Introduction to Machine Learning

Odysseas Pentakalos and Alex Gilgur

sysnet INTERNATIONAL

Overview

- Define AI and Machine Learning
- See a few examples of Machine Learning
- Talk about some key concepts
- Review the Machine Learning Process
- Take a peek at a few algorithms
- Look at an example in detail

sysnet INTERNATIONAL

What is Artificial Intelligence

John McCarty coined the term in 1956

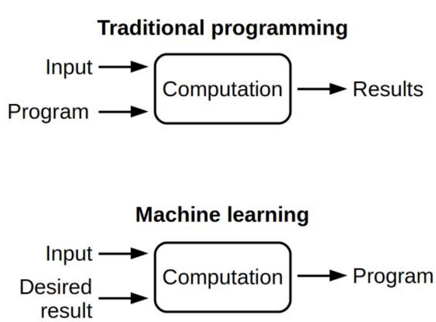
Merriam-Webster defines artificial intelligence this way:

- A branch of computer science dealing with the simulation of intelligent behavior in computers.
- The capability of a machine to imitate intelligent human behavior.

Bernard Marr,
https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance

What is Machine Learning

"Machine learning is the science of getting computers to act without being explicitly programmed, but instead letting them learn a few tricks on their own." by Danko Nikolic



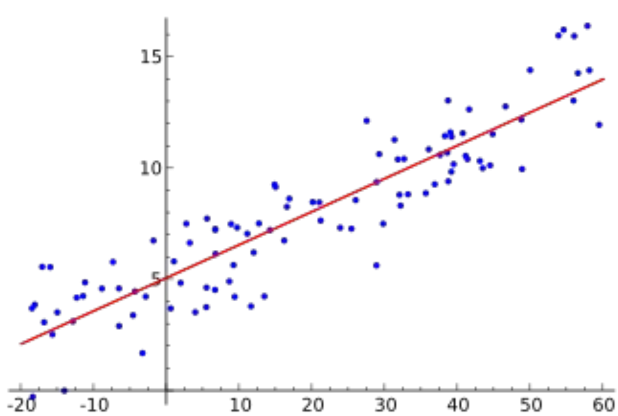
Applications

Machine Learning is everywhere!

- Financial
- Healthcare
- Retail
- ...

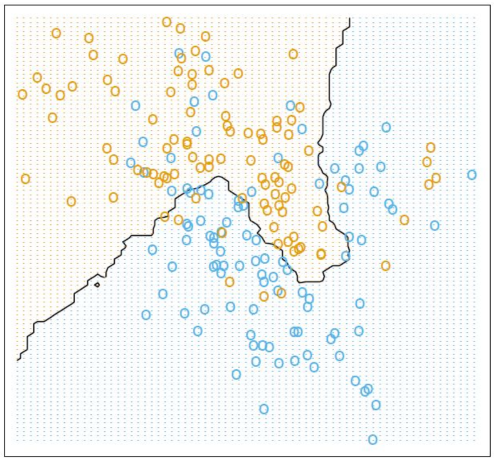
Algorithms - Regression

- Regression Analysis: estimate the relationships between variables



Algorithms - Classification

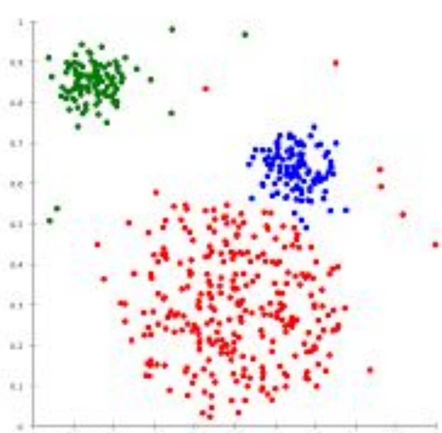
- Classification: identify category an observation belongs to



