Machine Learning Introduction

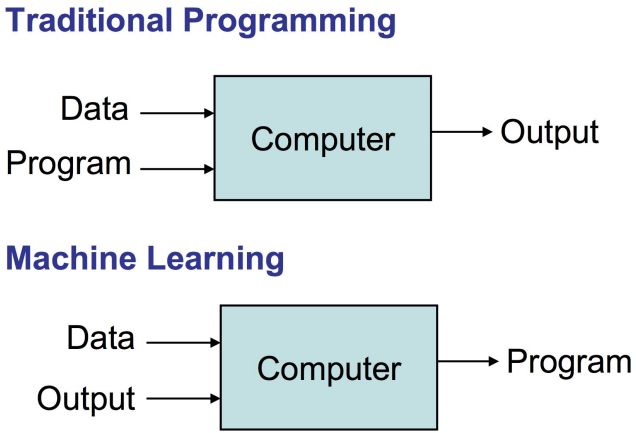
Machine Learning Introduction

1. What is ML ? Type of Learnings
2. Basic ML concepts
3. Supervised Learning
4. Unsupervised Learning
5. Other Common Terminologies

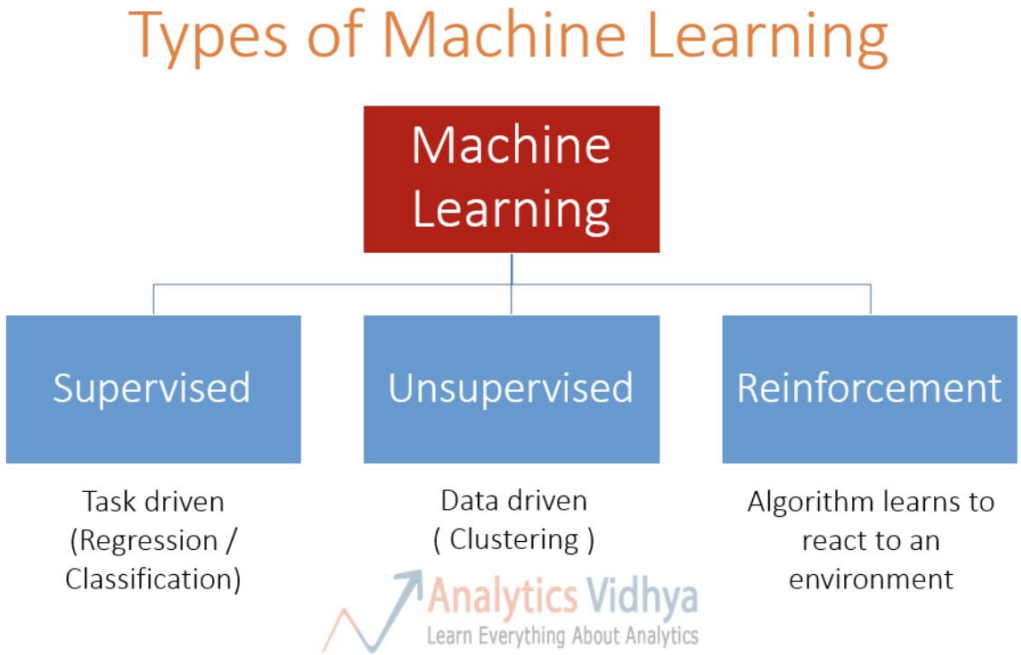
1. Machine Learning

- Getting computers to program themselves

(Writing software is the bottleneck, let the data do the work instead)



Data we have decide algorithms we use



Type of Learning - Supervised Learning

- Where a program is “trained” on a pre-defined dataset.
- Example: Using a training set of human tagged positive, negative and neutral tweets to train a sentiment analysis classifier. (Classification)

Type of Learning - Unsupervised Learning

- Where a program, given a dataset, can automatically find patterns and relationships in that dataset.
- Example: Analyzing a dataset of emails and automatically grouping related emails by topic with no prior knowledge or training. (Clustering)

  Or: Find similarities of people sitting here now.

