26th May 2017

INTRO TO MACHINE LEARNING – NOCA, UK

Gaurang Mehta, FIA

INTRO TO MACHINE LEARNING (ML)

AGENDA

- Background on ML
- Introduction to ML
- KNN Algorithm
- Naïve Base Algorithm
- Further Sources
- Q & A

INTRO TO MACHINE LEARNING (ML)

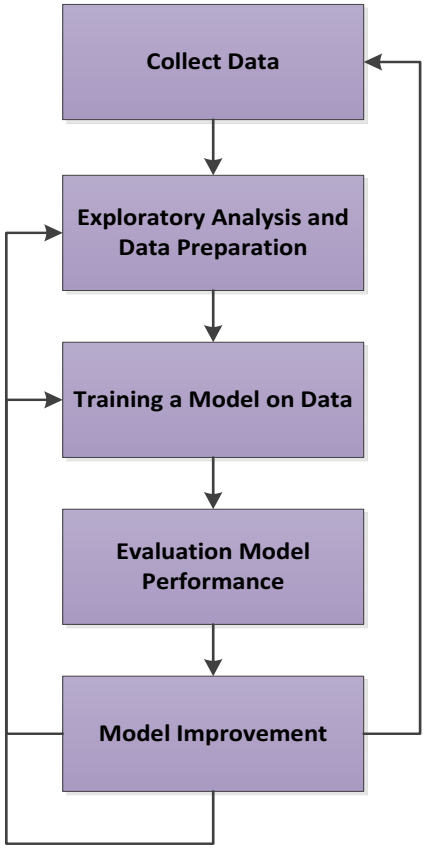
- **Background and Introduction to ML**

MACHINE LEARNING - BACKGROUND

- "Every big start-up over the next 5 years will have one thing in common: machine learning" - Eric Holt (Google)

- "ML is as big of a breakthrough as personal computers, the internet, or electricity itself" - Pedro Domingos (Uni. Washington)

HOW DO MACHINES LEARN?



MACHINE LEARNING – DEFINITION AND TERMINOLOGY

Machine Learning deals with the design of programs that can learn rules from

- *Data;*
- *Adapt to changes; and*
- *Improve performance with experience.*

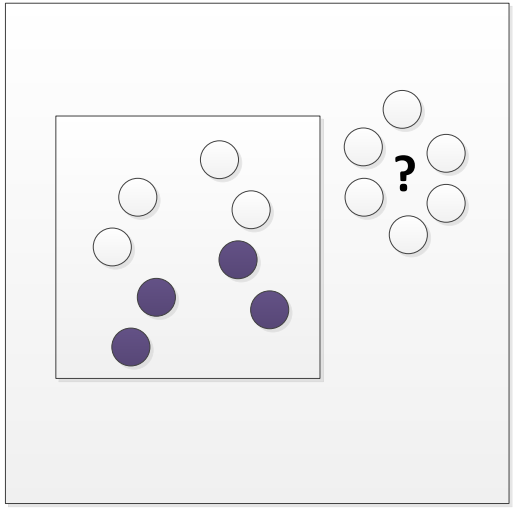
ML is often used in situations where the

- Predictive accuracy of a model is **more** important than the interpretability of a model.
- More variables than observations and / or Many correlated variables
- Unstructured data, e.g. texts or emails

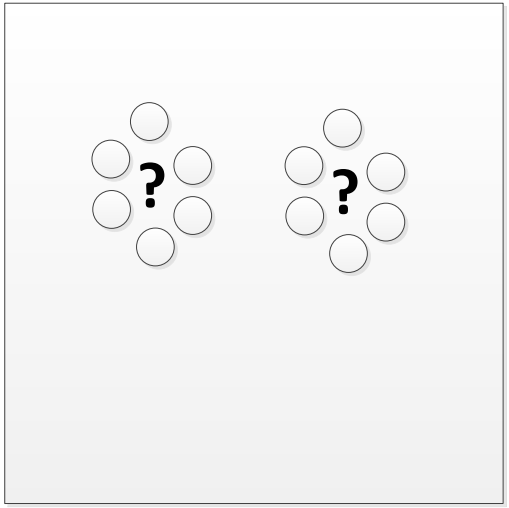
ML Term	Statistics Term
Case, instance, example	Observation, record, row, data point
Feature, input	Independent variable, variable, column
Label	Dependent variable, target
Class	Categorical target variable level
Train	Fit
Score	Predict

