Intro to ML

The Hello World of Machine Learning

Who am I?

Who am I?

Backend and Feature Developer for

"Google Analytics for healthcare data"

by Ivbar

Machine Learning and Touch Rugby Enthusiast

christine.winter.42@gmail.com

github.com/christinewinter

Lightning talks & study group

Wednesday, June 19, 2019

5:30 PM to 8:30 PM

Foo Café (Hammarby kaj 10D · Stockholm)

www.meetup.com/PyLadiesStockholm/events

Who is listening?

Who knows what machine learning is?

Who programmed in python before?

What is Learning ?

What is Learning ?

The acquisition of knowledge or skills through study, experience, or being taught.

Oxford dictionary

What is Learning ?

The acquisition of knowledge or skills through study, experience, or being taught.

Oxford dictionary

Understand what's going on with different methods and use this information the next time.

My interpretation

How do we learn ?

How do we learn ?

Get instructions from others

How do we learn ?

Get instructions from others

Interpret a situation and try to define rules or draw conclusions alone

How do we learn ?

Get instructions from others

Interpret a situation and try to define rules or draw conclusions alone

Learn from previous mistakes or success

How do Machines learn ?

How do Machines learn ?

Get instructions from others

Interpret a situation and try to define rules or draw conclusions alone

Learn from previous mistakes or success

How do Machines learn ?