Type of Learning - Reinforcement Learning

- Learning what to do and how to map situations to actions. Example: AlphaGo Zero (Zero developed its Go skills by competing against itself)

1. What is Machine Learning: Recap

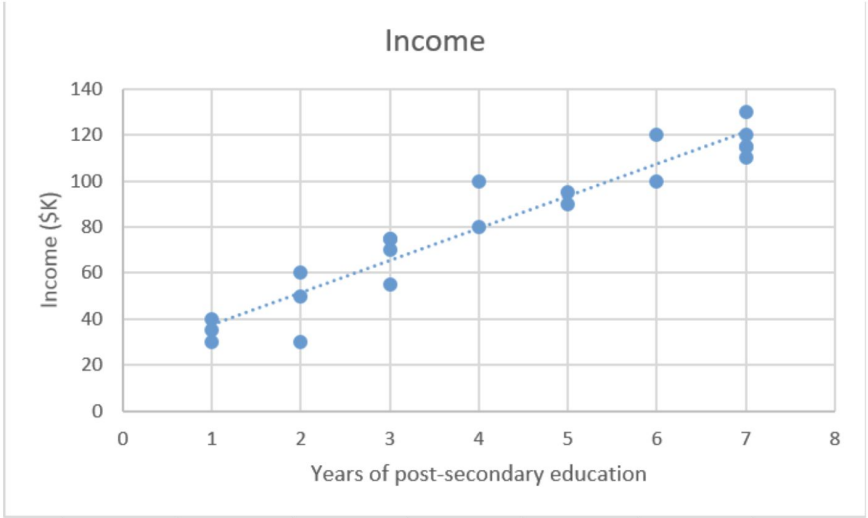
- ML is how we use data to build programs
- Basically, there are 3 types of ML: Supervised, Unsupervised, Reinforcement

2. ML Concepts: Case study

To find the best parameters:

$$Y = f(X) + \epsilon \qquad \hat{y} = \beta_0 + \beta_1 * x + \epsilon$$

1. Choose a **knowledge presentation**
2. Define a **cost function**, or **loss function**, that measures how inaccurate our model's predictions are.
3. Find the parameters that **minimize loss**, i.e. make our model as accurate as possible.



Example: Linear Regression

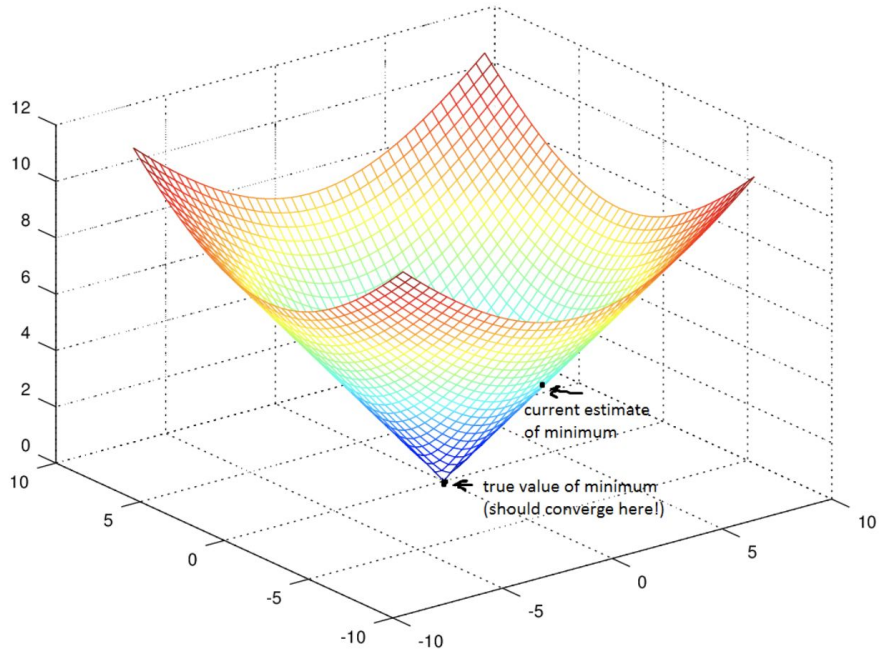
ML Concepts - Cost function/ Gradient Descent

$$Cost = \frac{\sum_1^n ((\beta_1 x_i + \beta_0) - y_i))^2}{2 * n}$$

How to minimize the cost ?

