DJS Synapse Learning Period
Week 2

Instructions:

1. Try to go through the resource list below before working on the tasks.
2. Make sure you try to understand the algorithm. A complete mathematical understanding is not necessary; an intuition will do.
3. You are given 2 datasets,
    Housing_data.csv: For Task 1 - linear regression
    Classified_data.txt: For Task 2 - kNN

Resources:

1. Train Test Split - Sklearn: https://www.youtube.com/watch?v=BUkqYGPnLZ8
2. Regression Metrics: https://www.youtube.com/watch?v=Ti7c-Hz7GSM
3. Understanding Boxplots:
https://medium.com/analytics-vidhya/introduction-to-box-plots-and-how-to-interpret-them-22464acbcba7
4. Comparing two boxplots: https://www.nagwa.com/en/explainers/812192146073/
5. Classical ML Models:

Linear regression

Theory : https://youtu.be/E5RjzSK0fvY
    https://www.youtube.com/watch?v=1-OGRohmH2s
Implementation : https://youtu.be/b0L47BeklTE

Logistic regression

Theory and implementation : https://youtu.be/VCJdg7YBbAQ
Theory: https://www.analyticsvidhya.com/blog/2021/07/an-introduction-to-logistic-regression

KNN

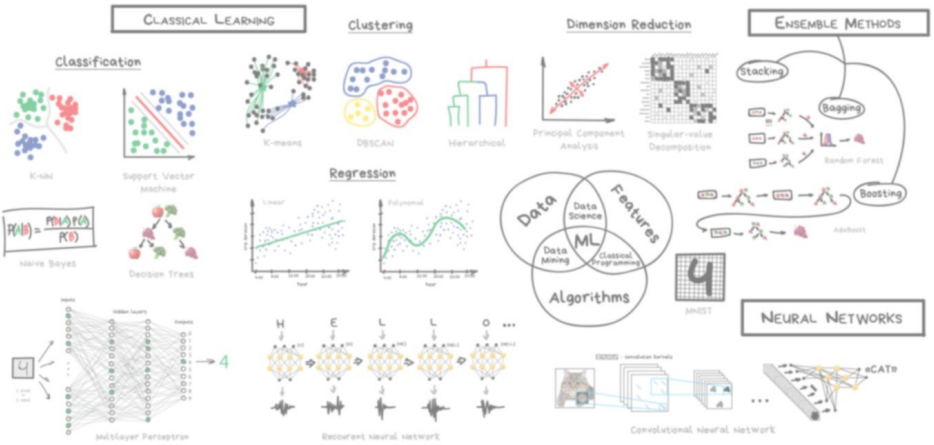
Theory and implementation : https://youtu.be/wTF6vzS9fy4
Theory: https://www.analyticsvidhya.com/blog/2018/08/k-nearest-neighbor-introduction-regression-python/

Overfitting and Underfitting [VERY IMPORTANT]

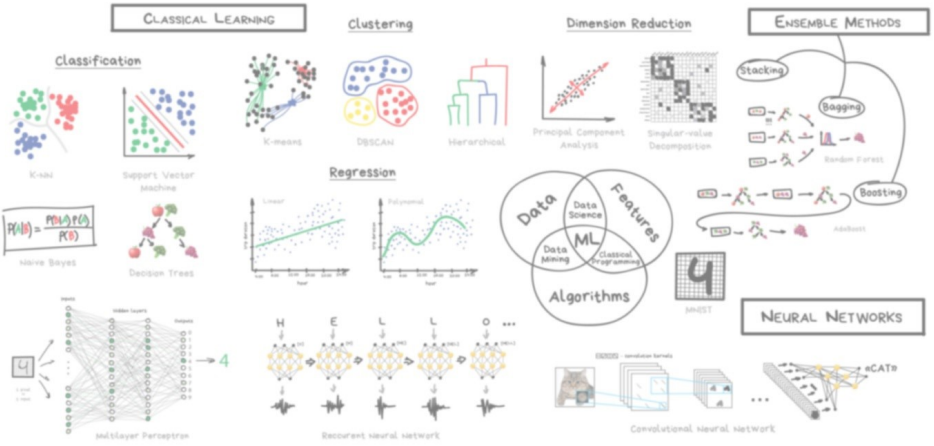
https://www.youtube.com/watch?v=T9NtOa-IITo