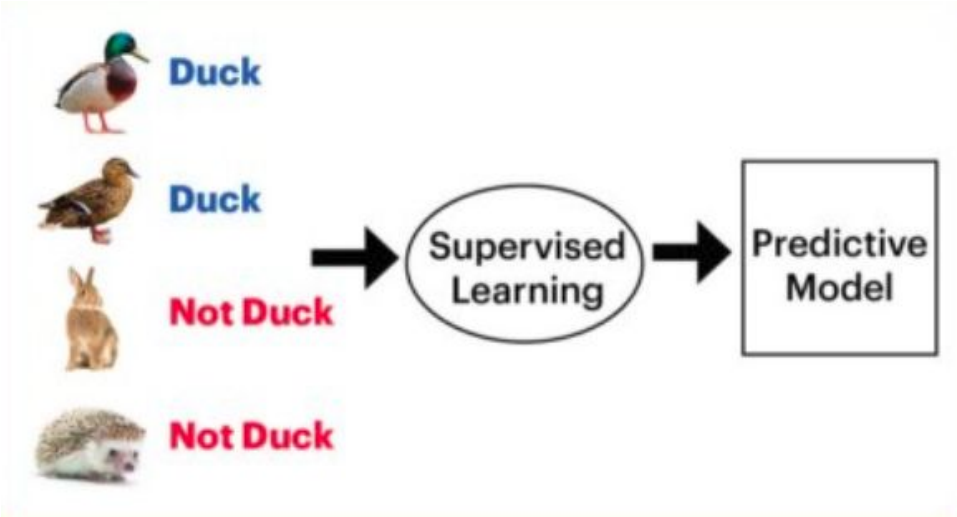
Get instructions from others

→ **supervised learning**

Interpret a situation and try to define rules or draw conclusions alone

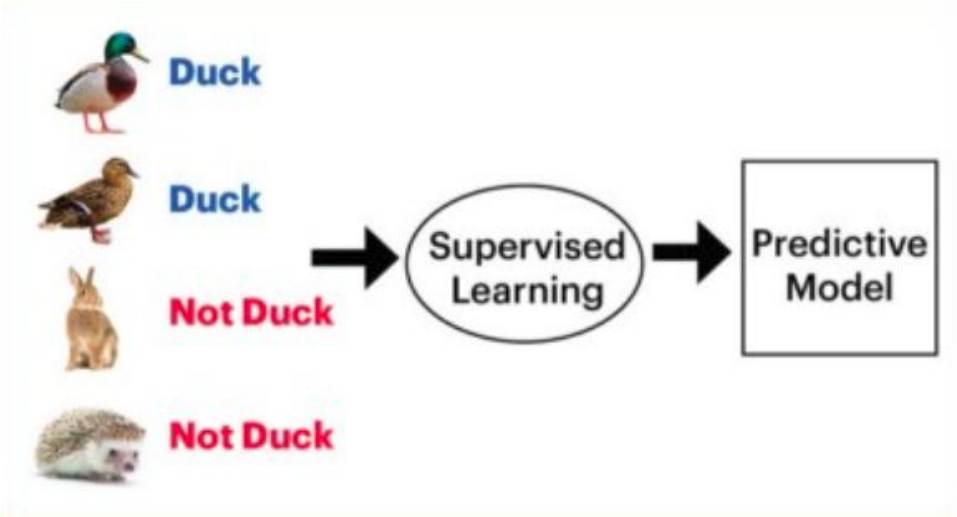
Learn from previous mistakes or success

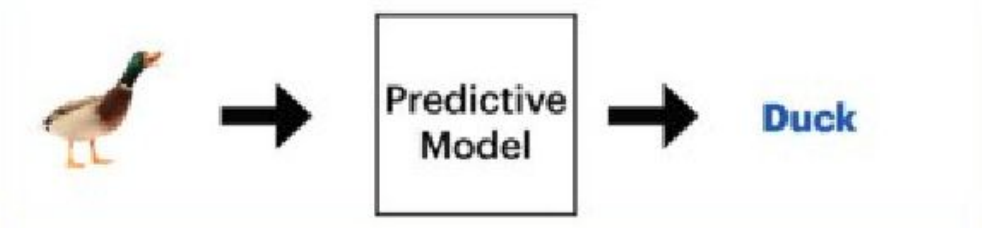
Supervised



https://medium.com/datadriveninvestor/differences-between-ai-and-machine-learning-and-why-it-matters-1255b182fc6

Supervised





https://medium.com/datadriveninvestor/differences-between-ai-and-machine-learning-and-why-it-matters-1255b182fc6

How do Machines learn ?

Get instructions from others

→ **supervised learning**

Interpret a situation and try to define rules or draw conclusions alone

Learn from previous mistakes or success

How do Machines learn ?

Get instructions from others

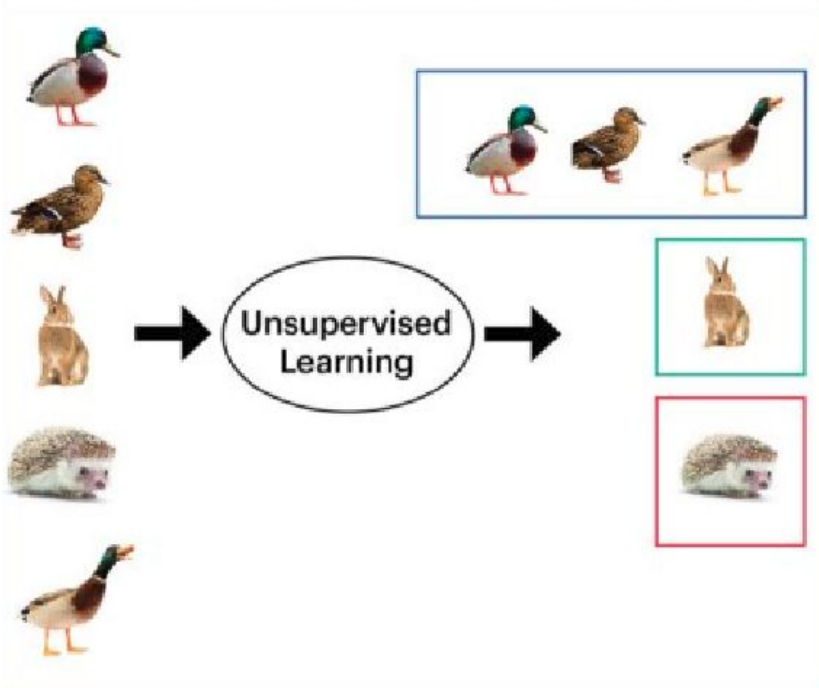
→ supervised learning

Interpret a situation and try to define rules or draw conclusions alone

→ **unsupervised learning**

Learn from previous mistakes or success

Unsupervised



https://medium.com/datadriveninvestor/differences-between-ai-and-machine-learning-and-why-it-matters-1255b182fc6

How do Machines learn ?

