ECS 171 Machine Learning

Lecture 1

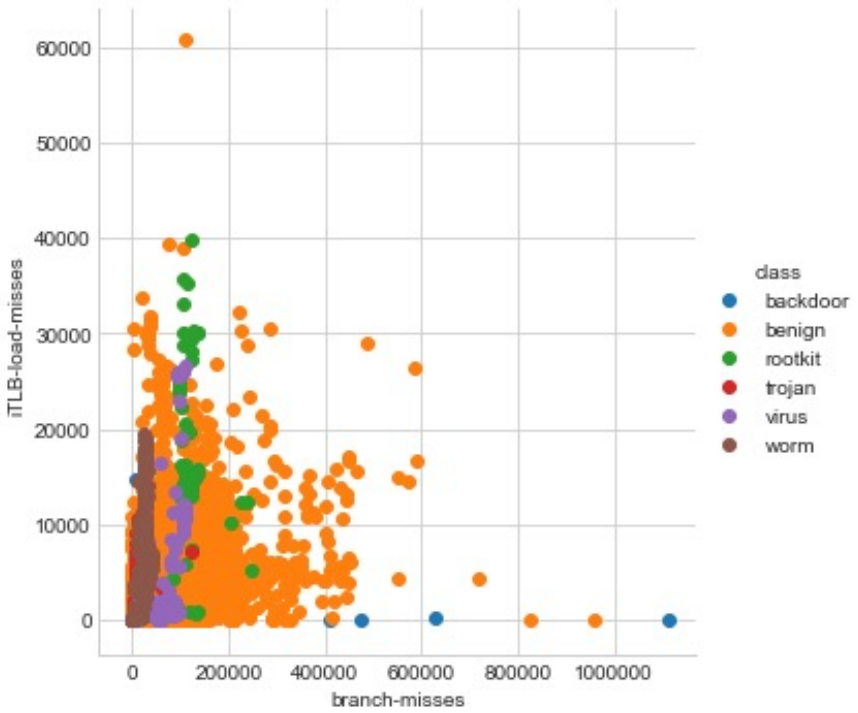
Introduction to Machine Learning
Instructor: Dr. Setareh Rafatirad

Outline

- What is Machine Learning?
- Machine Learning Process
- Machine Learning Categories

What is Machine Learning?

- Subfield of AI
- Enables systems to derive meaning from huge volume of data
- Learning from the data

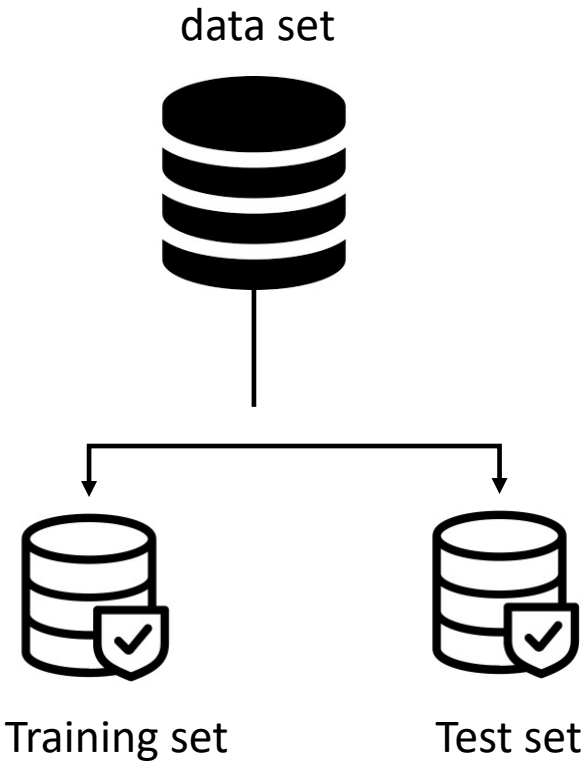


SOURCE: http://www.texasenterprise.utexas.edu/2014/01/27/innovation/does-big-data-mark-end-being-leadership-savvy

Definition (Machine Learning)

- Given a collection of observations (training set)
  - Each observation contains a set of attributes, where one of the attributes is the class.
- Task: find a model for class attribute as some function of the values of other attributes.
- The Goal is to assign a class to previously unseen records, as accurately as possible.
  - For this, we need a test set to evaluate the accuracy and robustness of the model.