sysnet INTERNATIONAL

Algorithms - Clustering

- Clustering: group together objects that are similar

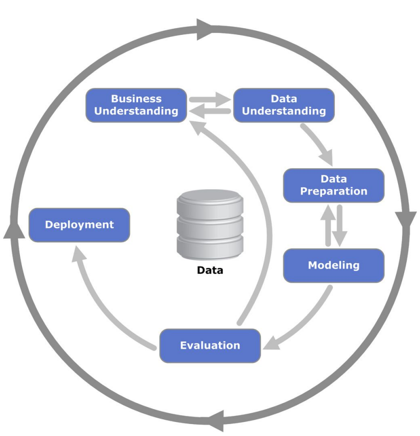


sysnet INTERNATIONAL

Algorithms - Recommender

- Recommender System: produce recommendations for a given selection

Machine Learning Process



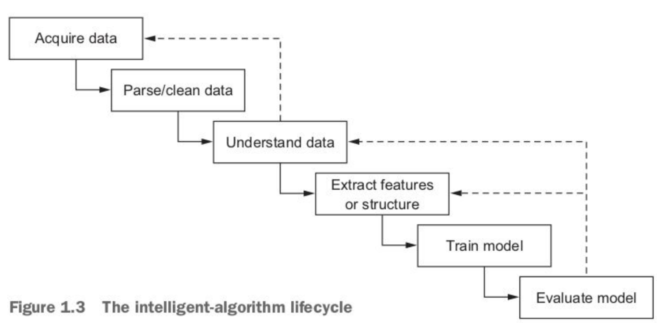
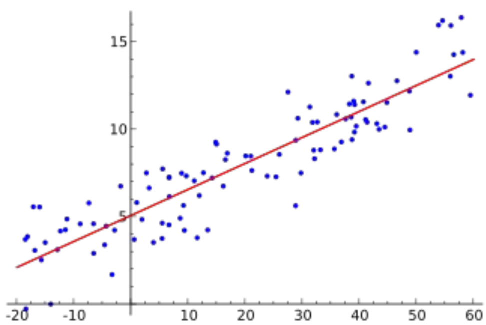


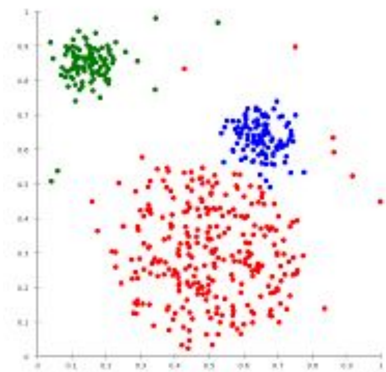
Figure 1.3 The intelligent-algorithm lifecycle

[6] Ben Fry, PhD thesis, *Computational Information Design* (MIT, 2004).