Get instructions from others

→ supervised learning

Interpret a situation and try to define rules or draw conclusions alone

→ **unsupervised learning**

Learn from previous mistakes or success

How do Machines learn ?

Get instructions from others

→ supervised learning

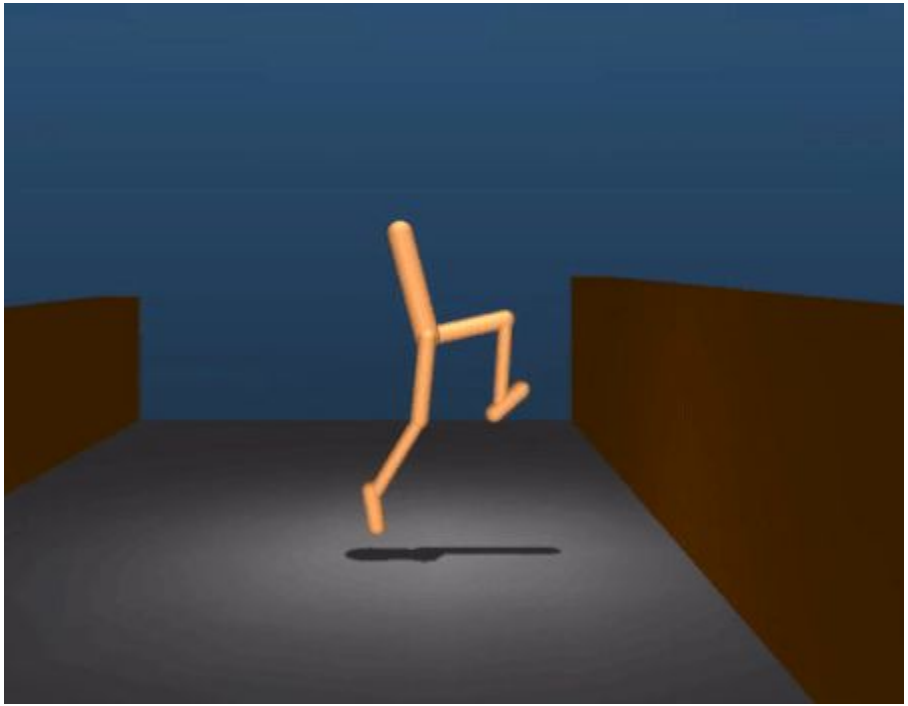
Interpret a situation and try to define rules or draw conclusions alone