SVM

Theory : https://youtu.be/H9yACitf-KM
Theory: https://www.analyticsvidhya.com/blog/2021/06/support-vector-machine-better-understanding/
Implementation : https://youtu.be/FB5EdxAGxQg



Machine Learning for Everyone

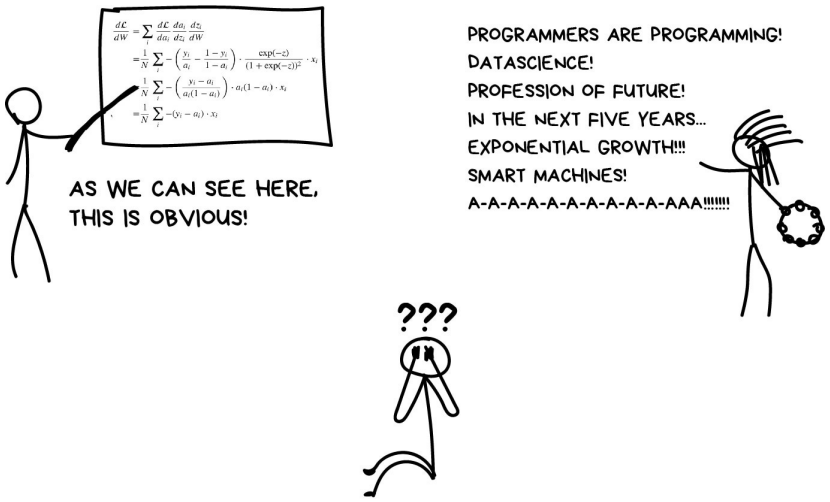


Machine Learning is like sex in high school. Everyone is talking about it, a few know what to do, and only your teacher is doing it. If you ever tried to read articles about machine learning on the Internet,

most likely you stumbled upon two types of them: thick academic trilogies filled with theorems (I couldn't even get through half of one) or fishy fairytales about *artificial intelligence*, *data-science magic*, and *jobs of the future*.

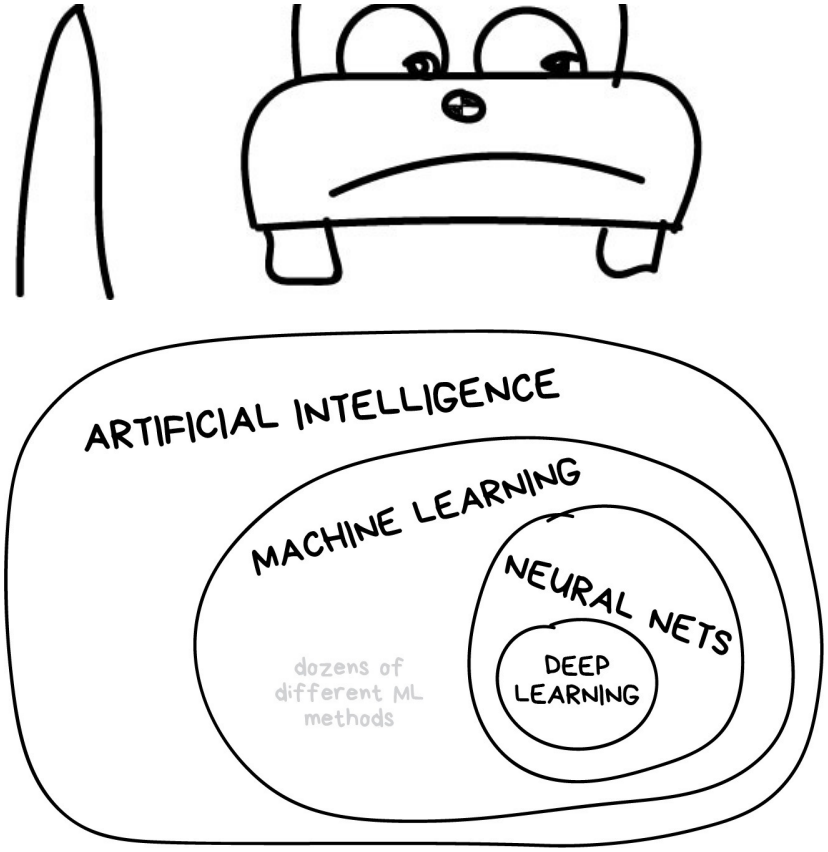
I decided to write a post I've been wishing existed for a long time. A simple introduction for those who always wanted to understand machine learning. Only real-world problems, practical solutions, simple language, and no high-level theorems. One and for everyone. Whether you are a programmer or a manager.