ML – LEARNING SCENARIOS

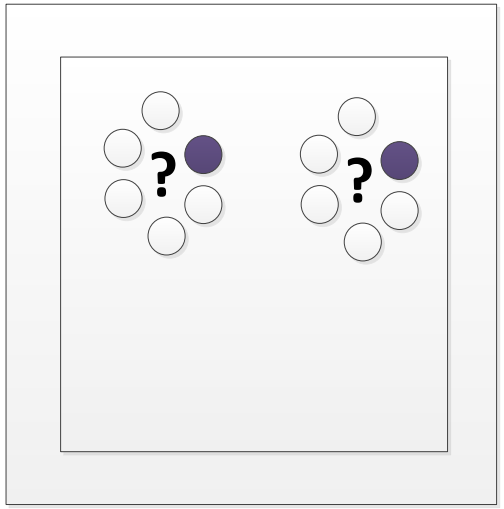
Supervised Leaning



Unsupervised Leaning



Semi- Supervised Learning



Supervised Learning
• Classification • Regression • Ranking Problems • Neural Networks • Decision Trees

Unsupervised Learning
• Clustering • Kernel Density Estimates • Dimensional Reduction Techniques such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD)

Semi-Supervised Learning
• Prediction and Classification Problems with Clustering • Autoencoders • Expectation Maximisation

- **K-Nearest Neighbours**

K-NEAREST NEIGHBOURS ("LAZY LEARNING")

- **Knn identifies k records in the training data set that are "nearest" in similarity.**

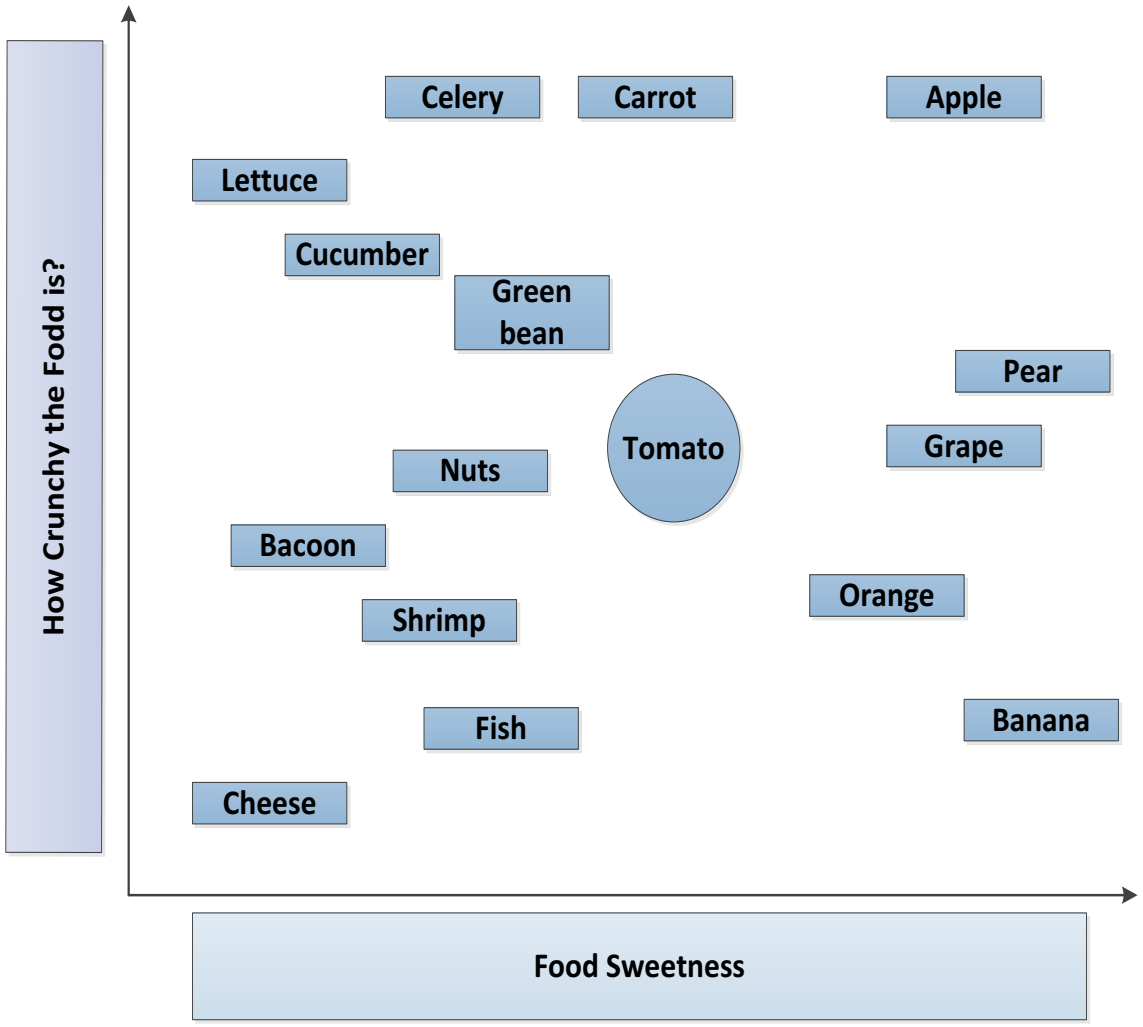
Applications:

