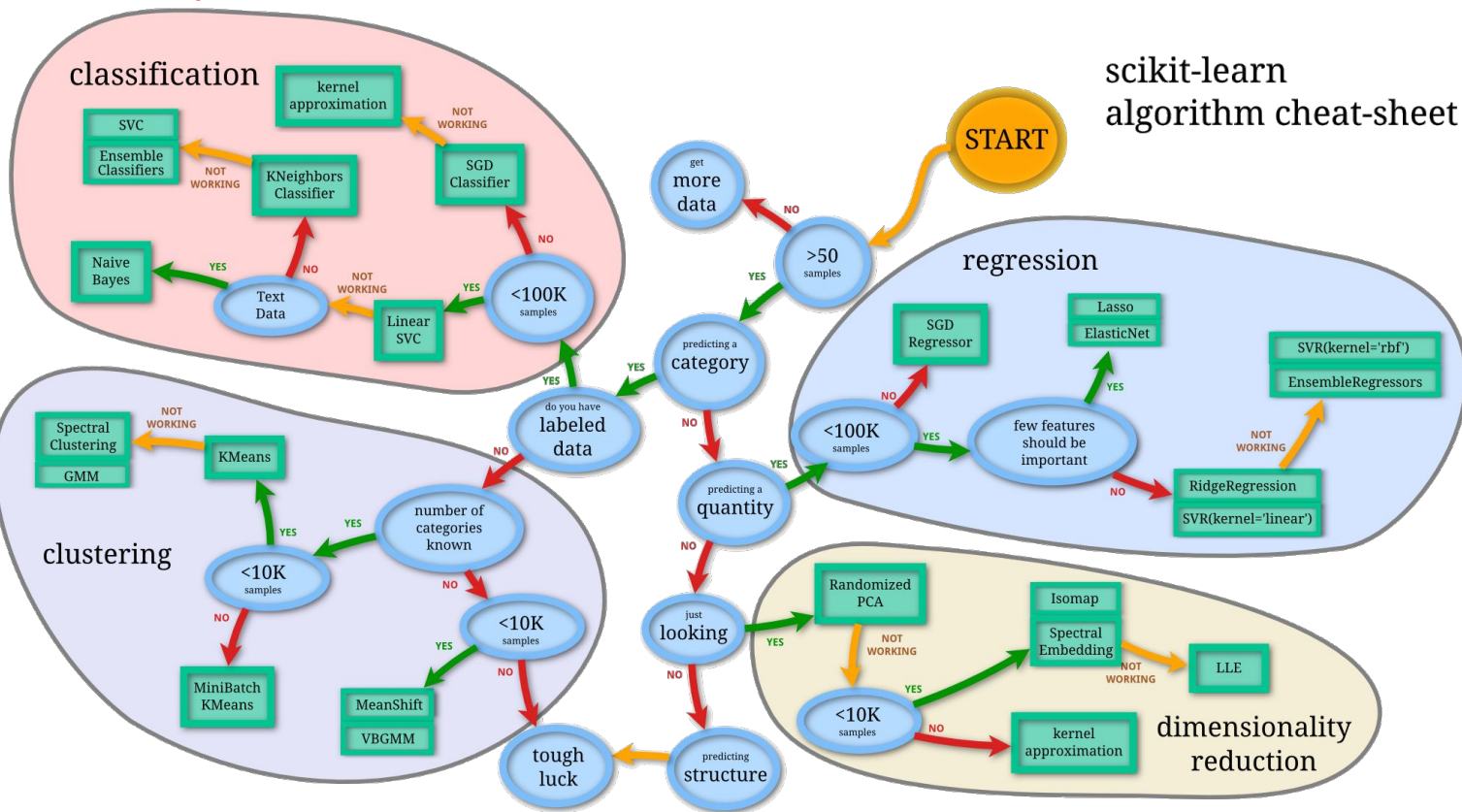


- Open source library for Machine Learning in Python
- Built on top of NumPy, SciPy, Matplotlib
- Improved continuously, active community
- Provides easy to use API for training, and making predictions
- Collection of the best, most popular, algorithms in one place.
- Great documentation!