Let's roll.



TWO TYPES OF ARTICLES ABOUT MACHINE LEARNING

add a comment here



The map of the machine learning world

Part 1. Classical Machine Learning

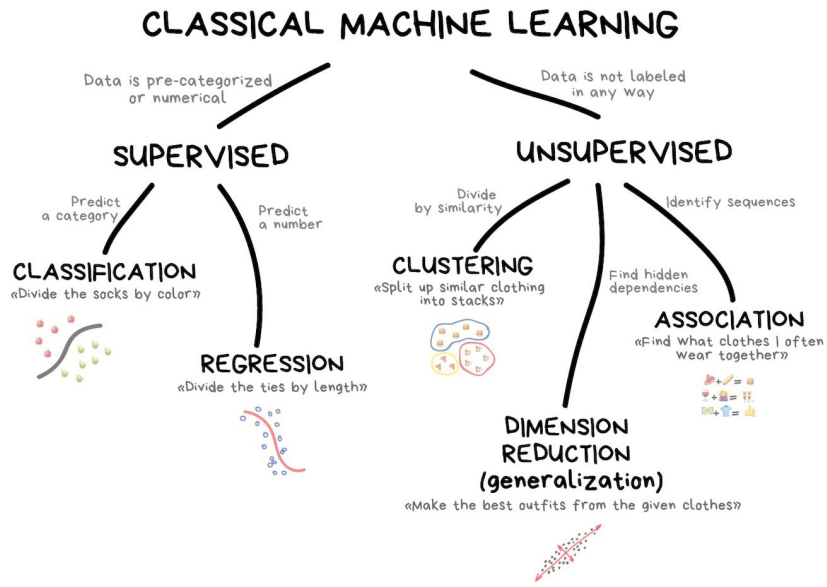
The first methods came from pure statistics in the '50s. They solved formal math tasks — searching for patterns in numbers, evaluating the proximity of data points, and calculating vectors' directions.

Nowadays, half of the Internet is working on these algorithms. When you see a list of articles to "read next" or your bank blocks your card

at random gas station in the middle of nowhere, most likely it's the work of one of those little guys.

Big tech companies are huge fans of neural networks. Obviously. For them, 2% accuracy is an additional 2 billion in revenue. But when you are small, it doesn't make sense. I heard stories of the teams spending a year on a new recommendation algorithm for their e-commerce website, before discovering that 99% of traffic came from search engines. Their algorithms were useless. Most users didn't even open the main page.

Despite the popularity, classical approaches are so natural that you could easily explain them to a toddler. They are like basic arithmetic — we use it every day, without even thinking.



[collapse all] [show all] **2** comments

1.1 Supervised Learning

Classical machine learning is often divided into two categories – **Supervised** and **Unsupervised Learning**.

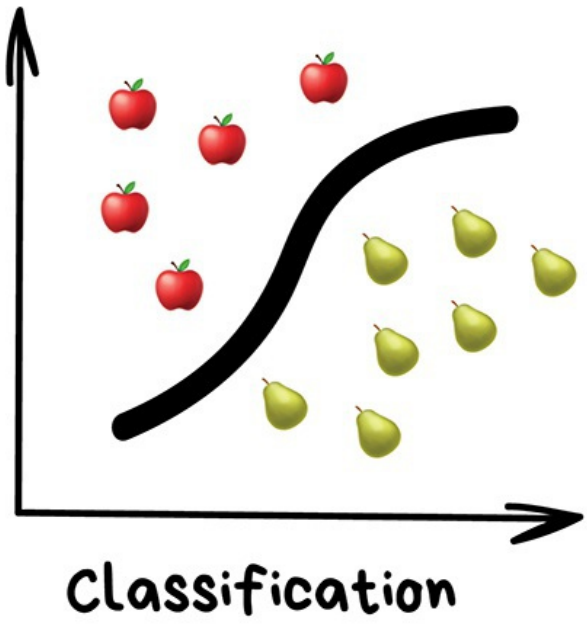
In the first case, the machine has a "supervisor" or a "teacher" who gives the machine all the answers, like whether it's a cat in the picture or a dog. The teacher has already divided (labeled) the data into cats and dogs, and the machine is using these examples to learn. One by one. Dog by cat.