Algorithm Types

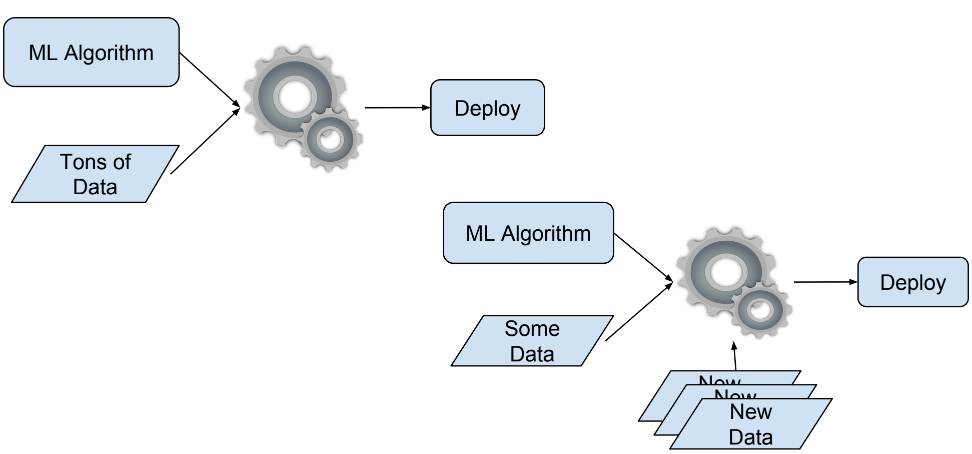
Supervised vs Unsupervised





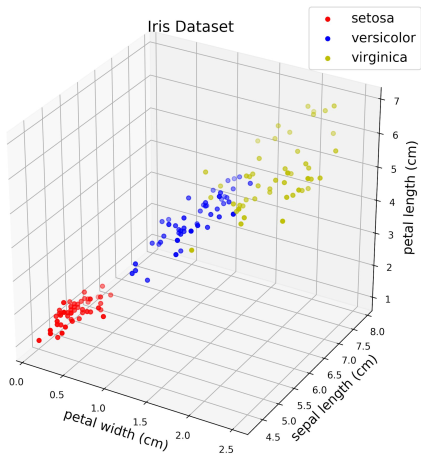
Algorithm Types

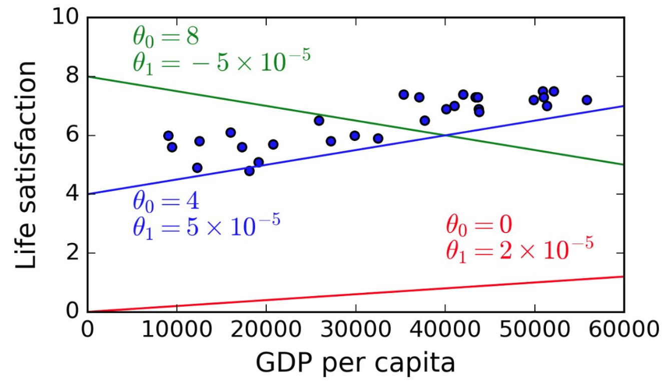
Batch vs Online



Algorithm Types

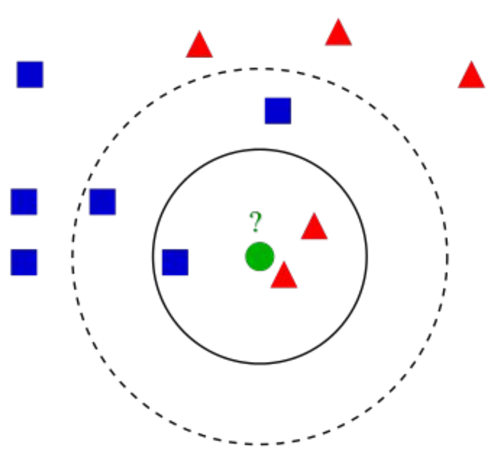
Instance Based vs Model Based





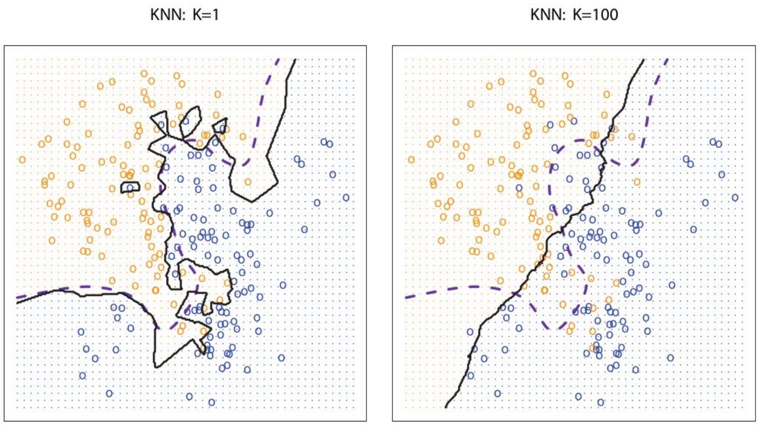
k-Nearest Neighbors

- A sample is most likely similar to its neighbors



k-Nearest Neighbors

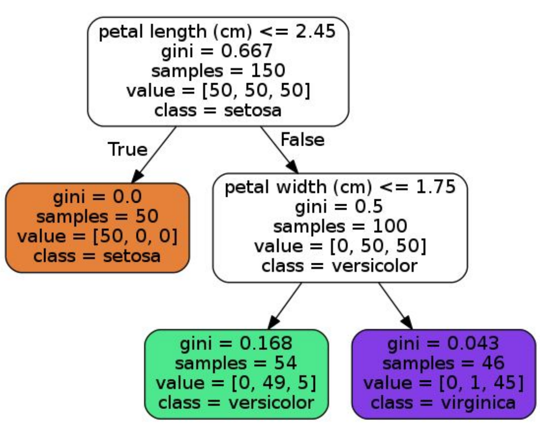
The parameter k defines the decision boundary



From: “Introduction to Statistical Learning” by Gareth James, et al

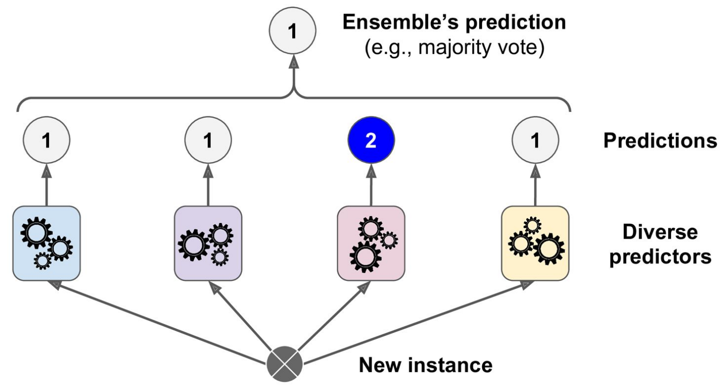
Decision Trees

- Classify samples based on a sequence of decisions on attribute values
- Many different ways to build a tree