<http://scikit-learn.org/stable/documentation.html>

# Machine Learning Algorithms: where to start?

map by A. Mueller, scikit-learn



Some Acronyms

SGD: Stochastic Gradient Descent (linear classifiers)

SVC: Support Vector Classifier

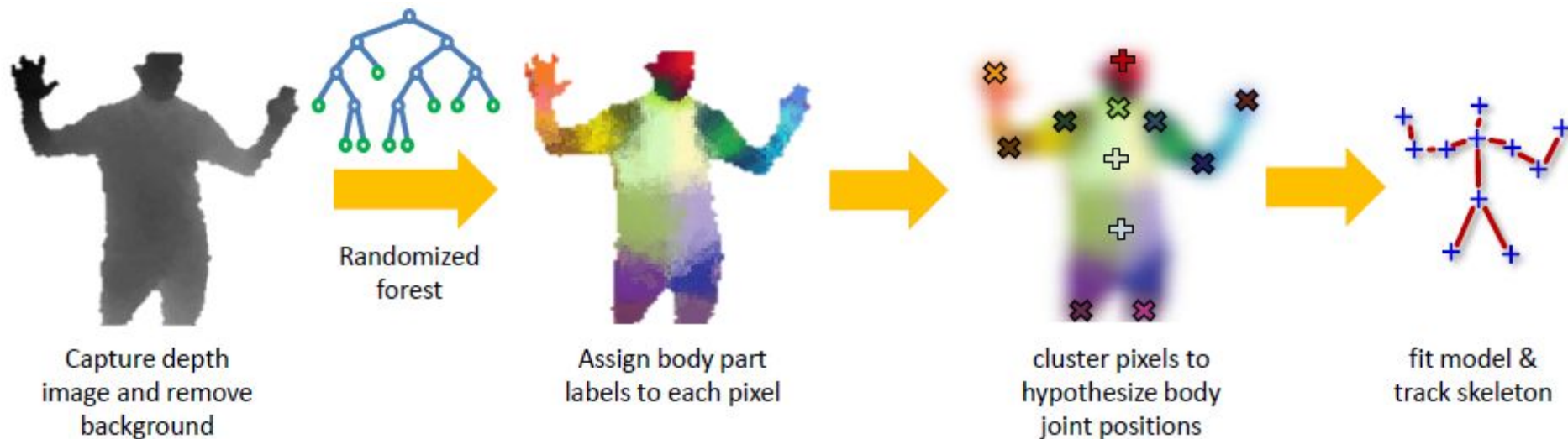
SVR: Support Vector Regression

PCA: Principal Component Analysis



## Supervised learning and the Kinect

# Supervised learning and the Kinect



Random forest: use many decision trees and pick the class which is the general consensus among the trees.

# Supervised learning and the Kinect



Figure 2. **Synthetic and real data.** Pairs of depth image and ground truth body parts. Note wide variety in pose, shape, clothing, and crop.

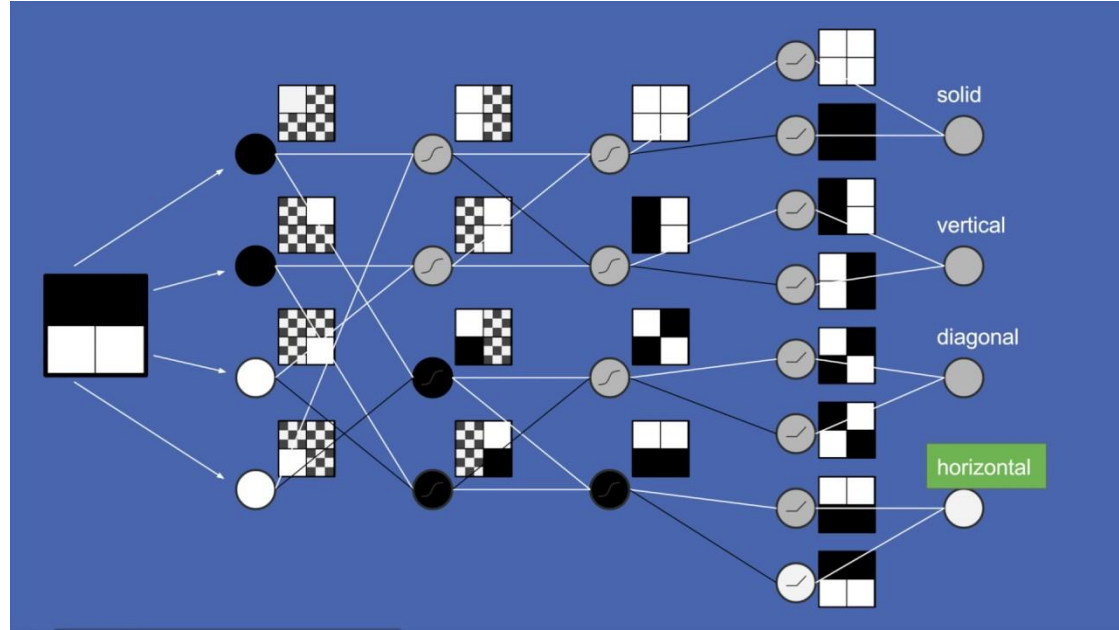
Random forest: use many decision trees and pick the class which is the general consensus among the trees.