=> We use **optimization algorithms**

The most common one is **Gradient Descent** (First-order optimization algorithm)



ML Concepts - 3 key components

- **Representation**: how to represent knowledge (decision trees, rules, graphical models, neural networks, support vector machines, model ensembles ...).
- **Evaluation**: the way to evaluate candidate programs (hypotheses). (accuracy, prediction and recall, squared error, likelihood, posterior probability, cost, margin, entropy k-L divergence ...).
- **Optimization**: the way candidate programs are generated (convex optimization, constrained optimization).

2. ML Concepts: Recap

- A ML algorithm can be constructed using 3 components: representation, evaluation and optimization
- Given a problem, to build an AI program, we first choose a representation, then a way to evaluate performance, and finally, how to optimize our program

3. Supervised Learning

To learn a function f mapping between input and output

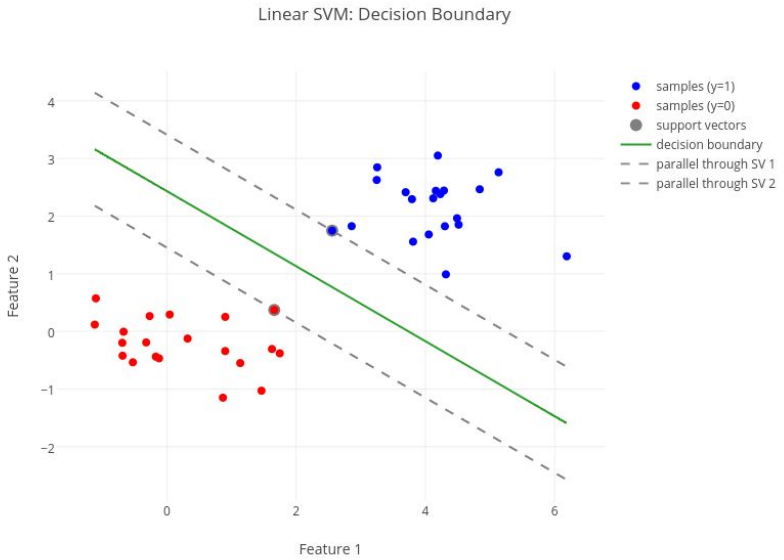
$$Y = f(X) + \epsilon$$

Regression: predict a continuous numerical value. *How much will that house sell for?*