→ unsupervised learning

Learn from previous mistakes or success

→ **reinforcement learning**

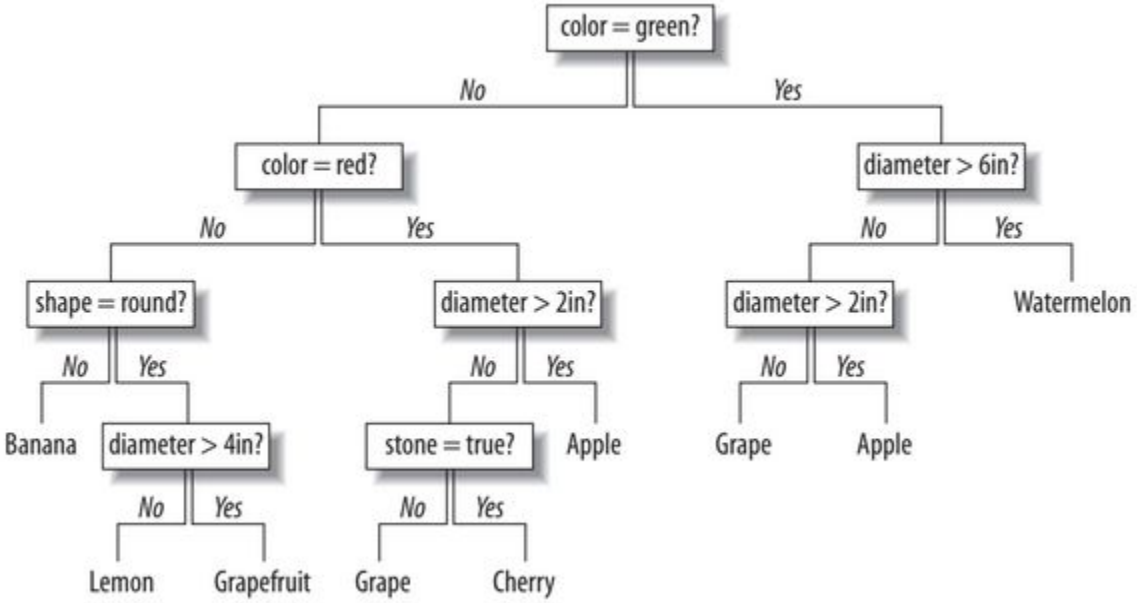
Reinforcement



https://www.theverge.com/tldr/2017/7/10/15946542/deepmind-parkour-agent-reinforcement-learning

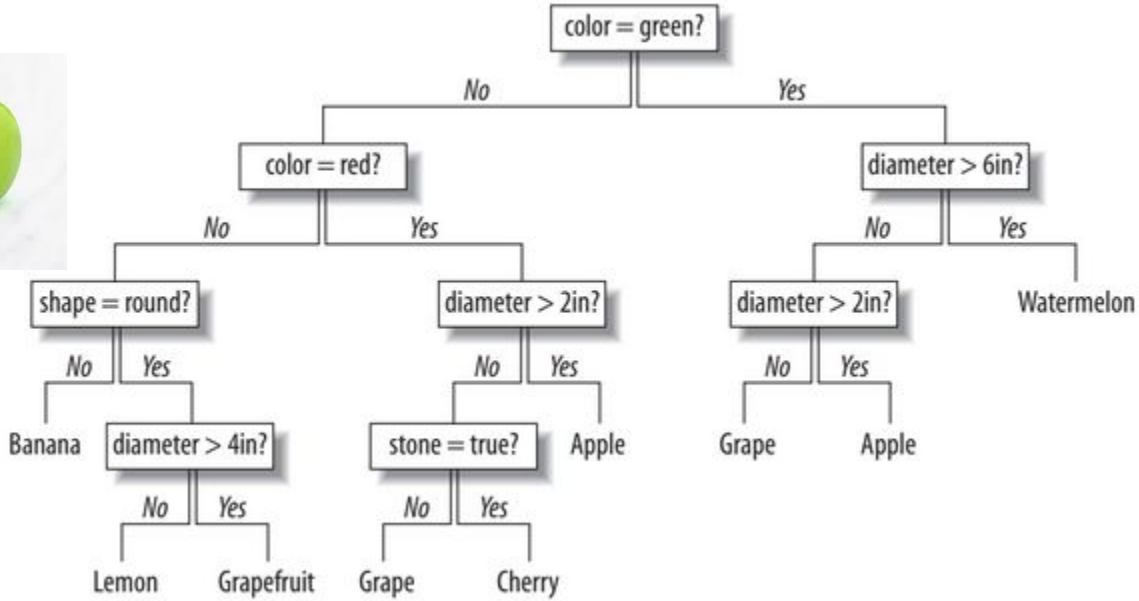
So, what will we do?

So, what will we do? - Learn with decision trees



https://github.com/ShawnLeee/the-book/blob/master/pybooks/Programming%20Collective%20Intelligence.pdf

So, what will we do? - Learn with decision trees



https://github.com/ShawnLeee/the-book/blob/master/pybooks/Programming%20Collective%20Intelligence.pdf