Partition size:
- 70:30
- 80:20

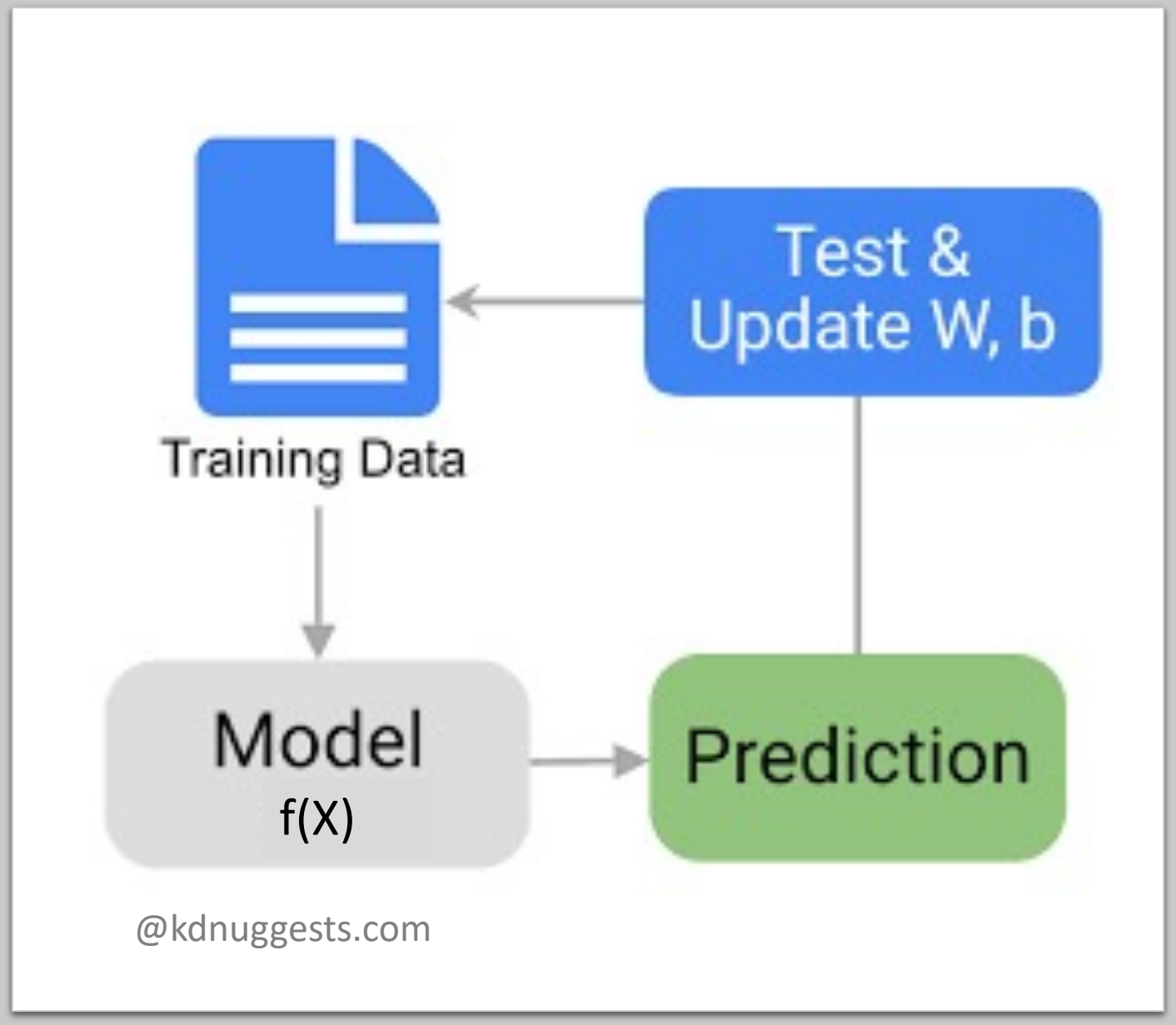


Model Evaluation Metrics

- TP,TN, FP, FN
- Precision and Recall
- Receiver Operator Characteristic (ROC) and Area under curve (AUC)
  from sklearn.metrics import roc_curve
  from sklearn.metrics import roc_auc_score
- Accuracy
- F1 Score
- Useful links
  - https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative
  - https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall
  - https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc
  - https://developers.google.com/machine-learning/crash-course/classification/accuracy
  - https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

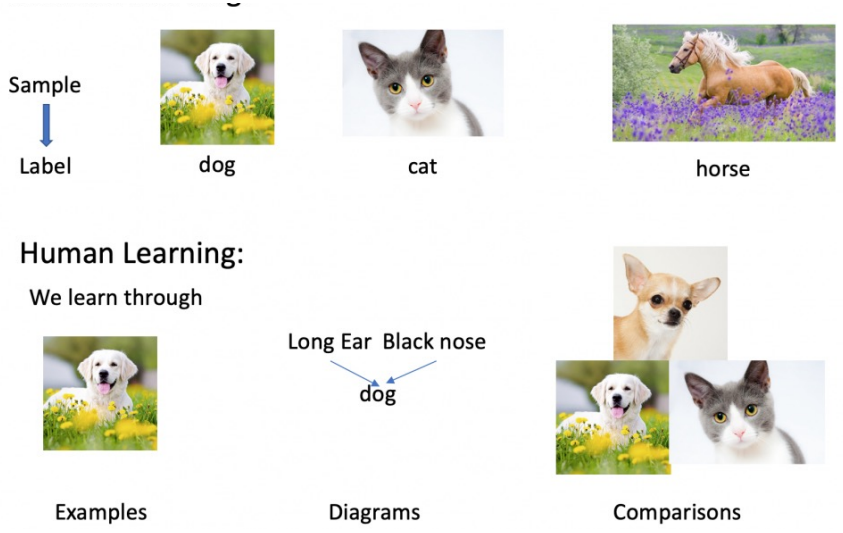
Machine Learning Process

- Machine learning algorithms are described as learning a target function (f) that best maps input variables (X) to an output variable (Y).
- Y = f(X)