Classification: assign a label. *Is this a picture of a cat or a dog?*

Supervised Learning: Classification

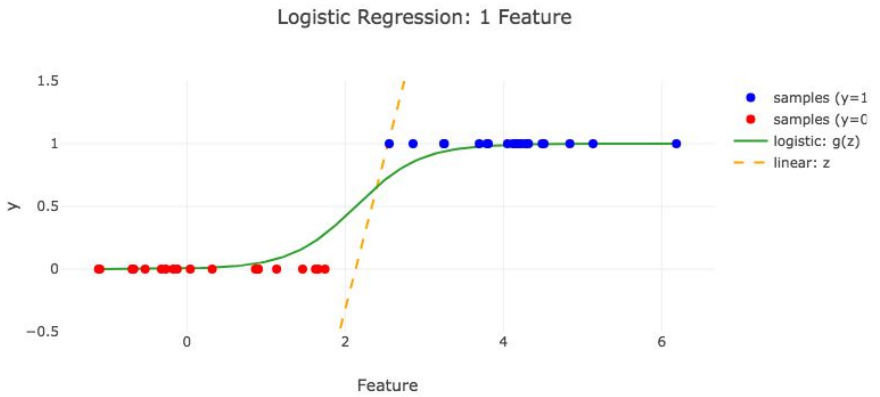
SVM: Support Vector Machine



Logistic Regression: quite tricky to understand at the beginning

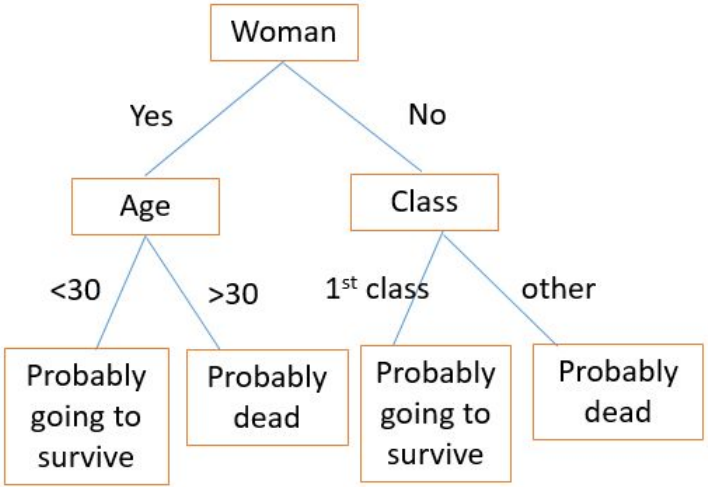
$$g(z) = \frac{1}{1+e^{-z}} \qquad \text{with}$$

$$z = f(x) = w_0 + w_1 x_1 + \ldots + w_n x_n$$



Supervised Learning: Classification

Decision Tree with Titanic data



Data Dictionary

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

Source: Kaggle

4. Unsupervised Learning

- Clustering: k-means, hierarchical clustering, Latent Dirichlet Allocation
- Data Compression: Principal Component Analysis (PCA), Singular-Value Decomposition (SVD)

PCA / SVD looks for a low-dimensional representation of the observations that explains a good fraction of the variance

Clustering looks for homogeneous subgroups among the observations.

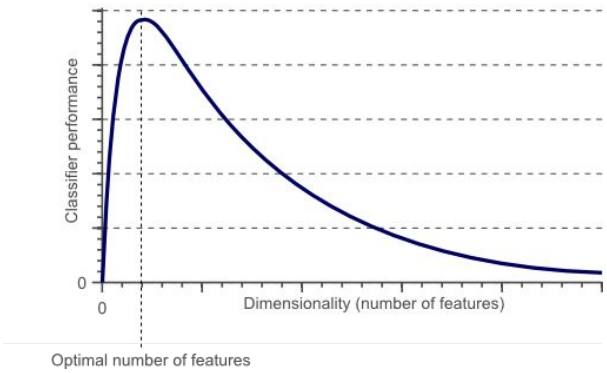
Unsupervised Learning

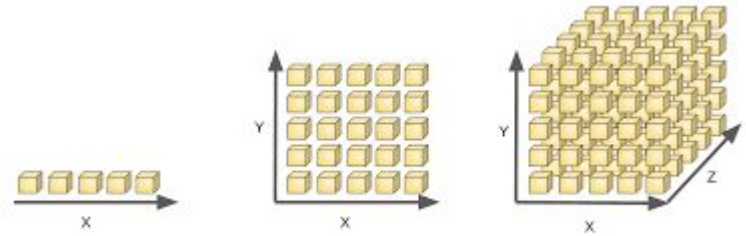
LDA and Topic Modelling: there are some latent topics of documents that are not observed. Each document has a distribution over these topics

Trig/Family/Inspiration: family, web, mail, god, son, from, congratulations, children, life, child, down, trig, baby, birth, love, you, syndrome, very, special, bless, old, husband, years, thank, best, …
Wildlife/BP Corrosion: game, fish, moose, wildlife, hunting, bears, polar, bear, subsistence, management, area, board, hunt, wolves, control, department, year, use, wolf, habitat, hunters, caribou, program, denby, fishing, …
Energy/Fuel/Oil/Mining: energy, fuel, costs, oil, alaskans, prices, cost, nome, now, high, being, home, public, power, mine, crisis, price, resource, need, community, fairbanks, rebate, use, mining, villages, …
Gas: gas, oil, pipeline, agia, project, natural, north, producers, companies, tax, company, energy, development, slope, production, resources, line, gasline, transcanada, said, billion, plan, administration, million, industry, …
Education/Waste: school, waste, education, students, schools, million, read, email, market, policy, student, year, high, news, states, program, first, report, business, management, bulletin, information, reports, 2008, quarter
Presidential Campaign/Elections: mail, web, from, thank, you, box, mccain, sarah, very, good, great, john, hope, president, sincerely, wasilla, work, keep, make, add, family, republican, support, doing, p.o

