Machine Learning Regression

@ Azure Advocates

Regression → mathematical methods that let us predict a continuous value

Linear Regression

$y = a + bx$

Polynomial Regression

Authoring environment + tools

- Python
- .ipynb Notebook
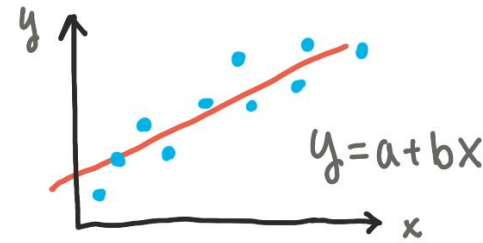- VSCode
- import data set

Scikit-Learn open-source ML library
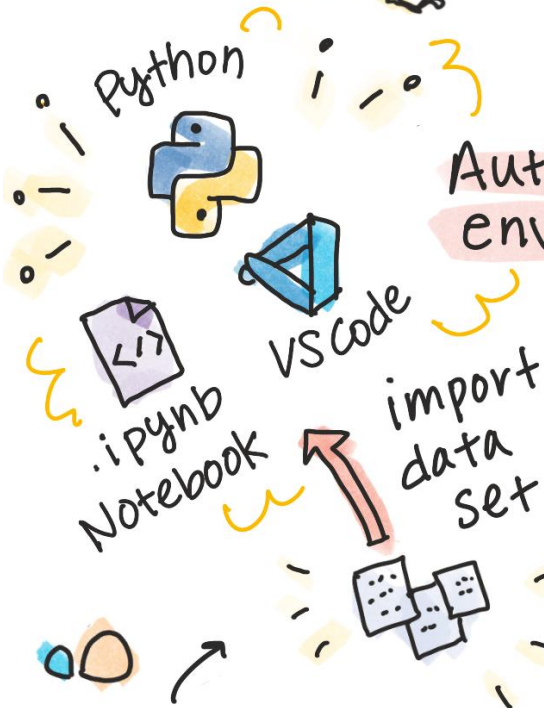
Libs
- ⭐ matplotlib
  - graphing tool
- ⭐ numpy
  - handling numeric data
- ⭐ pandas
  - analyzing + manipulating data

Read CSV data → CSV
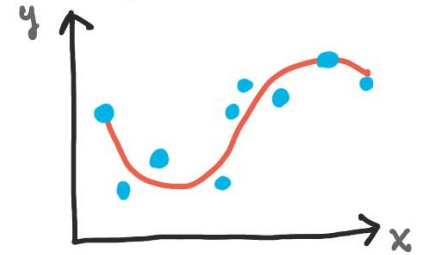
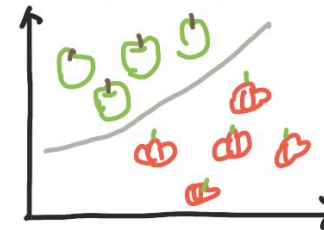## Logistic Regression

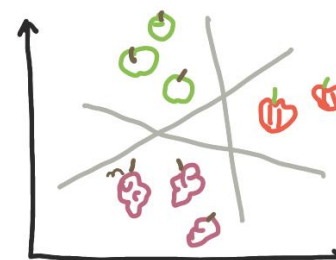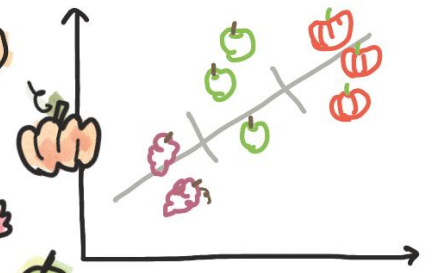: ♥ Binary Classification