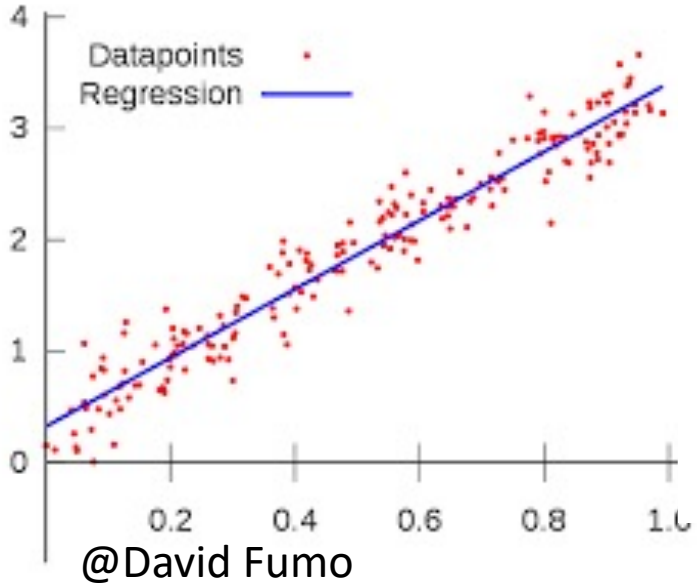
Machine Learning Tasks

Classification



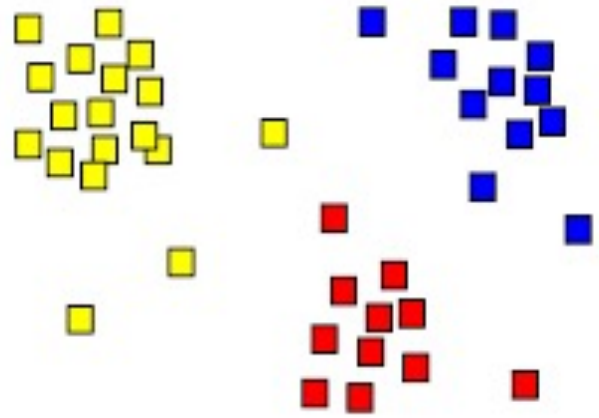
@CMU ML Blog

Regression



@David Fumo

Clustering



@Wikipedia

Illustrating Example

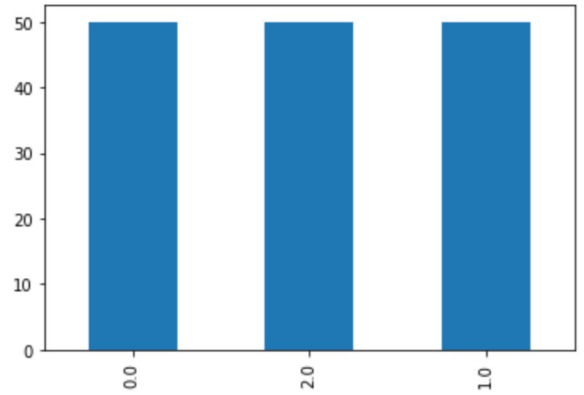
Tid	Refund	Marital Status	Taxable Income	Cheat
	categorical	categorical	continuous	class
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Refund	Marital Status	Taxable Income	Cheat
No	Single	75K	?
Yes	Married	50K	?
No	Married	150K	?
Yes	Divorced	90K	?
No	Single	40K	?
No	Married	80K	?