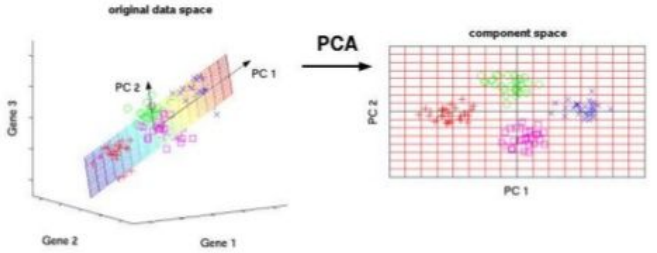
Unsupervised Learning

PCA and Curse of Dimensionality





Use of PCA
(**P**rincipal **C**omponent **A**nalysis)



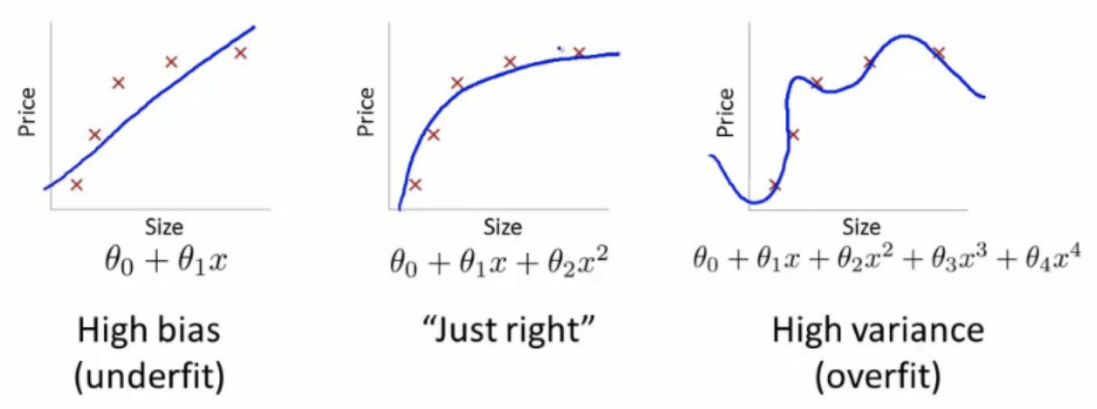
Picture from *http://www.nlpca.org/pca_principal_component_analysis.html*

3. and 4. Supervised and Unsupervised Learning Recap

- The number of algorithms for both is big
- For a problem, there are many algorithms that can be used => try and error.

5. ML Terminologies - Overfit/Underfit

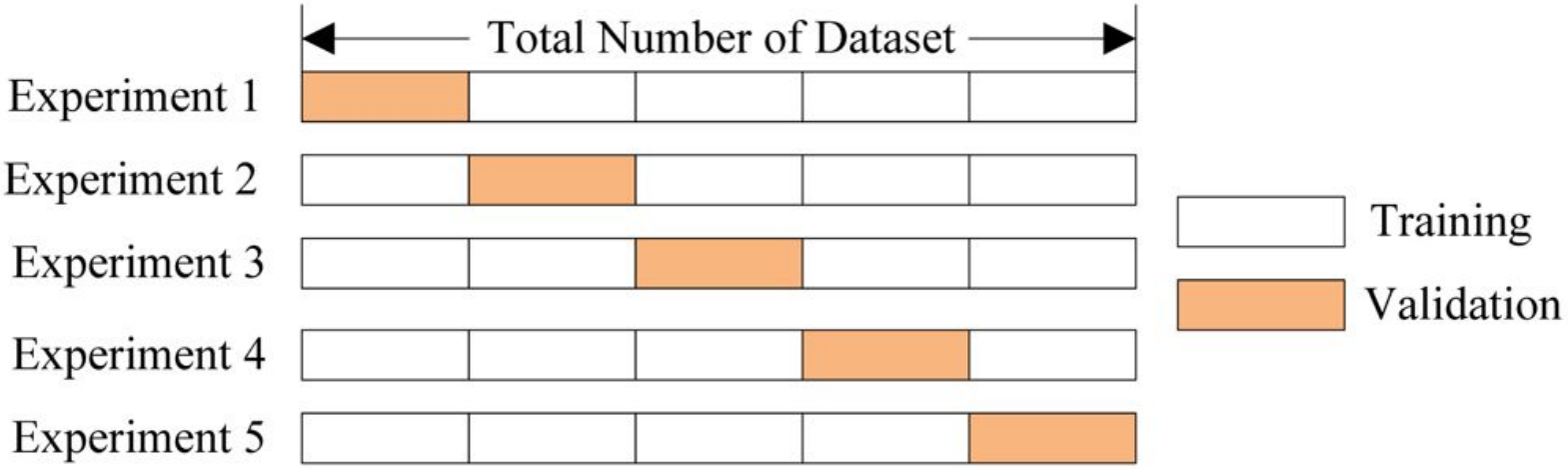
To learn a function f mapping between input and output