# Supervised Learning with Neural Networks

Multiple layers of neurons which encode different levels of information in order to predict some class.

Deep network: many hidden layers.

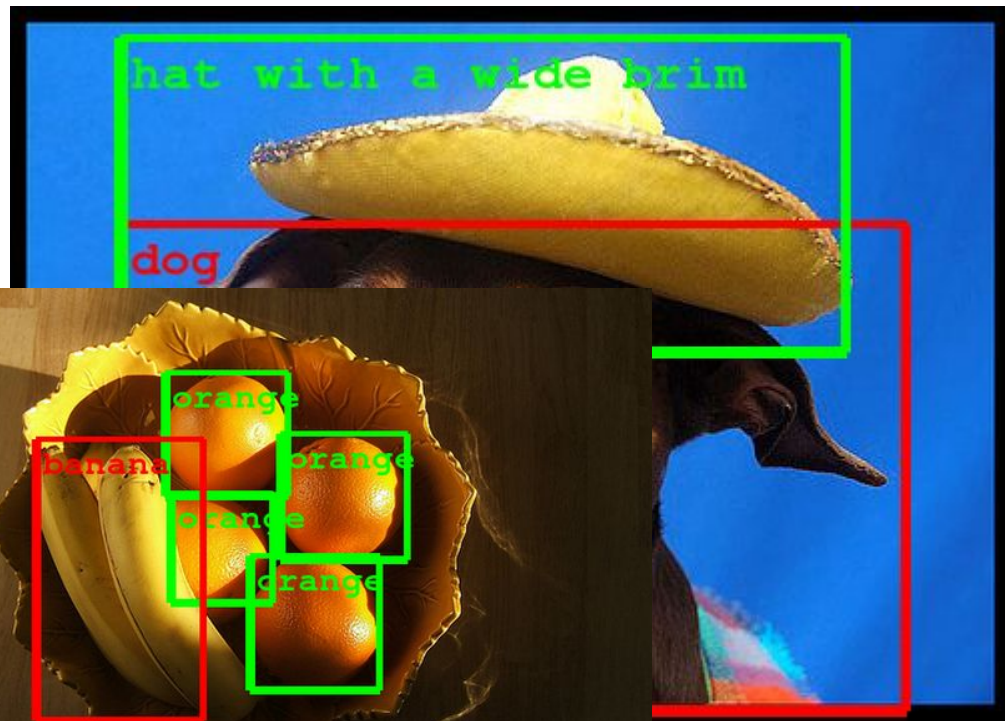
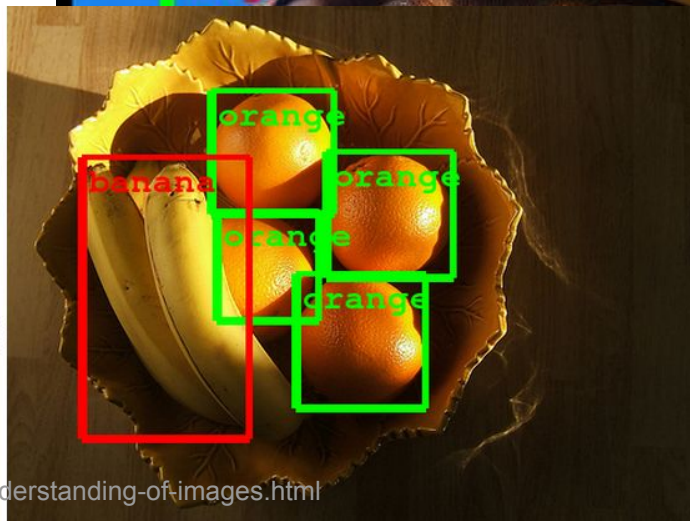
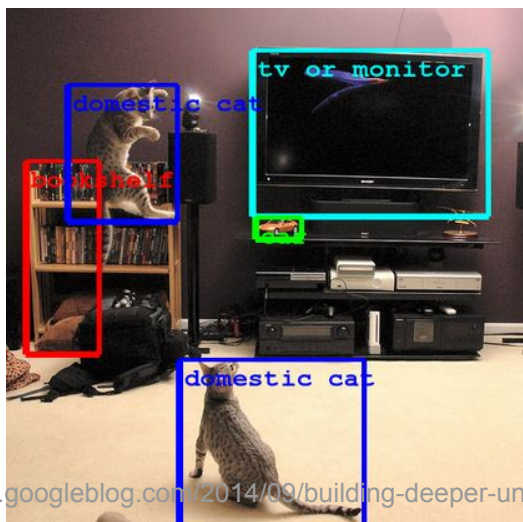
Finding the right structure for a neural network is not trivial.