Python Example

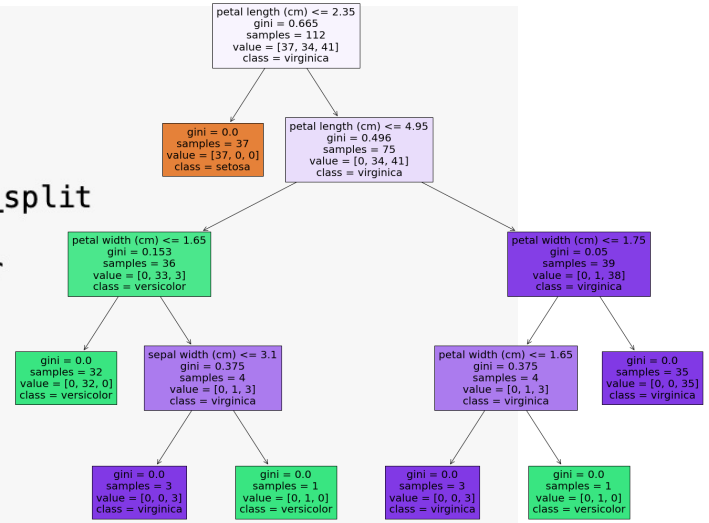
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   sepal length (cm)  150 non-null    float64
 1   sepal width (cm)   150 non-null    float64
 2   petal length (cm)  150 non-null    float64
 3   petal width (cm)   150 non-null    float64
 4   target             150 non-null    float64
dtypes: float64(5)
memory usage: 6.0 KB
```

Out[71]: <AxesSubplot:>



Dataset : Iris https://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html

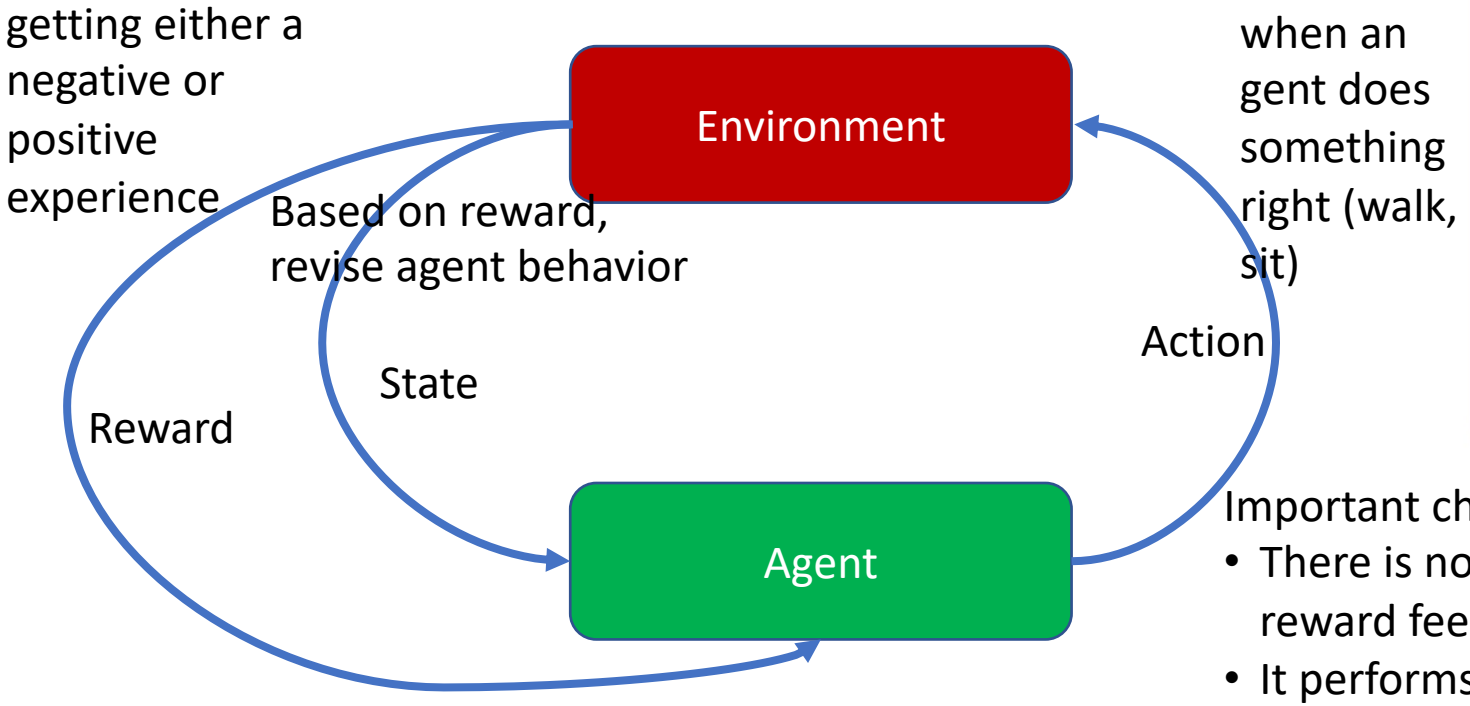
```python
#Classification Problem Example

import numpy as np
import pandas as pd
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from matplotlib import pyplot as plt
from sklearn.tree import DecisionTreeClassifier
from sklearn import tree
from sklearn import datasets

iris = datasets.load_iris()

X = iris.data
y = iris.target

X_train, X_test, y_train, y_test = train_test_split(X, y, random_state = 0)

from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier(random_state=1234)
dtree_model = clf.fit(X_train, y_train)
dtree_predictions = clf.predict(X_test)

cm = confusion_matrix(y_test, dtree_predictions)
print(cm)

