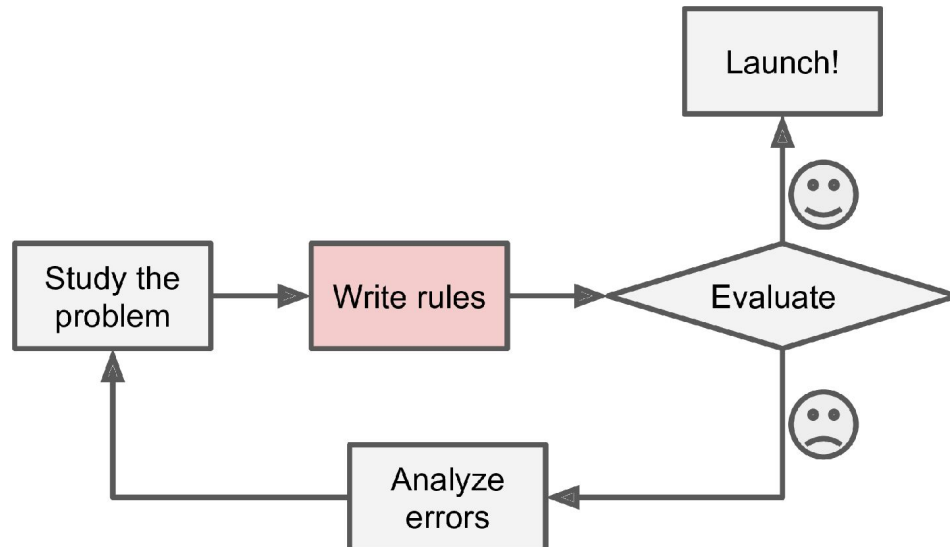


1. Why Machine Learning?

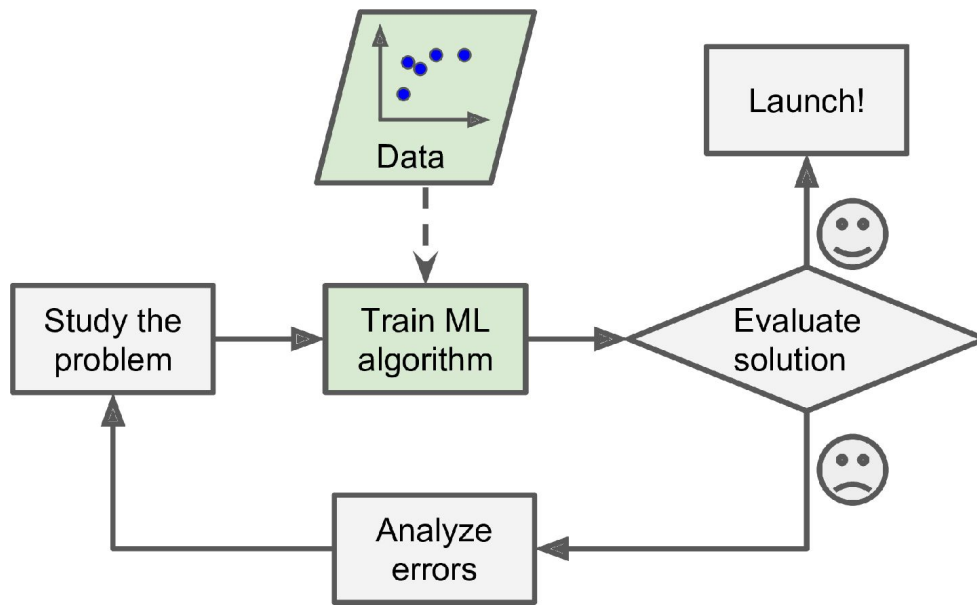
a. Traditional approach:

- i. Find out rules and detect patterns
- ii. Specify the rules in the program (complex)
- iii. Evaluate results based on this program and the rules embedded
- iv. Hard to maintain
- v. Repeats i. & ii. until program is good enough



b. Machine Learning approach:

- i. Algorithm learns the patterns by itself
- ii. Training set & testing set
- iii. Program is much shorter and easier to maintain
- iv. More accurate



2. Ex: Spam Email Filter - Long list of rules looking for words such as “Credit Card,” “Free,” and “4U” would need to be hard coded. With ML, however, the algorithm can continuously learn about new words and detect patterns by itself.