Ensembles

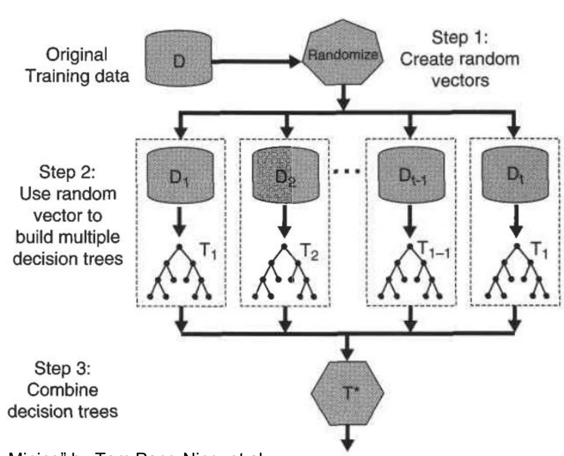
- Use multiple models to improve performance



From: "Hands-on Machine Learning with Python, Keras, and Tensor Flow" by Aurélien Géron.

Decision Trees (Next Generation)

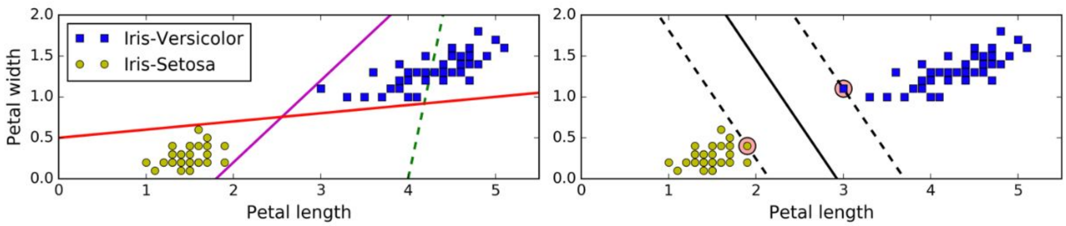
- Random Forests, Bagging and Boosting



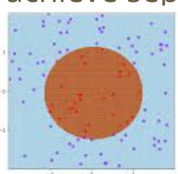
From: "Introduction to Data Mining" by Tam Pang-Ning, et al

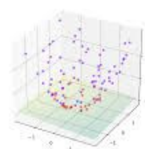
Support Vector Machines

- Goal is to achieve maximum separation



- Add dimensions to achieve separation



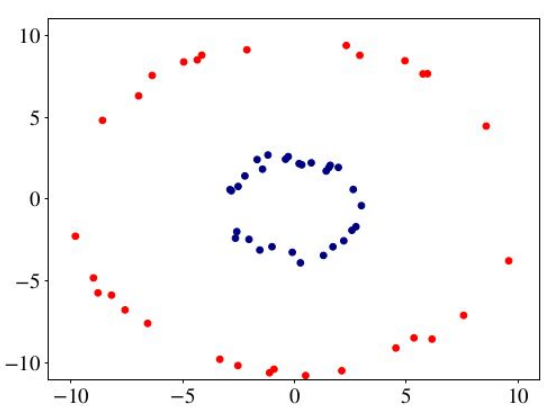


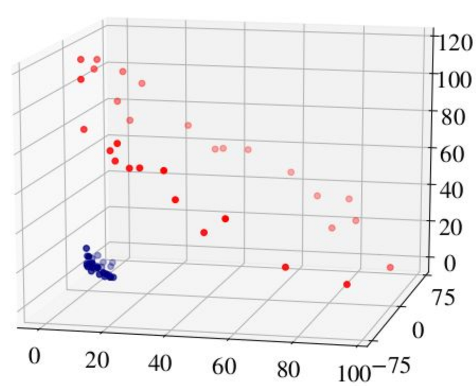
From: “Hands-on Machine Learning with Python, Keras, and Tensor Flow” by Aurélien Géron.