fig = plt.figure(figsize=(25,20))
_ = tree.plot_tree(clf,
                   feature_names=iris.feature_names,
                   class_names=iris.target_names,
                   filled=True)
```



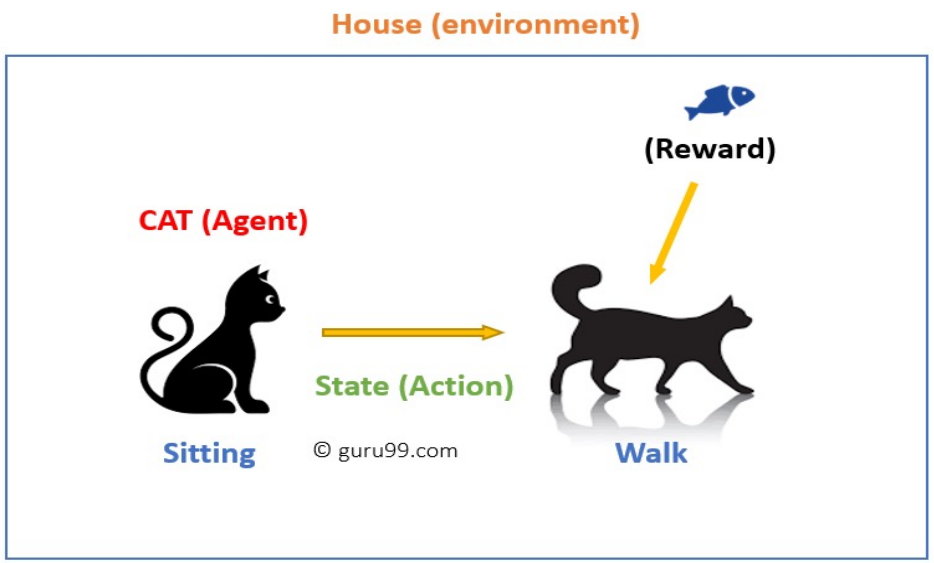
Machine Learning Categories

- Supervised Machine Learning
  - Labeled data set
  - Good for prediction
  - Example: Classification
- Unsupervised Machine Learning
  - Unlabeled data set
  - Good for data exploration and association rule discovery
  - Example: Clustering (discover similarities and differences)
- Reinforcement Learning
  - Interacts with its environment producing actions and discovers errors or rewards through trial and error search.
  - Example algorithm: Q-Learning

Reinforcement Learning



RL Examples

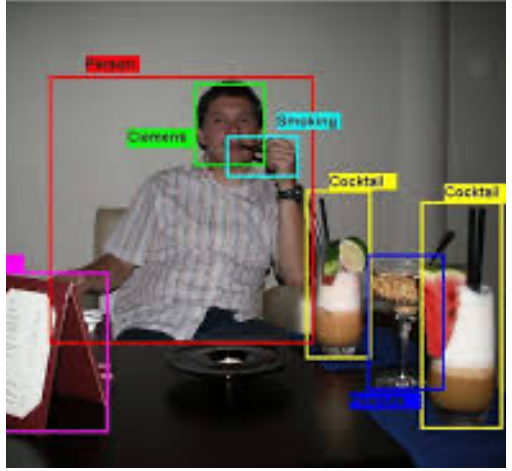


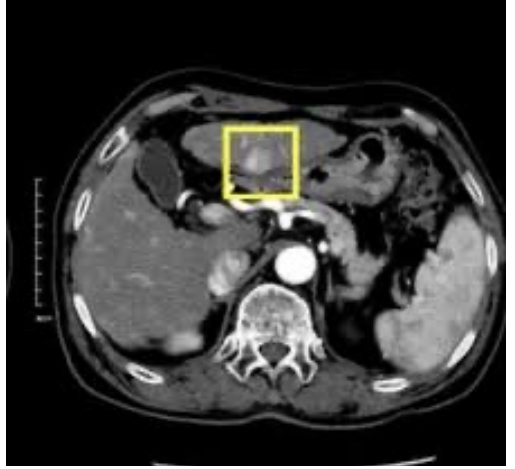
Important characteristics of RL:

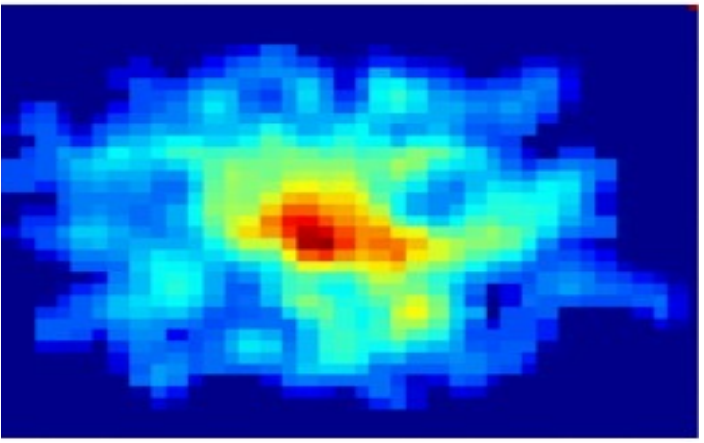
- There is no supervisor unlike supervised learning, only a reward feedback is used to train the algorithm
- It performs Sequential decision making
- Time has a crucial role in RL problems
- Feedback is always delayed, so it does not happen instantaneously to wait a while to see the result of the actions.
- Agent's actions determines the subsequent data it receives, e.g., moving a robot to the left or right.