3. Summary:

- a. One ML algorithm can simplify code and perform better
- b. ML algorithms can find out solutions for complex and unsolvable problems
- c. Flexible to adapt to new data
- d. Insights and analysis for large and complex data

4. Types of Machine Learning Systems

a. Supervised Learning

- i. Training data fed to the algorithm includes the desired solutions, called *labels*
- ii. *Classification* tasks (e.g. class = spam or ham)
- iii. *Predictors*: Predict *target* numeric value through inspecting *features* (e.g. price of a car given a set of features such as mileage, age, brand, etc.)
- iv. Most important Supervised Learning Algorithms (will cover most in this club):
 1. k-Nearest Neighbors
 2. Linear Regression
 3. Logistic Regression
 4. Support Vector Machines
 5. Decision Trees and Random Forests
 6. Neural Networks (could also be unsupervised; many variations)

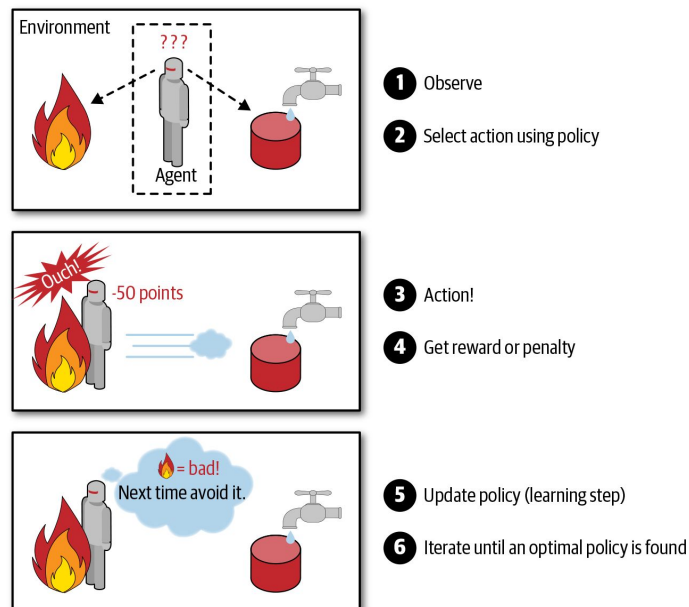
b. Unsupervised Learning

- i. Training data is NOT labeled; the system learns without a teacher
- ii. Most important Unsupervised Learning Algorithms:
 1. Clustering: detect groups with high similarities

- a. k-Means
- b. Hierarchical Cluster Analysis
- c. Expectation Maximization
- 2. Visualization and Dimensionality Reduction: outputs 2D/3D representation of data; simplify data without losing too much information (feature extraction)
 - a. Principal Component Analysis
 - b. Kernel PCA
- 3. Association Rule Learning
 - a. Apriori
 - b. Eclat

c. Reinforcement Learning

- i. Learning system observes the environment, select and perform actions, and get rewards in return (or penalties for negative rewards)



ii.