The Data

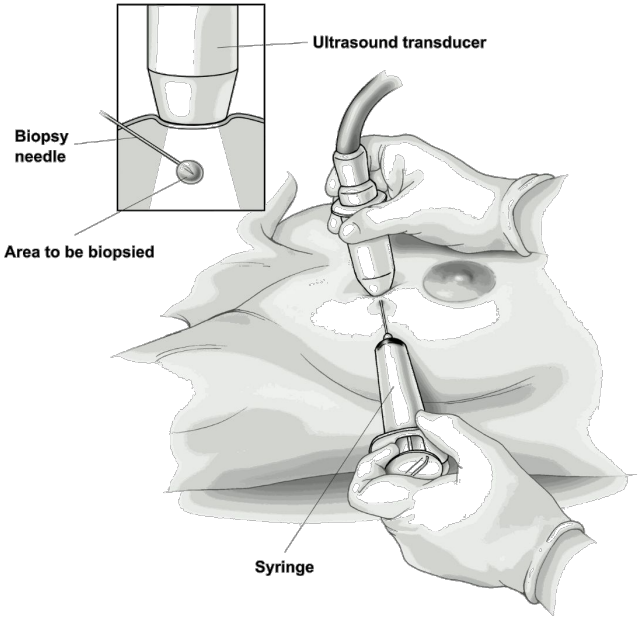
- Breast Cancer Wisconsin (Diagnostic) Data Set acquired by researchers University of Wisconsin (Dr. W.H. Wolberg, N. Street, O.L. Mangasarian)

- Features computed from a digitized image of a fine needle aspirate (FNA) of a breast mass

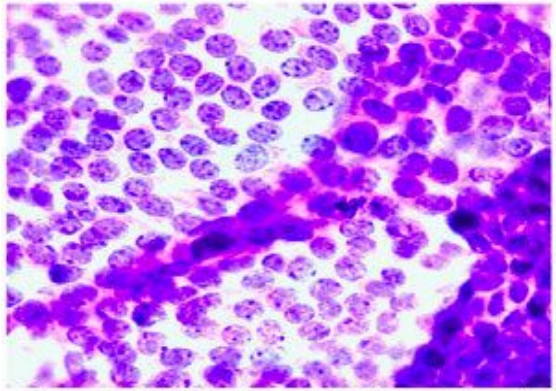
- Describe characteristics of cell nuclei present in the image

https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)
https://www.kaggle.com/uciml/breast-cancer-wisconsin-data

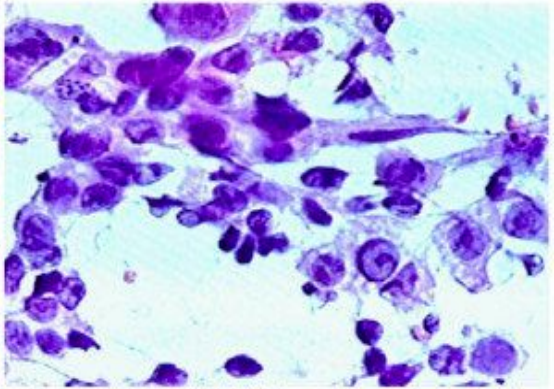
The Data



Fine needle aspiration using ultrasound



Smear with BENIGN diagnosis – uniform nucleus of cells, symmetrical, homogeneous, with areas within normal size



Smear with MALIGNANT diagnosis – nucleus of cells without uniformity, asymmetrical, not homogeneous (multiple sizes) and with areas above normal size

https://www.researchgate.net/figure/Captured-images-of-layers-of-glass-with-smears-of-breast-mass-obtained-by-FNA-the-parts_fig3_232811011
https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/breast-biopsy/fine-needle-aspiration-biopsy-of-the-breast.html

Lets Go !