sysnet INTERNATIONAL

SVM Kernel Trick

Move data into higher dimensions to make linear separability possible

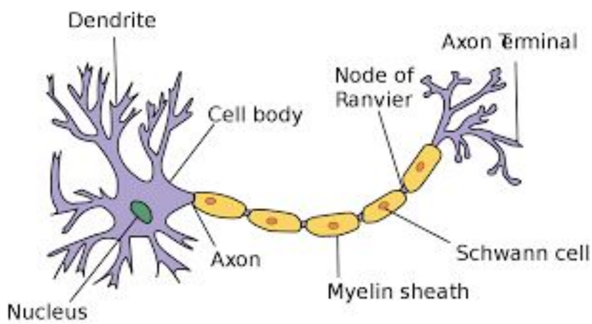


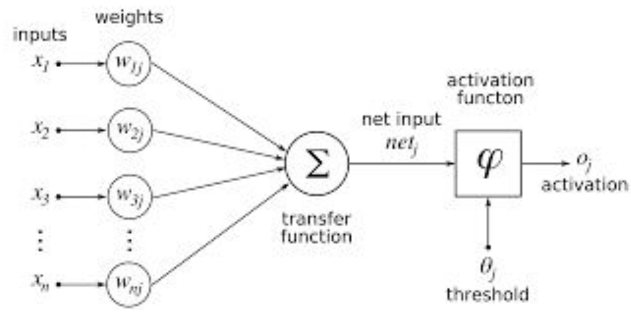


sysnet INTERNATIONAL

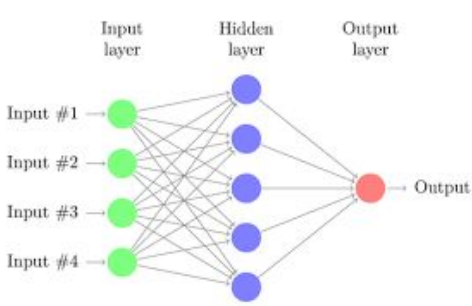
Neural Networks

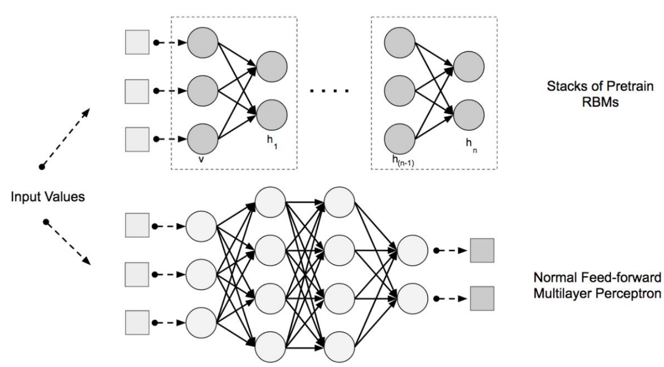
Why not simulate the human brain?





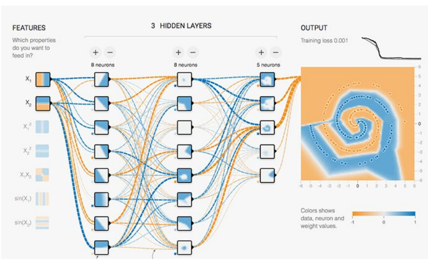
Deep Learning

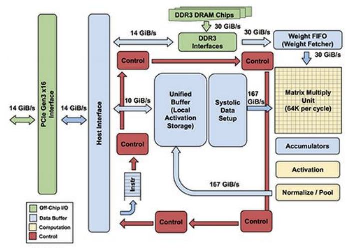


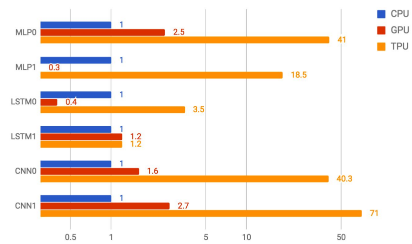


Hardware support for Deep Learning

- Custom made TPUs with Quantization and CISC architecture

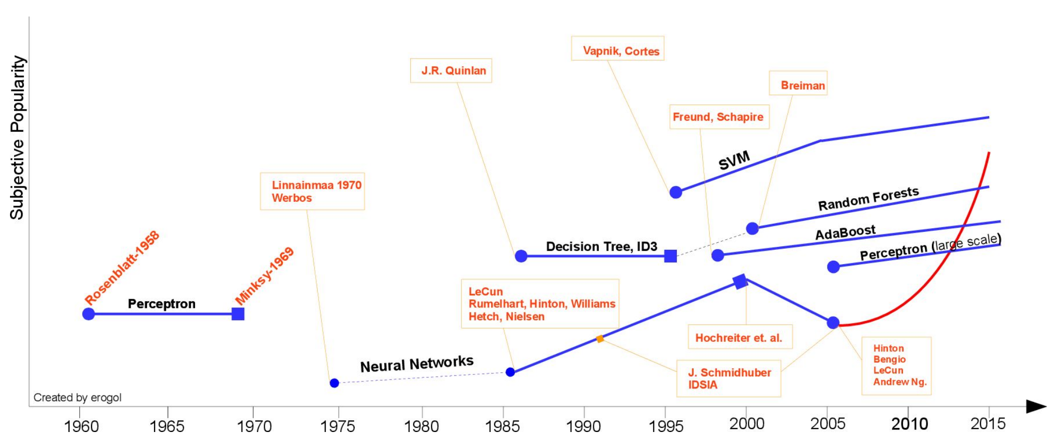






Source: “https://cloud.google.com/blog/products/gcp/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu”