- **Identification of patterns in genetic data**
- **Facial & optical character recognition**
- **Classification problems**

Strength	Weakness
Simplicity	Does not produce a model and therefore limited understanding of model features
Quick training	Slow classification phase
Impartial to underlying data distributions	

K-NEAREST NEIGHBOURS – "TOMATO" EXAMPLE

- **Question : Is tomato a "Fruit" or a "Vegetable"?**

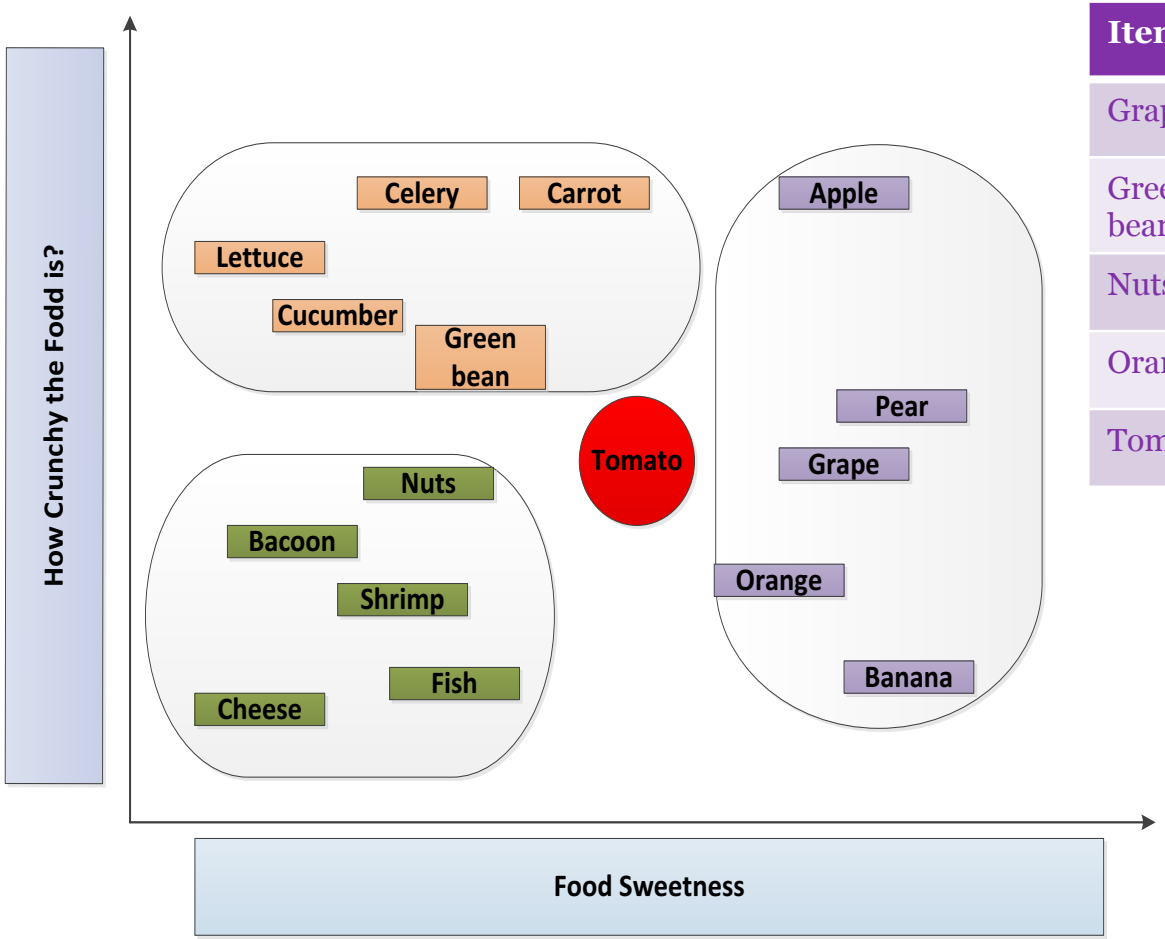


Item	Sweetness	Crunchiness
Grape	8	5
Green bean	3	7
Nuts	3	6
Orange	7	3
Tomato	6	4