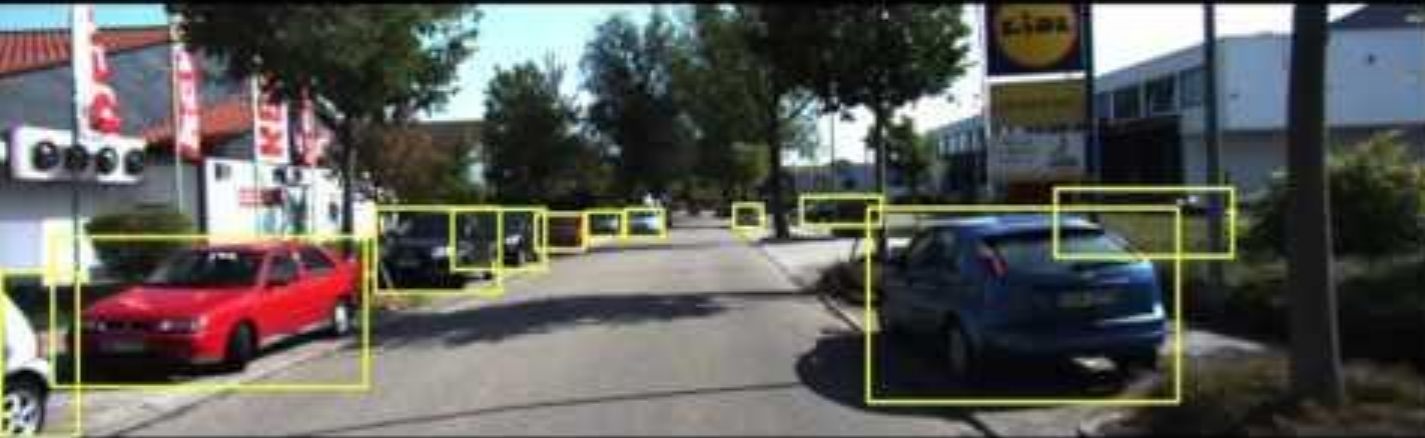
# Supervised Learning with Neural Networks

Deep learning for classification with localisation.



## Supervised Learning and Self-driving cars

Deep network to  
recognise different  
objects (classes).





# Summary

- Machine learning is the science of getting computers to act without being explicitly programmed.
- ML builds a model that can be used for predicting the future or making decisions
- **ML Types:** Supervised, Unsupervised, Reinforcement
- **ML Tasks:** Classification, Regression, Clustering
- No single algorithm is the best suited to all tasks and problems!