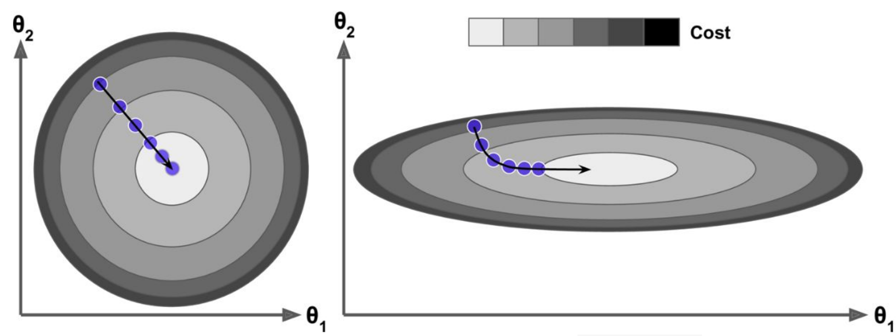
History of Machine Learning



Source: “http://www.erogol.com/brief-history-machine-learning/”

Feature Engineering

- Select features
- Transform features: feature scaling
- Example: Gradient Descent and the need for Feature Scaling



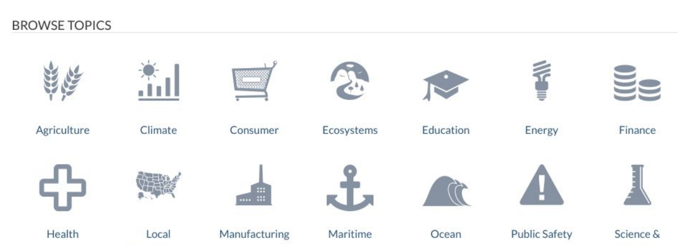
From: “Hands-on Machine Learning with Python, Keras, and Tensor Flow” by Aurélien Géron.

sysnet INTERNATIONAL

Datasets for Learning ML

Plenty of open datasets to work with:

- UC Irvine Machine Learning Repository
- Kaggle datasets
- Amazon AWS datasets
- US Government open data



sysnet INTERNATIONAL

Importance of Data

Quantity is more important than algorithm tuning

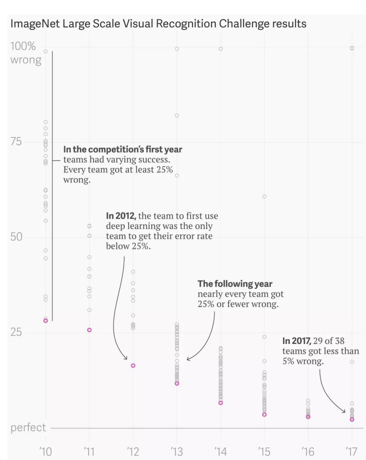
- Deep Learning needs data
- Some problem domains benefit from more data

… but

- The quality of the data is important
- Look out for bias in the data

sysnet INTERNATIONAL

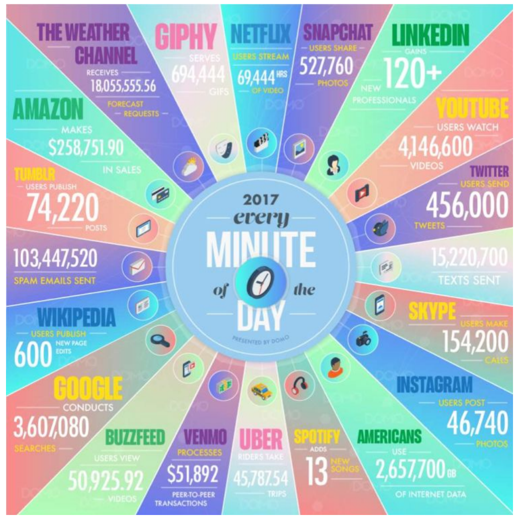
ImageNet Competition



https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/

sysnet INTERNATIONAL

...and we are generating a lot of it



Metrics for Evaluation

Evaluating classification algorithms

- Accuracy: $$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

Let's try calculating accuracy for the following model that classified 100 tumors as malignant (the positive class) or benign (the negative class):