Setting up

Local: clone repo, install requirements.txt

https://github.com/christinewinter/intro_ml_bc

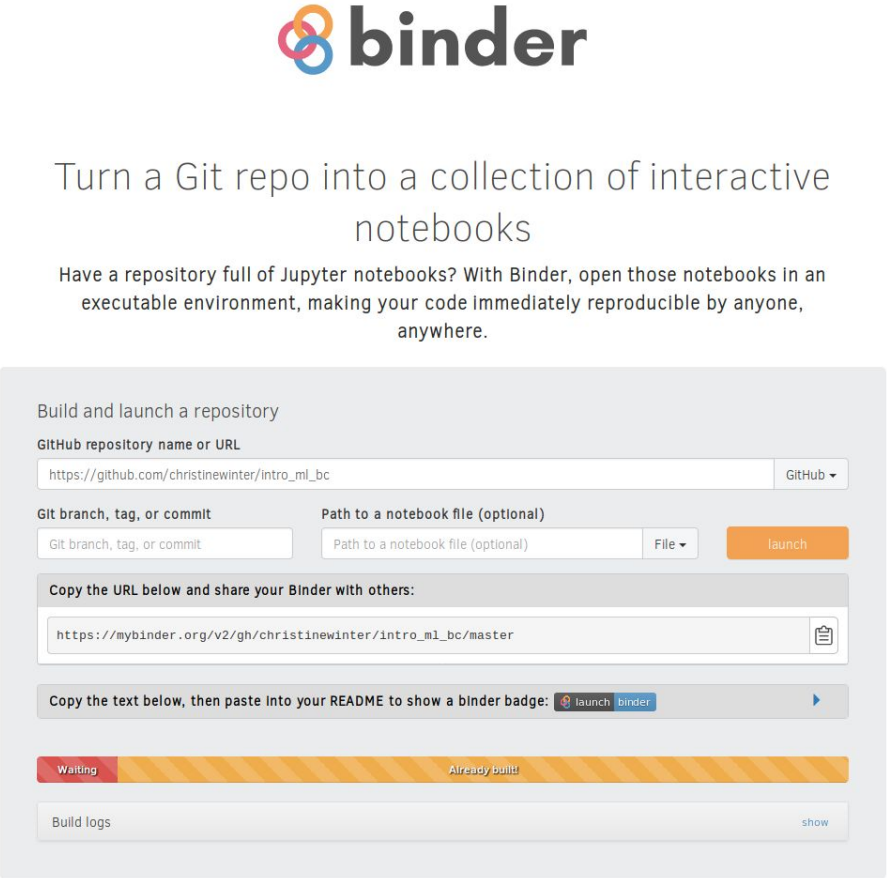
Remote:

Go to: https://mybinder.org/

Insert:

https://github.com/christinewinter/intro_ml_bc

Will take ~ 5 minutes

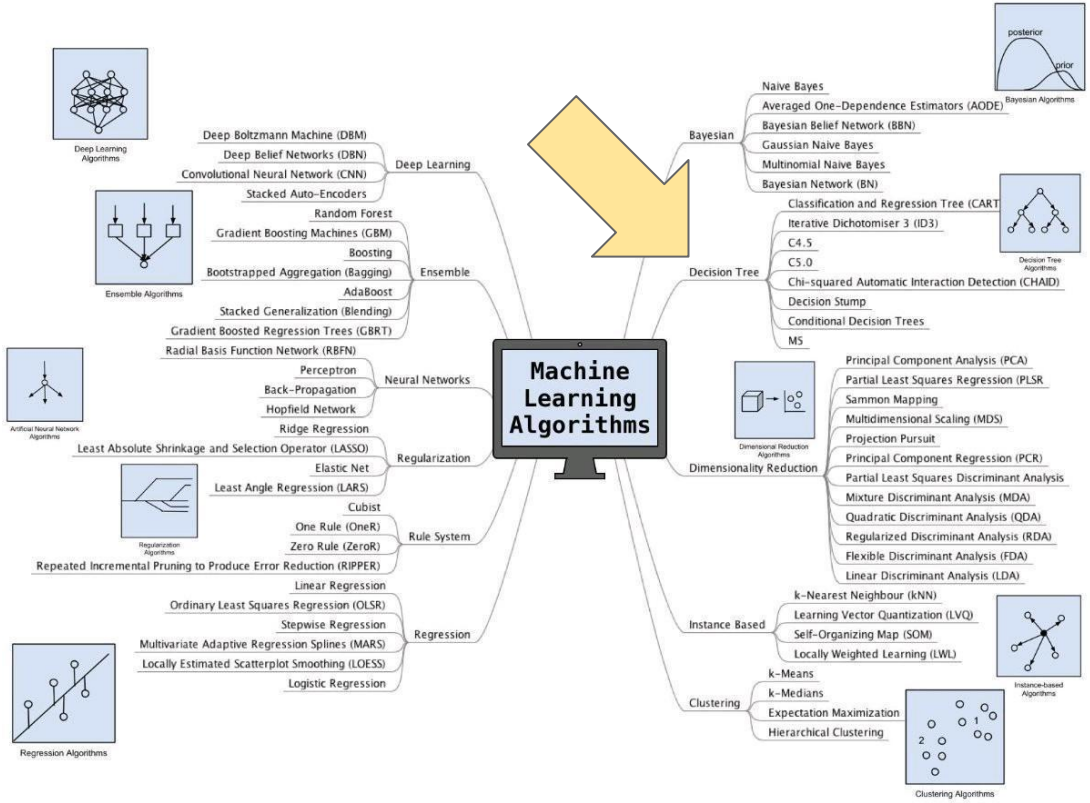


The actual machine Learning part

```
model = DecisionTreeClassifier()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
metrics.accuracy_score(y_test, y_pred)
y_pred = model.predict(X_unknown)
```

scikit-learn.org

Decision trees are just one way to learn



What is important for ML ?

What is important for ML ?

- ★ Data driven
  - Density, diversity, structure

What is important for ML ?

- ★ Data driven
  - Density, diversity, structure
- ★ Problem definition
  - Supervised, unsupervised, reinforcement learning

What is important for ML ?

- ★ Data driven
  - Density, diversity, structure
- ★ Problem definition
  - Supervised, unsupervised, reinforcement learning
- ★ Model choice
  - Decision tree, deep neural networks, support vector machines...

What is important for ML ?

- ★ Data driven
  - Density, diversity, structure
- ★ Problem definition
  - Supervised, unsupervised, reinforcement learning
- ★ Model choice
  - Decision tree, deep neural networks, support vector machines...
- ★ Metrics to evaluate the model
  - Prediction score, prevention of overfitting

What is important for ML ?

- ★ Data driven
  - Density, diversity, structure
- ★ Problem definition
  - Supervised, unsupervised, reinforcement learning
- ★ Model choice
  - Decision tree, deep neural networks, support vector machines...
- ★ Metrics to evaluate the model
  - Prediction score, prevention of overfitting
- ★ Feature selection
  - Relevance, correlations

What is important for ML ?

- ★ Data driven
  - Density, diversity, structure
- ★ Problem definition
  - Supervised, unsupervised, reinforcement learning
- ★ Model choice
  - Decision tree, deep neural networks, support vector machines...
- ★ Metrics to evaluate the model
  - Prediction score, prevention of overfitting
- ★ Feature selection
  - Relevance, correlations
- ★ Iteration & architecture
  - Implementation, backup

What you learned today:

What you learned today:

- ★ Types of learning:
  - supervised, unsupervised, reinforcement

What you learned today:

- ★ Types of learning:
  - supervised, unsupervised, reinforcement
- ★ One machine learning model:
  - Decision trees

What you learned today:

- ★ Types of learning:
  - supervised, unsupervised, reinforcement
- ★ One machine learning model:
  - Decision trees
- ★ Data is the key component
  - Clean up is 80 % of the work

What you learned today:

- ★ Types of learning:
  - supervised, unsupervised, reinforcement
- ★ One machine learning model:
  - Decision trees
- ★ Data is the key component
  - Clean up is 80 % of the work
- ★ Implementation
  - Python, jupyter, Scikit learn

Buzzword Bingo :

Artificial Neural Network

Machine Learning - ML

Artificial Intelligence - AI

Cloud AI

Data Mining

Deep Learning - DL

DNN

ANN

CNN

AIOps - automated DevOps

BIG data

Data driven

Tensorflow

Keras

AI in IoT

Artificial Neocortex

Questions ?

Gini Impurity

Gini impurity is the expected error rate if one of the results from a set is randomly applied to one of the items in the set.

If every item in the set is in the same category, the guess will always be correct, so the error rate is 0. If there are four possible results evenly divided in the group, there's a 75 percent chance that the guess would be incorrect, so the error rate is 0.75.

This function calculates the probability of each possible outcome by dividing the number of times that outcome occurs by the total number of rows in the set. It then adds up the products of all these probabilities. This gives the overall chance that a row would be randomly assigned to the wrong outcome.

The higher this probability, the worse the split. A probability of zero is great because it tells you that everything is already in the right set.

https://github.com/ShawnLeee/the-book/blob/master/pybooks/Programming%20Collective%20Intelligence.pdf