K-NEAREST NEIGHBOURS – "TOMATO" EXAMPLE CONTINUED

- **Measure: "Euclidean" Distance**

Distance = $sqrt(\sum_{i=0}^{n}(s_{Tomato} - s_i)^2 + (c_{Tomato} - c_i)^2)$



Item	Sweetness	Crunchiness	Dist
Grape	8	5	2.2
Green bean	3	7	4.2
Nuts	3	6	3.6
Orange	7	3	1.4
Tomato	6	4	0

k	Conclusion
1	Fruit
2	Fruit
3	Still Fruit
4	Still Fruit

K-NEAREST NEIGHBOURS – BREAST CANCER EXAMPLE

- Data : Wisconsin Breast Cancer Data, total 569 observations
- Attributes: Radius, Texture, Perimeter, Area, Smoothness, etc.
- Subset: Training 469 obs. , Test 100 obs.

Total Observations in Table: 100

wbcd_test_labels	wbcd_test_pred_k20 Benign	Malignant	Row Total
Benign	77 (TN)	0 (FP)	77
	1.000	0.000	0.770
	0.975	0.000	
	0.770	0.000	
Malignant	2 (FN)	21 (TP)	23
	0.087	0.913	0.230
	0.025	1.000	
	0.020	0.210	
Column Total	79	21	100
	0.790	0.210	

- Settings:
- K = 20
- Accuracy = 98%

K	Accuracy
1	93%
5	97%
10	99%
15	98%

K-NEAREST NEIGHBOURS – SYNOPSIS

- Knn is a classification algorithm that stores the training data
- Knn does not learn
- Knn uses distance function to classify unlabelled example
- "Bias-Variance" trade-off is important in determining K

$$Err(x) = \left(f(x) - \frac{1}{k} * \sum_{i=1}^{k} f(x_i) \right)^2 + \frac{\sigma_\epsilon^2}{k} + \sigma_\epsilon^2$$

$$Err(x) = Bias^2 + Variance + Irreducible\ Error$$

- High Var & Low Bias vs. Low Var & High Bias
- Bagging or resampling techniques help to reduce variance
- Bias reduces generally as the size of the data increases

- **“Naïve” Bayes Algorithm**

"NAÏVE" BAYES

- **It is "naïve" because it makes "naïve" assumptions about the data.**

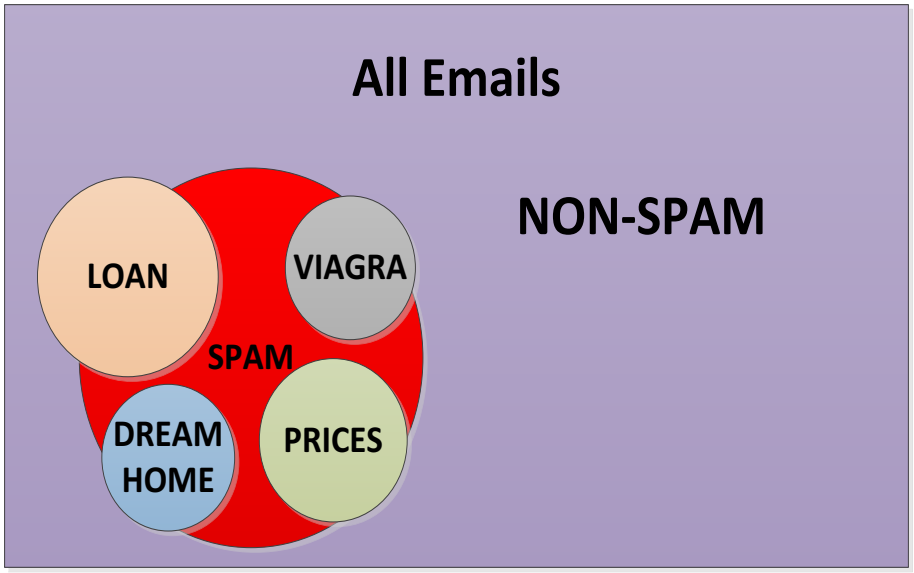
Applications:

