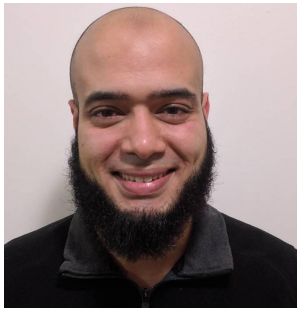
Machine Learning

What are the Features in ML

Mostafa S. Ibrahim
Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher
PhD from Simon Fraser University - Canada
Bachelor / MSc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)



Housing Price: The Human

- You are an expert in estimating a price for a house in your country
- Your friend asked you: “What is the price of a 3 bedroom apartment of 180 Meter”?
- What are the possible answers from an expert?

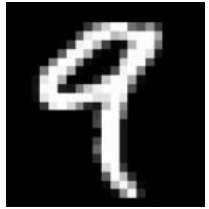
- You only provided 2 factors (aka **features**). We need more information such as the **location**, age of the building, condition (fully renovated?), facilities, school district, etc
    - Clearly an apartment in Boulak El Dakrour is way cheaper than Sheikh Zayed!

Housing Price: The Machine

- Which features are better for a machine learning algorithm?
  - 1) Number of bedrooms
  - 2) Location, number of bedrooms, facilities
  - 3) Location, number of bedrooms, facilities, age of the building, condition
  - 4) Location, number of bedrooms, facilities, age of the building, condition, color of each room

- **Location**, Number of bedrooms, facilities, age of the building, condition
  - Color is not added value
- Tip:
  - Features are very critical
  - If you missed a critical feature, it may harm the performance
  - If you added a useless feature, it might mislead the performance

Digit Recognition Features



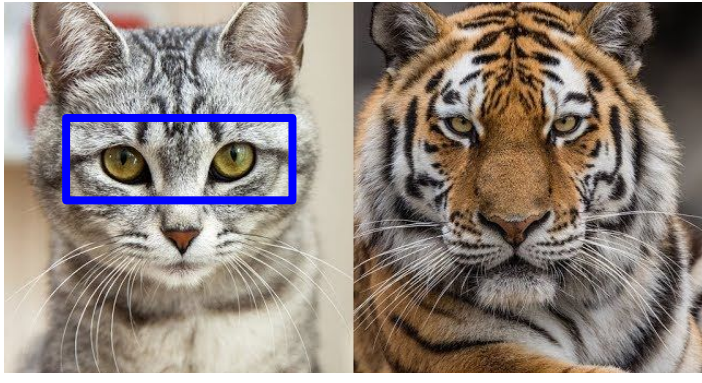
- Assume we have dataset of images
  - Each image has size 28x28 represeenting a digit from 0 to 9
  - Each pixel either white or black
- Assume we want to learn how to classify the image to a digit
- What are the features?!

- The image is small, we can use its pixels as features
- So a boolean vector of length 784 pixels is an input
  - 784 = 28 * 28

Animal Features

- What are the factors human use to differentiate a cat from a tiger?
- What are good features for a machine learning algorithm?

- For human
  - Our brain focuses on **specific features** such as eyes, nose and texture
- For machine learning
  - This is the job of computer vision field
  - **Classical machine learning**:
    - Let's find way to extract eyes, nose and texture
      - We call that **features extraction**
  - **Deep Learning**
    - Just give me the whole image pixels!
      - That is why deep learning wins