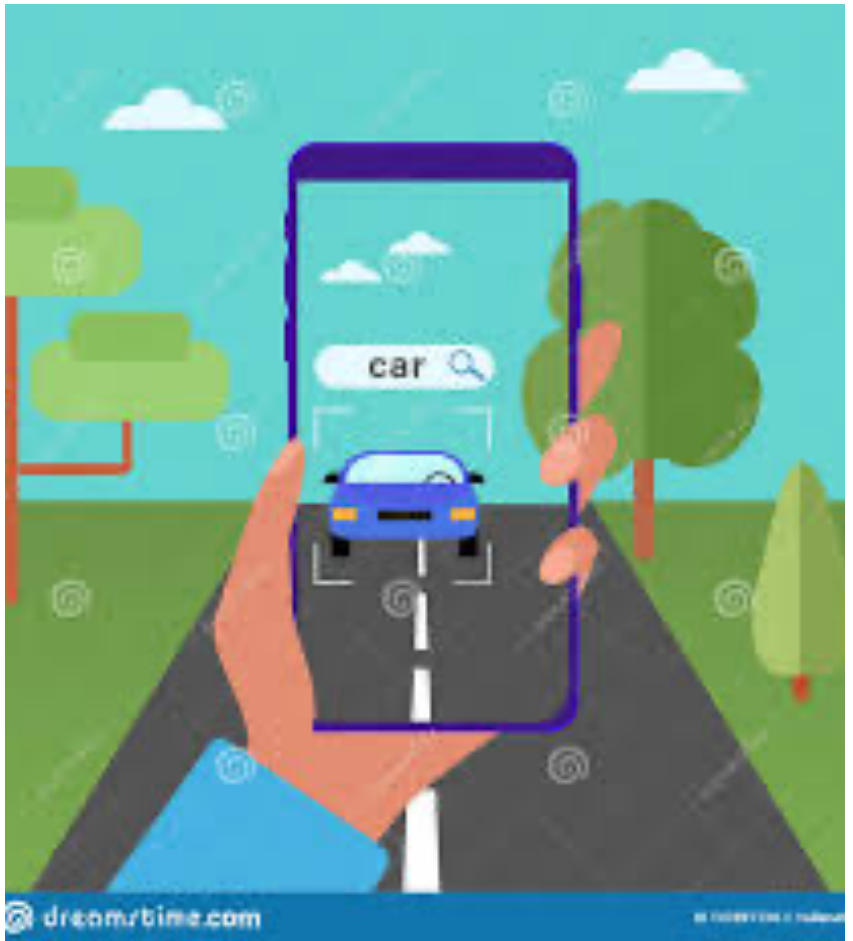
Two important learning models in reinforcement learning:

1. Markov Decision Process
2. Q learning







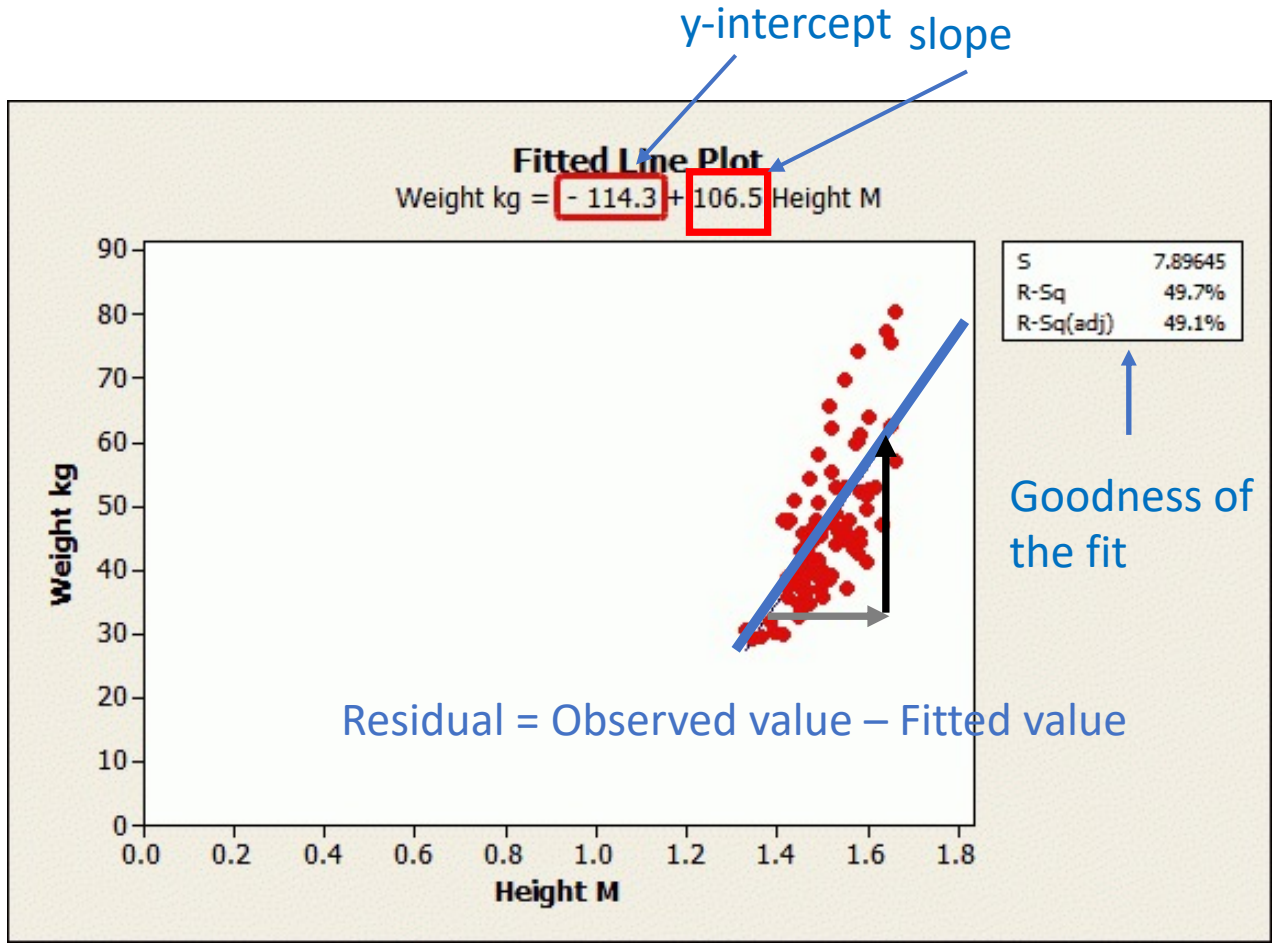


Classification

- Object Recognition
- Face Detection
- Face Recognition
- Scene Recognition
- Malware Detection
- Many more

Regression

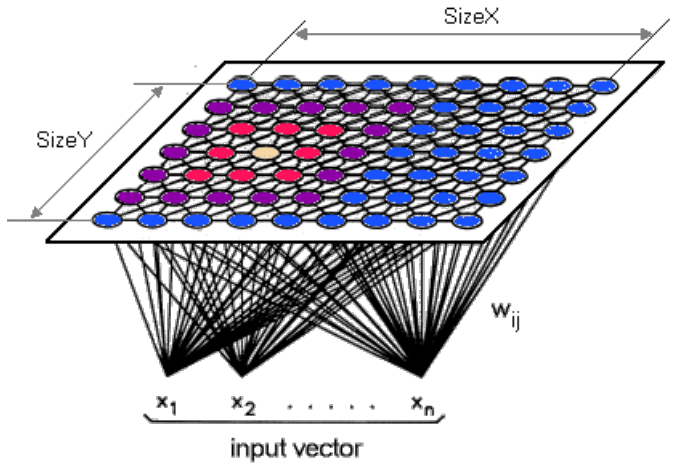
- Example Applications
  - Engine Performance Prediction
  - Business
  - Real-Estate Market Prediction
  - Stock Market Prediction
  - Weather Data Analysis



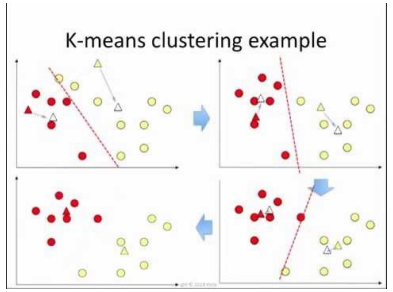
$$y = \beta + \alpha x$$

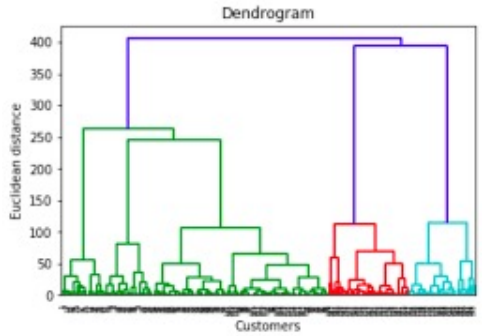
Unsupervised Learning

- Knowledge Discovery
- Data Exploration
- Descriptive Task
- Un-labeled Dataset
- Common Algorithms
  - Clustering
    - Example: K-means Clustering, Hierarchical Clustering, Self-Organizing Map
  - Association Rule Discovery



from http://www.lohninger.com/helpcsuite/kohonen_network_-_background_information.htm>





Unsupervised Learning: Association Rule Discovery

- Given a set of records each of which contain some number of items from a given collection
  - Produce dependency rules which will predict occurrence of an item based on occurrences of other items.
    - Support & Confidence measures

ID	
1	bread, Pepsi, milk
2	beer, bread
3	Pepsi, beer, diaper, milk
4	beer, bread, diaper, milk