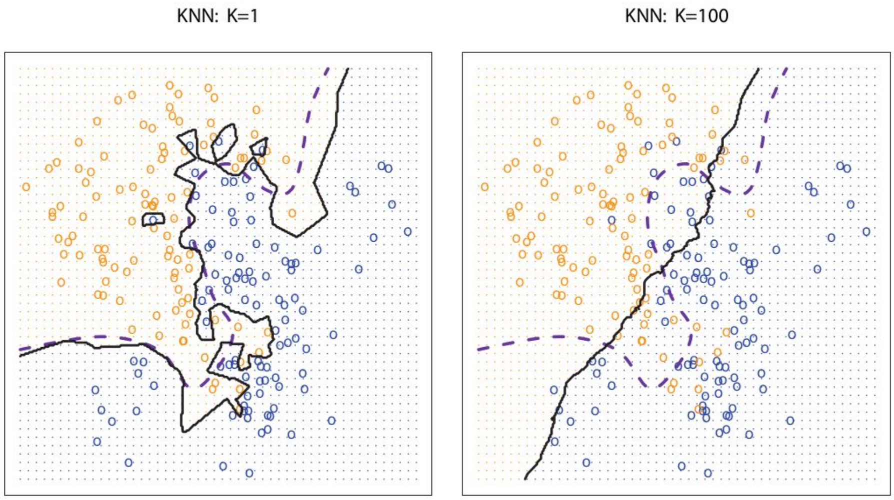
True Positive (TP): • Reality: Malignant • ML model predicted: Malignant • **Number of TP results: 1**	**False Positive (FP):** • Reality: Benign • ML model predicted: Malignant • **Number of FP results: 1**
False Negative (FN): • Reality: Malignant • ML model predicted: Benign • **Number of FN results: 8**	**True Negative (TN):** • Reality: Benign • ML model predicted: Benign • **Number of TN results: 90**

Regression:
Mean Squared Error

$$\sum_{i=1}^{n} \frac{\left(w^T x(i) - y(i)\right)^2}{n}$$

Source: https://developers.google.com/machine-learning/crash-course/classification/accuracy

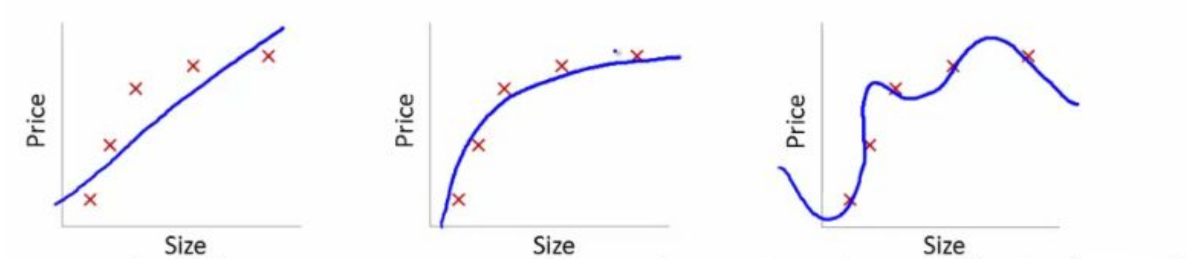
Overfitting/Underfitting



From: “Introduction to Statistical Learning” by Gareth James, et al

sysnet INTERNATIONAL

Underfitting vs Overfitting



$\theta_0 + \theta_1 x$

High bias (underfit)

$\theta_0 + \theta_1 x + \theta_2 x^2$

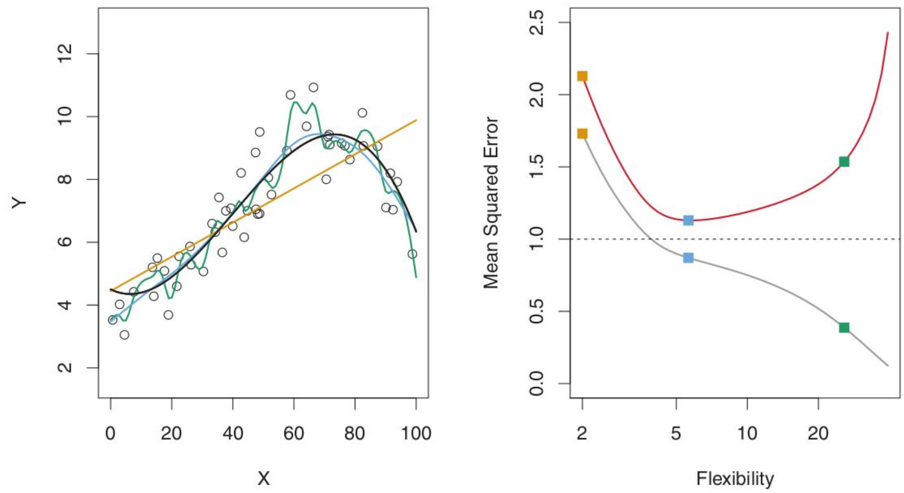
“Just right”