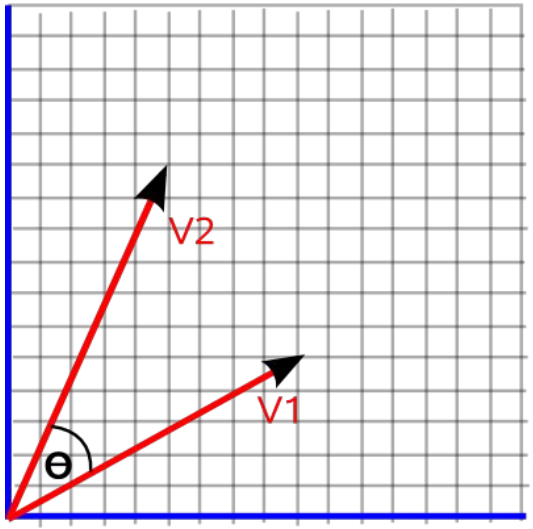
Features

- Features are input variables describing our data
  - Like the location of a home, words of a text, etc
- Feature vector: the features list
- X = {x1, x2, x3, …., xd} where d is the number of input features
  - The values are in real numbers representation (e.g. 12, 0.5, 0, 1, etc)
- In practice, the **raw** features can come in any data type
  - For example: country is Egypt/US
    - Index the countries: Egypt = 0, US = 1, Germany = 2, etc

Feature Vector

- In practice, we just use a list or a numpy array to represent the features of interest. Typically we use **real**/numerical values
- We can visualize these elements as a vector in the feature space

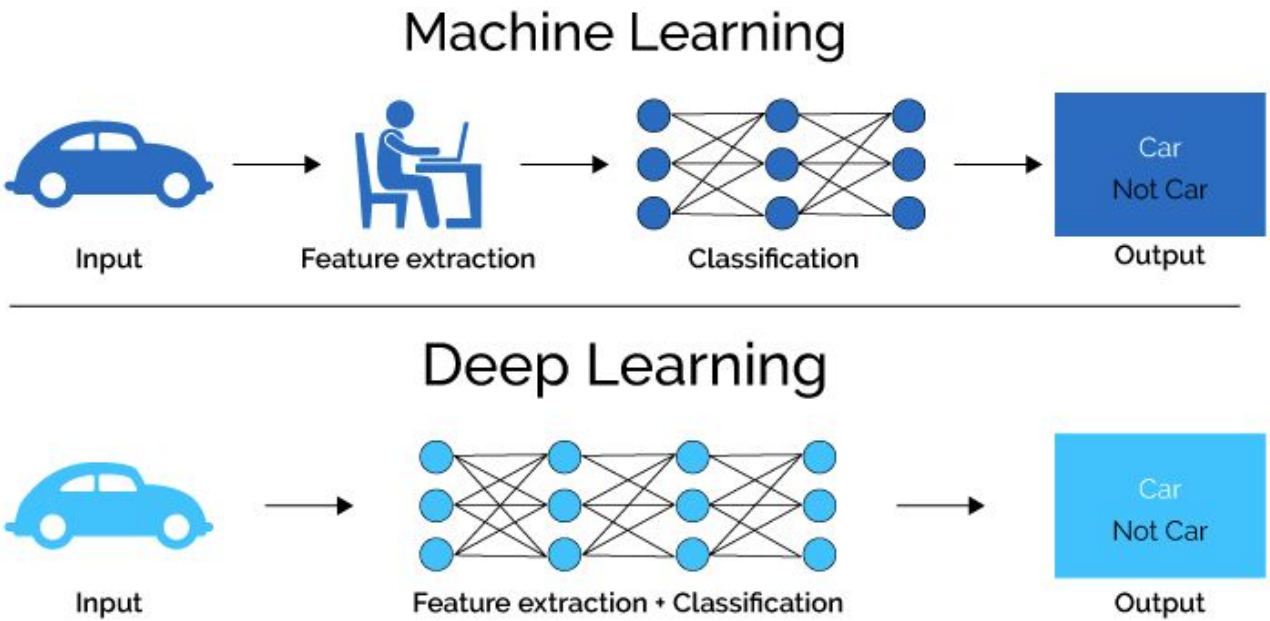
```
location = 11235     # Indexing the area
bedrooms = 3
age = 20             # 20 years old
has_air_conditioer = 1
has_balacony = 0

feature_vector = [location, bedrooms, age,
                  has_air_conditioer, has_balacony]
print(feature_vector)   # [11235, 3, 20, 1, 0]
```