iii. DeepMind's AlphaGo

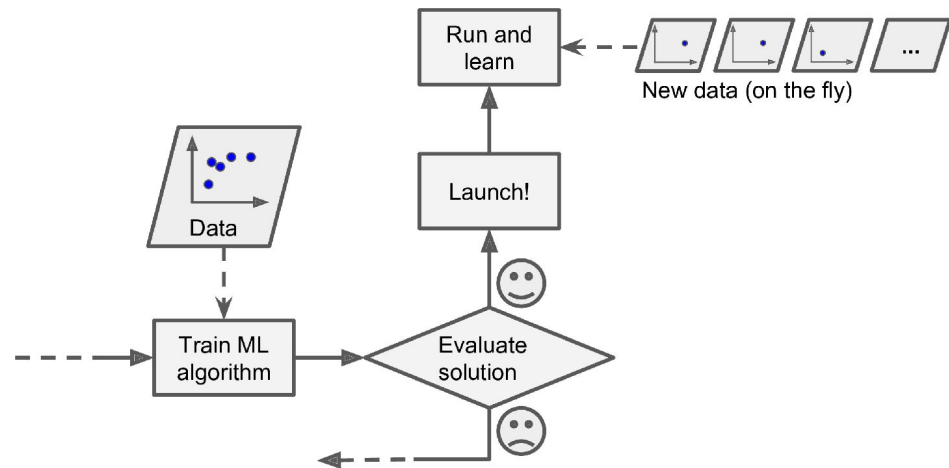
5. Batch & Online Learning

a. Batch

- i. Trained with all available data
- ii. Done offline for large amount of time and computing resources
- iii. Can be automated

b. Online (Incremental)

- i. Train the system incrementally by feeding it data instances sequentially
- ii. Each learning step is fast and cheap



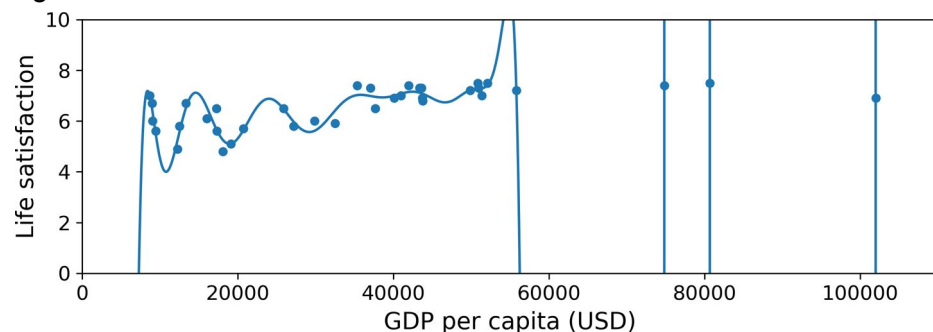
- iii.
- iv. Split large datasets into parts
- v. Run training step on the data and repeat the process until all data has been finished
- vi. Learning Rate: How fast the learning systems should adapt to changing data

6. Instance-Based & Model-Based Learning

- a. Instance: Measure of similarity; generalizes to new cases using a similarity measure
 - i. Number of words in common
- b. Model: Build a model of these examples, and use that model to make predictions

7. Common Issues:

a. Overfitting



- i.
- ii. When the model is too complex relative to the amount and noisiness of the training data
- iii. Can simplify the model by selecting one with fewer parameters
 - 1. Linear rather than polynomial
- iv. Gather more training data
- v. Fix noise (e.g. fix data errors and remove outliers)
- b. Underfitting:
 - i. Model is too simple to learn the underlying structure of the data
 - ii. Select a more powerful model with more parameters
 - iii. Feed better features to the learning algorithm

iv. Reduce the constraints on the model

Instructions on Environmental Set-up

1. Install Python: <https://www.python.org/downloads/>
2. Open Terminal
3. Check pip version by entering in: `pip3 --versions`
4. If you have pip installed, skip this step; otherwise, enter in your terminal:
 - a. `curl https://bootstrap.pypa.io/get-pip.py -o get-pip.py`
 - b. `python3 get-pip.py`
5. If you skipped step #4, enter in you terminal: `pip3 install --upgrade pip`
6. Now, it is time to install the packages we will be using for our workshops! In terminal, enter:
 - a. `pip3 install --upgrade jupyter matplotlib numpy pandas scipy scikit-learn`
7. Enter in your terminal: `jupyter notebook`