- **Anomaly detection in computer networks**
- **Text classification**
- **Diagnosing medical condition given a set of observed symptoms**

Strength	Weakness
Simplicity	Relies on equally important and independent features
Performs well with noisy and missing data	Not ideal for dataset with numeric values
Works well with large sample sizes	

NAÏVE BAYES ALGORITHM – EMAIL SPAM EXAMPLE

- **Question : How to detect which email is spam?**



"Valentine Day Special!!! Win £100,000 in our quiz and get a free dream retirement home"

"Congratulations!! you are awarded £10000 price and a free entry 210 weekly draw"

$$P\left(\frac{Spam}{L \cap \neg H \cap P \cap \neg V}\right) = \left(\frac{P(L \cap \neg H \cap P \cap \neg V / Spam)}{P(L \cap \neg H \cap P \cap \neg V)}\right) * P(Spam)$$

Conditional Independence under "Naïve" Bayes

$$P\left(\frac{Spam}{L \cap \neg H \cap P \cap \neg V}\right) = \left(\frac{P\left(\frac{L}{Spam}\right) * P\left(\frac{\neg H}{Spam}\right) * P\left(\frac{P}{Spam}\right) * P\left(\frac{\neg V}{Spam}\right)}{P(L \cap \neg H \cap P \cap \neg V)}\right) * P(Spam)$$

K-NEAREST NEIGHBOURS – DIABETES EXAMPLE

- **Data : Diabetes Database, total 768 observations**
- **Attributes: No. of pregnancies, Glucose level, Blood pressure, Skin thickness, Serum insulin, BMI, Age, etc.**
- **Subset: Training 600 obs. , Test 168 obs.**

Total Observations in Table: 168

predicted	actual N	actual P	Row Total
N	92 (TN) 0.793 0.852	24 (FN) 0.207 0.400	116 0.690
P	16 (FP) 0.308 0.148	36 (TP) 0.692 0.600	52 0.310
Column Total	108 0.643	60 0.357	168

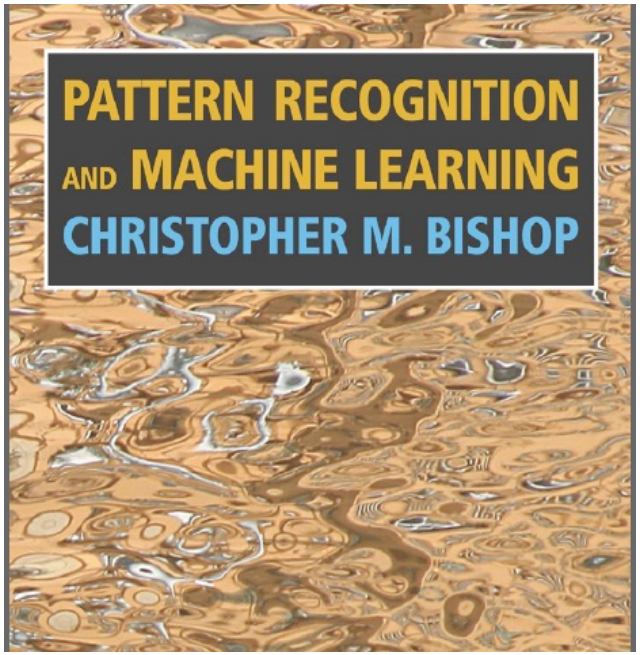
Total Observations in Table: 168

predicted	actual N	actual P	Row Total
N	94 (TN) 0.790 0.870	25 (FN) 0.210 0.417	119 0.708
P	14 (FP) 0.286 0.130	35 (TP) 0.714 0.583	49 0.292
Column Total	108 0.643	60 0.357	168

NAÏVE BAYES – SYNOPSIS

- **Naïve bayes is a classification algorithm that uses Bayes theorem**
- **It constructs tables of probabilities for classification**
- **Relies on naïve assumption of conditional independence**
- **Simple but versatile algorithm widely used for identification of spam "texts" and "emails"**

STILL INTERESTED?



http://machinelearningmastery.com/start-here/ → Blogs by Dr. Jason Brownlee

http://archive.ics.uci.edu/ml/ →University of California, Irvine

Machine Learning Course – Stanford University on Coursera by Andrew Ng

QUESTIONS?