$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

High variance (overfit)

sysnet INTERNATIONAL

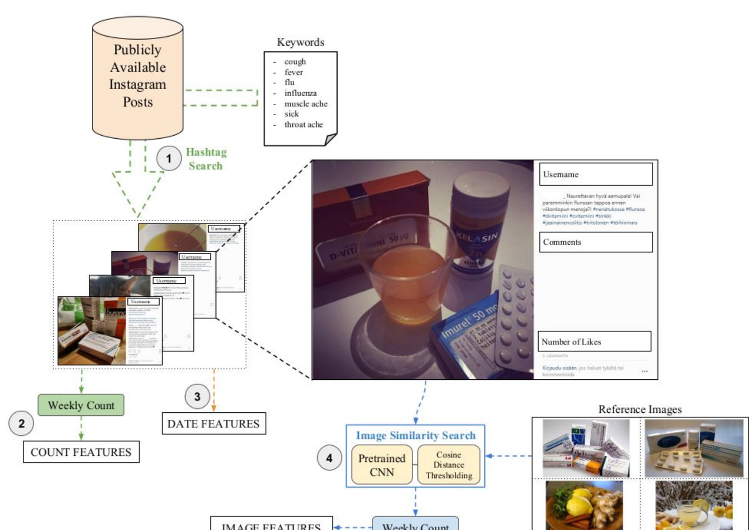
Overfitting/Underfitting



From: "Introduction to Statistical Learning" by Gareth James, et al

Machine Learning Application

Predicting the Flu



From: "Predicting the Flu from Instagram" by Gencoglu and Ermes

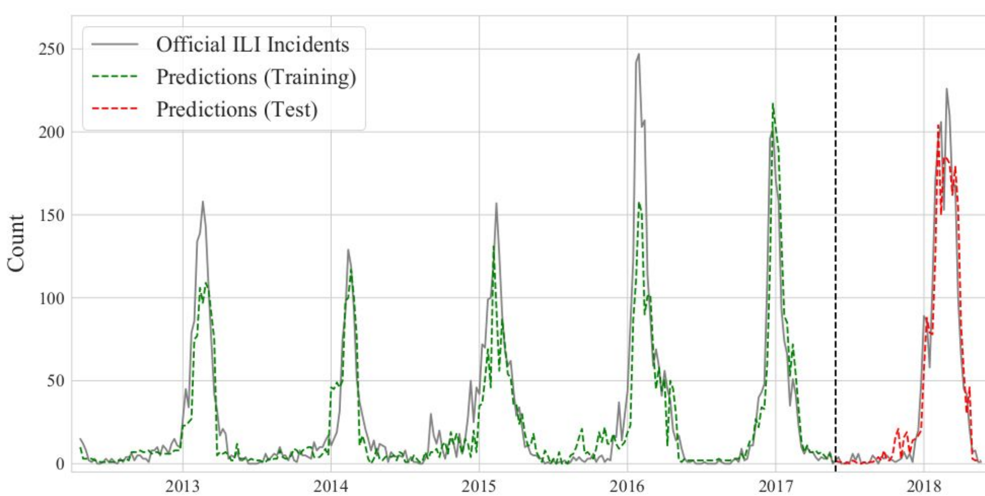
Algorithms Evaluated

NOWCASTING RESULTS ACHIEVED BY DIFFERENT MACHINE LEARNING MODELS AND CORRESPONDING INPUT FEATURES

Model	Feature Extraction	Number of Features	MAE (10-fold CV)	MAE (test)	R^2 (test)	Pearson's Correlation (test)
1 - Elastic Net	date + count	10	27.10	31.83	0.609	0.805
2 - XGBoost	date	3	28.28	29.33	0.552	0.763
3 - XGBoost	count	7	23.55	26.64	0.710	0.910
4 - SVM	date + count	10	18.72	24.30	0.657	0.826
5 - AdaBoost	date + count	10	18.11	18.35	0.781	0.915
6 - Random Forest	date + count	10	17.04	17.61	0.861	0.938
7 - LASSO	date + count	10	21.94	16.20	0.877	0.949
8 - Ridge Regression	date + count	10	22.12	14.75	0.892	0.942
9 - Linear Regression	date + count	10	22.55	14.66	0.886	0.943
10 - kNN Regression	date + count	10	18.11	14.00	0.895	0.948
11 - XGBoost	date + count	10	15.67	13.83	0.897	0.954
12 - Deep ConvNet + XGBoost	**date + count + image**	**14**	**13.14**	**11.33**	**0.925**	**0.963**

From: "Predicting the Flu from Instagram" by Gencoglu and Ermes

Algorithm Predictions



From: "Predicting the Flu from Instagram" by Gencoglu and Ermes

Summary and Q&A

- Talked about what Machine Learning is
- Reviewed some of the key problems machine learning solves
- Reviewed some of the key algorithms
- Looked at an example
- Any Questions?