5	Pepsi, diaper , milk

Rules Discovered:

{milk} → {Pepsi }
{Diaperm, milk} → {beer}

Machine Learning Workflow (Overview)

@Ilias Tagkopoulos

Step 1. Get enough data!

Dataset

Step 2. Do all of the data samples have labels?

$\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix}$ Yes

No $\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$

Step 3: The task is to predict a continuous variable, assign a new sample to a class, or perform an optimal action?

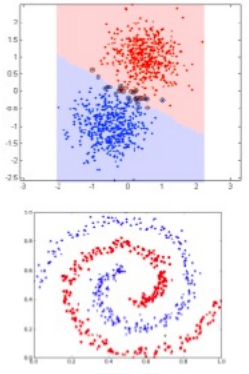
Supervised Learning

Assign to a class

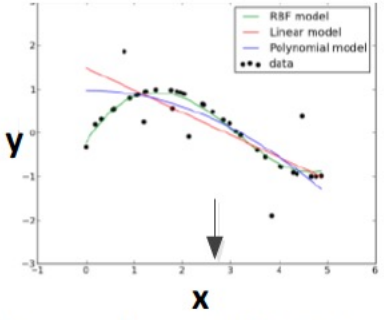
Predict a continuous variable

- Bayesian Classification (Naïve Bayes)
- Linear Discriminant Analysis
- Artificial Neural Networks
- Decision Trees
- Support Vector Machines

CLASSIFICATION



REGRESSION

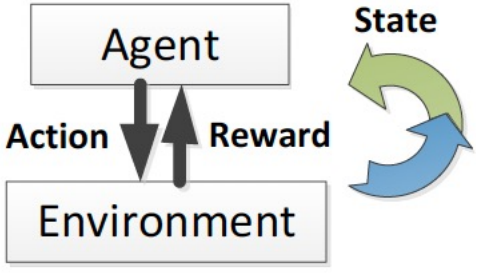


Linear, polynomial, logistic, ...

Reinforcement Learning

Perform an optimal action

REINFORCEMENT LEARNING (*)

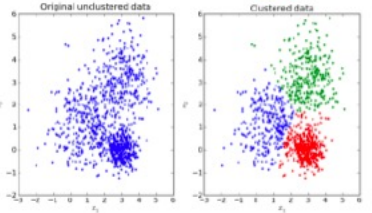


Markov Decision Process (MDP), POMDP, Q-learning,

Unsupervised Learning

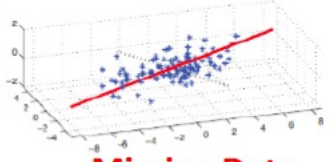
Step 3: The task is to cluster data together, find latent factors or complete missing data?

Clustering



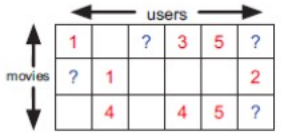
- K-means
- Hierarchical clustering
- SOM

Dimensionality Reduction