Unsupervised learning means the machine is left on its own with a pile of animal photos and a task to find out who's who. Data is not labeled, there's no teacher, the machine is trying to find any patterns on its own. We'll talk about these methods below.

Clearly, the machine will learn faster with a teacher, so it's more commonly used in real-life tasks. There are two types of such tasks: **classification – an object's category prediction, and regression – prediction of a specific point on a numeric axis**.

add a comment here

Classification



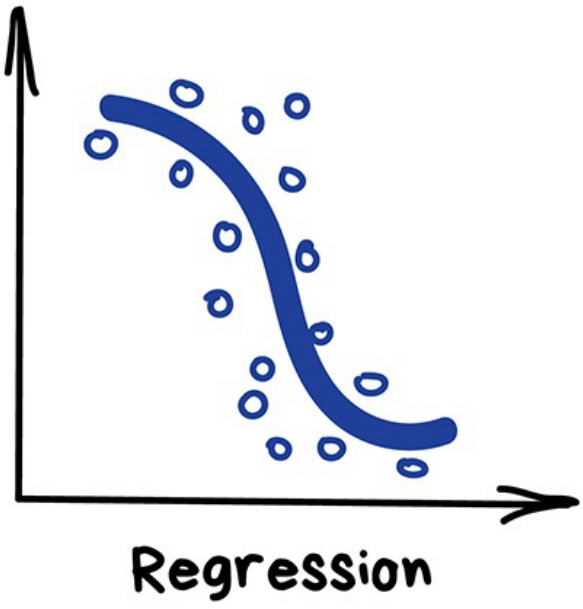
"Splits objects based at one of the attributes known beforehand. Separate socks by based on color, documents based on language, music by genre"

Today used for:
– Spam filtering
– Language detection
– A search of similar documents
– Sentiment analysis
– Recognition of handwritten characters and numbers
– Fraud detection

Popular algorithms: Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbours, Support Vector Machine

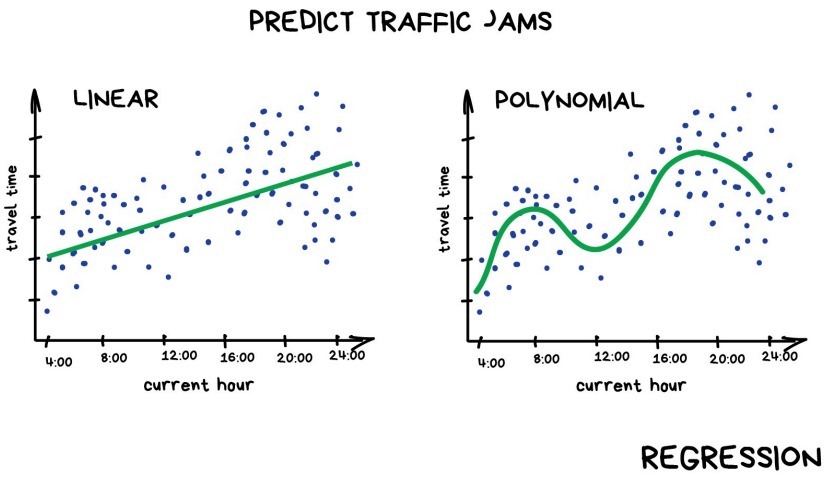
From here onward you can comment with additional information for these sections. Feel free to write your examples of tasks. Everything is written here based on my own subjective experience.

add a comment here



Regression is basically classification where we forecast a number instead of category. Examples are car price by its mileage, traffic by time of the day, demand volume by growth of the company etc. Regression is perfect when something depends on time.

Everyone who works with finance and analysis loves regression. It's even built-in to Excel. And it's super smooth inside — the machine simply tries to draw a line that indicates average correlation. Though, unlike a person with a pen and a whiteboard, machine does so with mathematical accuracy, calculating the average interval to every dot.



When the line is straight — it's a linear regression, when it's curved – polynomial. These are two major types of regression. The other ones are more exotic. Logistic regression is a black sheep in the flock. Don't let it trick you, as it's a classification method, not regression.

It's okay to mess with regression and classification, though. Many classifiers turn into regression after some tuning. We can not only define the class of the object but memorize how close it is. Here comes a regression.

If you want to get deeper into this, check these series: Machine Learning for Humans. I really love and recommend it!

add a comment here