Source: Coursera's ML course, taught by Andrew Ng

ML Terminologies - Cross Validation

Used for Model checking: to determine the performance of models



ML Terminologies - Batch/Online/Mini-batch

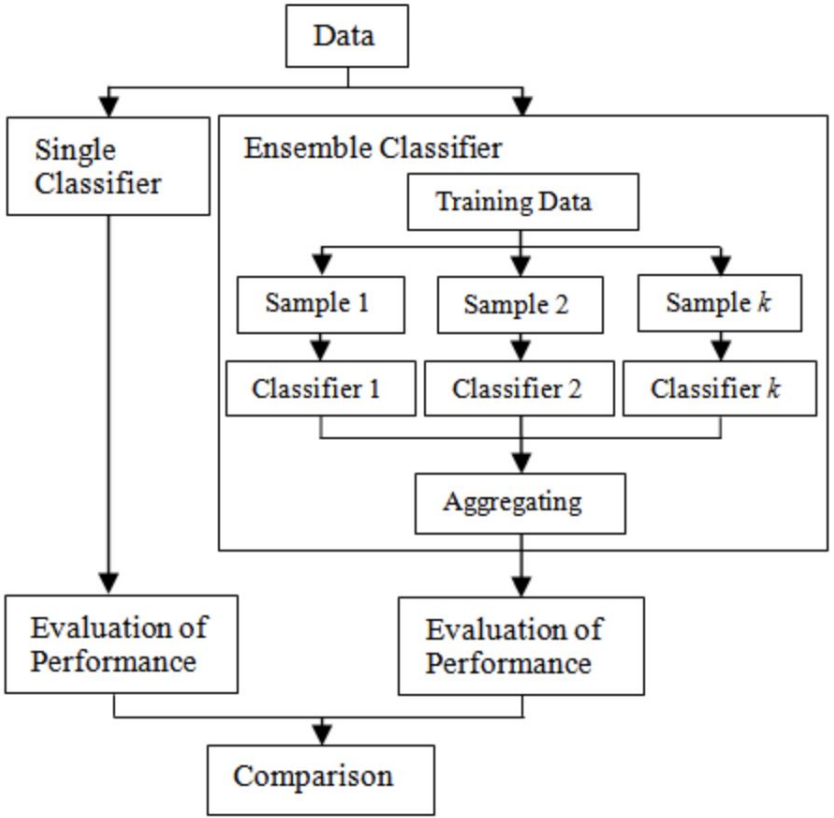
Given a training dataset of 1M samples:

- Full-batch learning: process all at once.
- Online-learning where for every iteration we randomly pick a training case
- Mini-batch learning where for every iteration we randomly pick 100 training cases

ML Terminologies - Ensemble Models

Ensemble is combining many different independent machine learning models

- Bagging: create subsets of training data set. Then, build a classifier for each subset.
- Boosting: Train a model on a set of data that is incorrectly predicted by previous models.
- Stacking: train a model based on output from other models



Thank you !

Q & A