Feature Extraction

- **Feature engineering/extraction** is the process of selecting, manipulating, and **transforming** raw data into useful features for helping the machine learning **finds more patterns** in the data
- How is it extracted?
  - **Classical ML**: **manual effort** to find interesting ways to extract such data
    - It requires intelligence and domain expertise
    - Special case: **Reducing** a large number of data to a small representative features list
  - **Deep Learning**: **automatically** let the algorithm find interesting features

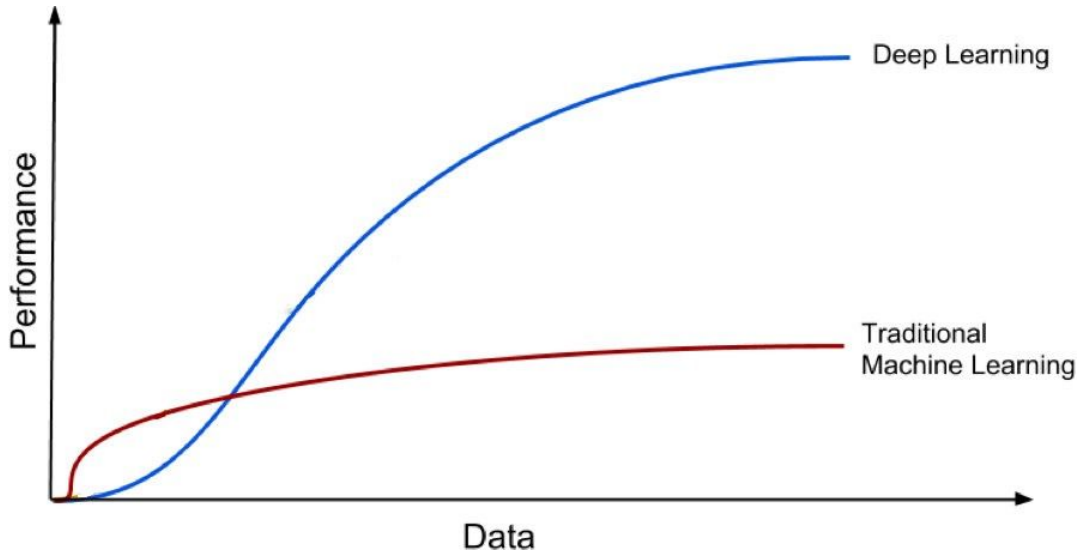
Classical Machine Learning vs Deep Learning



Img [src]

Classical Machine Learning vs Deep Learning

- If there are a lot of data, deep learning can extract strong features manually and boost the performance strongly



Img src

Facts

- Features never **fully describe** the input data!
  - Either extracted manually or automatically, features are an **approximation** for the data
  - In the best cases, most of the **sufficient features** will be available
- Regardless of how much algorithms continue to improve, feature engineering continues to be a difficult process that requires human **intelligence** with **domain expertise**.
- The **quality** of feature engineering often drives the **quality** of a machine learning model
- Great features represent **unique** characteristics that holds for **most** of the samples (not a few of them)

Question!

- Given a short video clip for volleyball game, classify the movement of the ball to either ‘**moving** from the **left** team to the **right** team’ or the opposite
  - A video is a set of frames (pictures)
- A deep learning classifier takes the **raw** video and learned the direction of the ball. How did the classifier learned that?!

Question!

- The algorithm just sees pixels (colors)
- It doesn’t understand people or the ball
  - Unless we did that in an explicit way
- The algorithm will learn the movement of the pixels (called flow)
  - Are most of the pixels moving to the right direction or left direction?!
- The moral of that:
  - Machine learning can learn complex patterns from the data that human even did not observe!
  - Sometimes, even ML can learn something descent from a buggy/noisy data
    - This is extremely hard to debug as you don’t notice something wrong!

Summary

- Features are **input variables** describing our data
- A **feature vector** is an ordered list of numerical properties representing an **input** that we feed to machine learning algorithm
- **Feature engineering** or (extraction/discovery) is the process of using **domain knowledge** to extract features (characteristics, properties, attributes) from raw data

"Acquire knowledge and impart it to the people."

"Seek knowledge from the Cradle to the Grave."