: ♥ Multinomial Classification

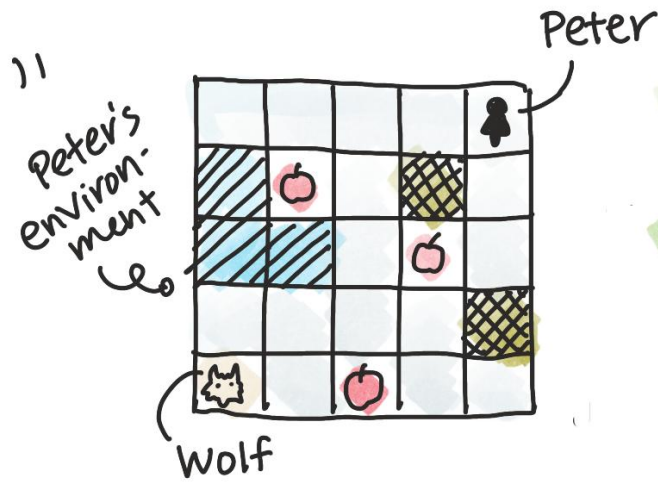: ♥ Ordinal Classification

ordered categories

Machine Learning
Reinforcement

Decision-making to take actions in order to maximize cumulative rewards

Peter's environment

Peter

- **Actions:**
  ⬆ ⬇ ⬅ ➡

- **Rewards:**
  +10 apple
  -10 wolf
  -5 move

Wolf

Peter: explores his environment + learns to avoid wolves and collect apples!

Q-Learning ← Algorithm to find the best way to achieve the goal!

? **Exploration** try experimenting on uncharted territory

! **Exploitation** known path

Find a good balance of both!

exploring...

⭐ Basic: **Q-Function**

$Q(\text{state}, \text{action}) \rightarrow \text{number}$

current state

At every position at the board, and every action you make, there would be some number

⭐ **Bellman Equation** ← Dynamic Programming

$$Q(s, a) = r(s, a) + \gamma \max_{a'} Q(s', a')$$

current state *action* *Reward* *discount factor* *estimate of optimal future value*
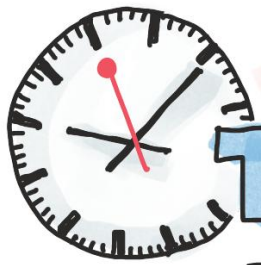
Machine Learning Time Series Forecasting

Looking at data overtime to forecast or predict future values based on patterns or recurring trends.

examples

- ☂️ Weather forecast
- 🎄 Holiday season sales
- 💲 Stock market
- online user traffic

I am predictable

Data Characteristics:
- ♡ Seasonality
- ♡ Outliers
- ♡ long-run cycle
- ♡ Constant variance
- ♡ abrupt changes

⭐ Using ARIMA — moving average

Auto regressive • integrated

Previous "lags" values in your data = long memory models
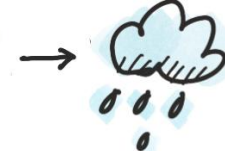
Stationarity

$y_t - y_{t-1}$

△ diff

Previous hidden shocks or errors in your model = short memory models

Differencing: look at the △ from one period to another

time

LIB Statmodels

Define model w/ SARIMAX()

@AzureAdvocates @girlie_mac

Fairness in Machine Learning

Training Set → input → BLACK BOX → output → Reject all raccoons!

Oh no!

Bias — Cats are great!