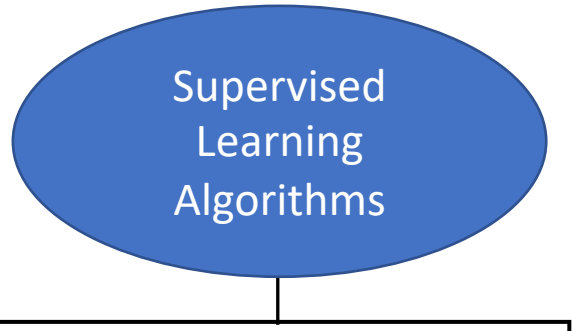
- PCA
- ICA

Missing Data



- Collaborative filtering
- Market Basket analysis
- ...

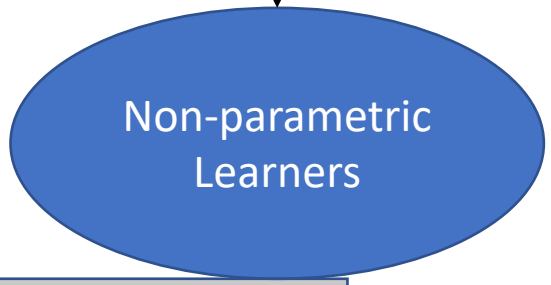
Parametric and Non-parametric Models in Machine Learning

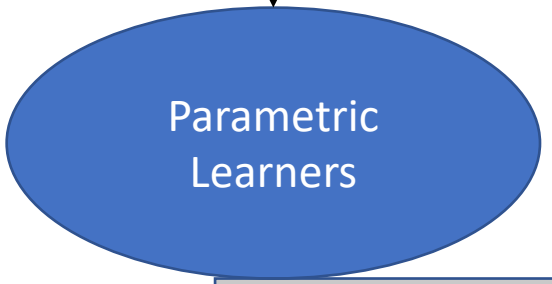


Decision Trees

RBF kernel SVMs

kNN





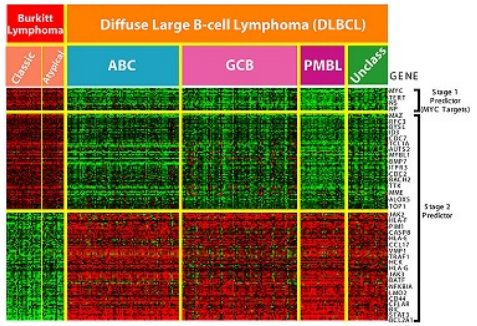
Linear models (linear regression, logistic regression, and linear SVM).

- (potentially) infinite number of parameters
- Lazy learners
- We need to store the training data
- does not estimate the parameters of a model during a training phase
- Can predict immediately
- Costly prediction

- finite number of parameters (fixed structure)
- Eager learners
- Once parameters (weights) of the model are learned, we no longer keep the training data.
- Training is computationally costly
- Inexpensive prediction

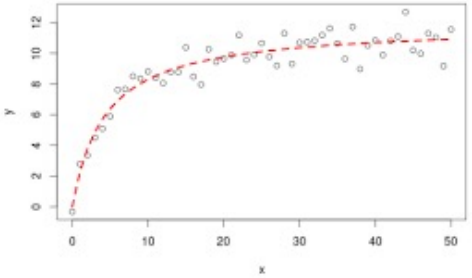
Basic Concepts in ML

- Linearity
- Dataset description
- Independent variables
- High-dimensional data
- Feature selection
- Overfitting , underfitting
- Error – variance trade-off
- Evaluation

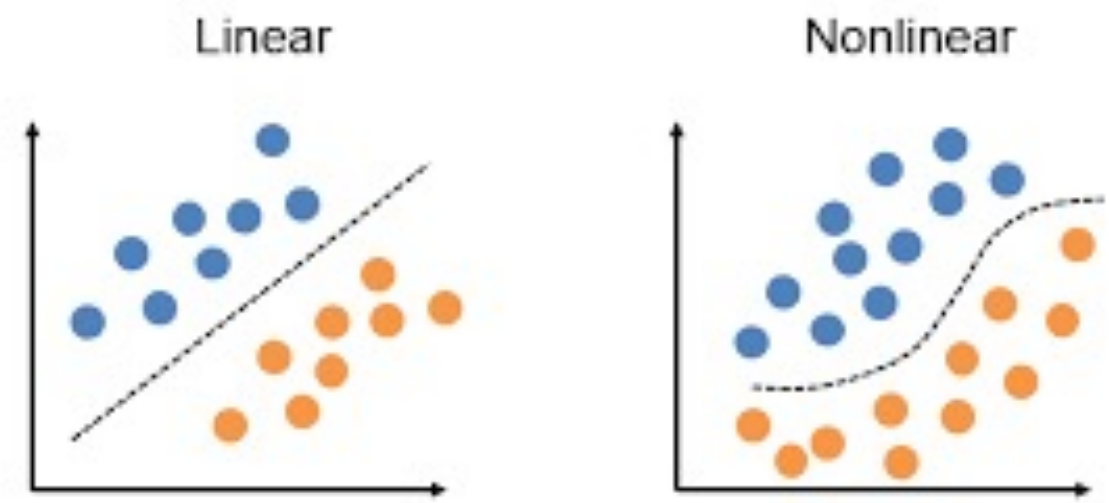


High dimensional Data

often in linear ML algorithms, we have a high bias but a low variance, or in Nonlinear ML algos, we often have a low bias but a high variance.



Nonlinear Regression



classification