Bias — Raccoons are bad!
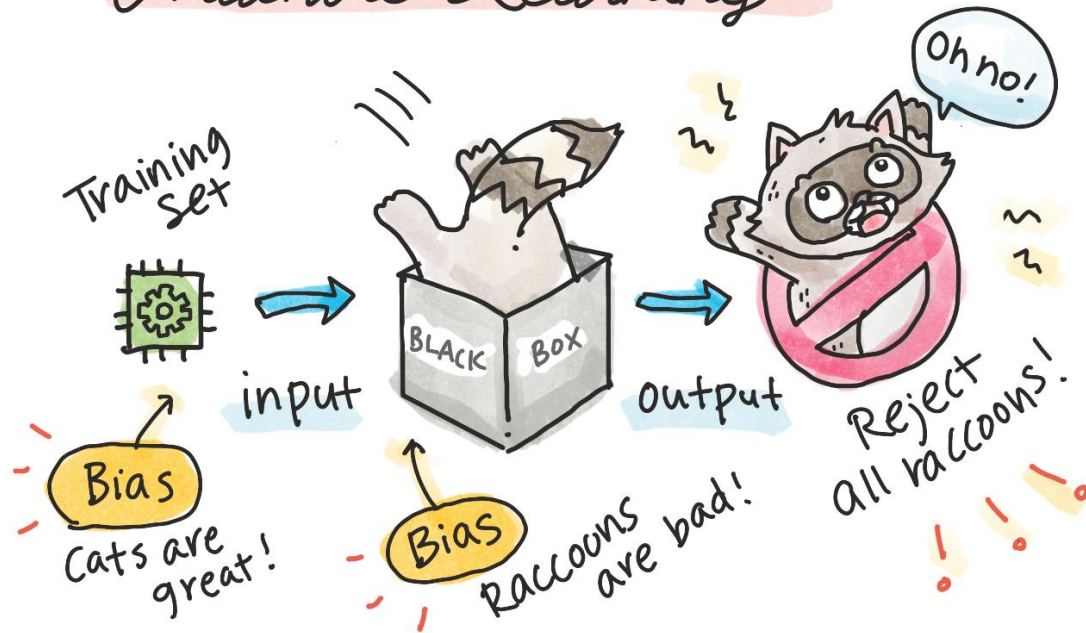
Fairness-related harms

Unfairness = negative impacts for group of people such as those defined in terms of
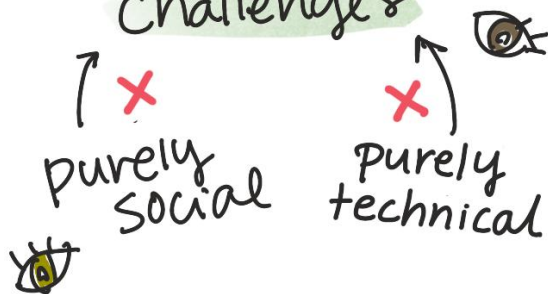- race
- gender
- age
- disability status

Harms:
- ☆ Allocation
- ☆ Quality of service
- ☆ Stereotyping
- ☆ Denigration
- ☆ Over- / under- representation

CEO

Complex sociotechnical challenges

✗ purely social

✗ purely technical

Assessment & mitigation

False negatives

False positives

- ♡ Identify the harm (+ benefits)
- ♡ Identify the affected groups
- ♡ Define fairness metrics

Fairlearn
fairlearn.github.io

	False-	False+	Counts
men	0.35	0.27	6239
women	0.29	0.35	3124

Machine Learning
in the real world

Finance
- Credit card fraud detection
  ↳ K-means clustering
- Wealth management
  ↳ Linear regression

Education
- Predicting student behavior → Coursera
- mitigating bias → Grammarly
  ↳ NLP

Retail
- personalizing the customer journey ← NLP
- inventory management → Stitch Fix

Healthcare
- Managing clinical trails
  ↳ Classifier
- Hospital readmission management
  ↳ Clustering
- Disease management
  ARIMA
  logistic curves
  linear regression
  SARIMA

Ecology & Green tech
- Forest management
  ↳ Reinforcement
- Motion sensing of animals
- Energy management
  ↳ Time series forecasting

Insurance
- Volatility management → MetLife
  ↑ ordinal + binary classification
  Forecasting

Arts, Culture, and Literature
- Fake news detection ↳ Classifiers
- Musium ML
  → Vatican archive
  → Art Institute of Chicago

Marketing
- Customer Segmentation
  ↳ Clustering