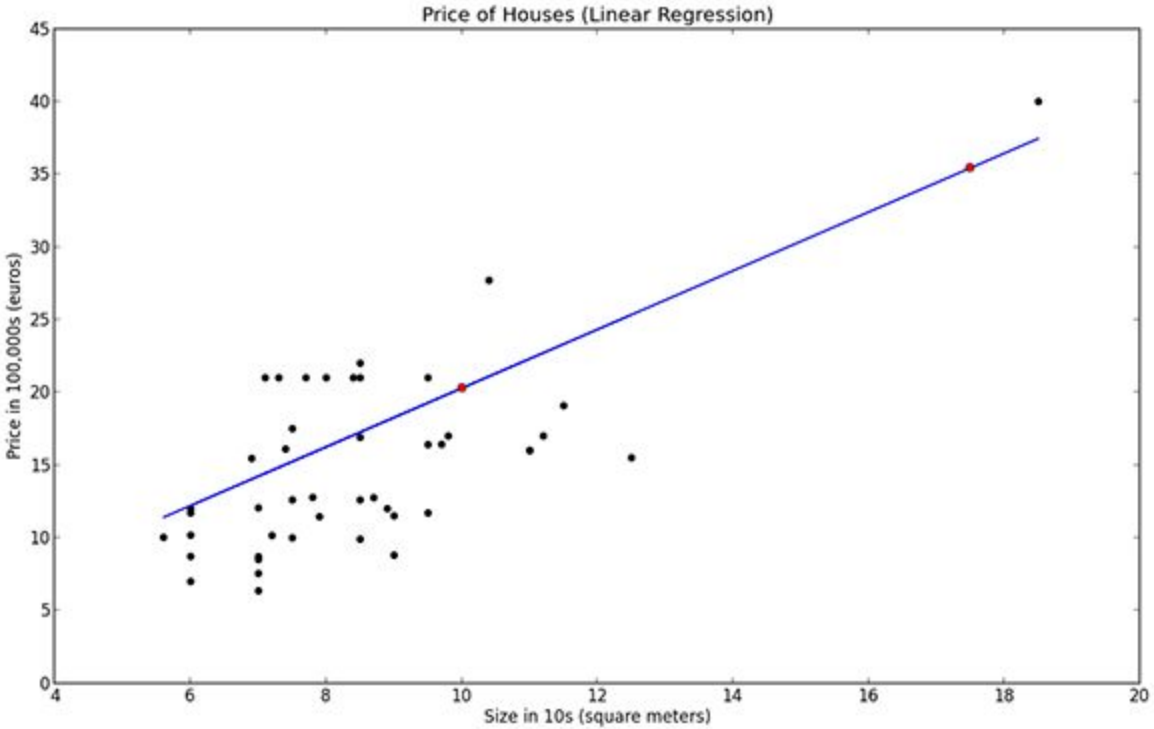
Basics of Machine Learning

By: Raja Hasnain Anwar

What is learning?

Machine's Learning



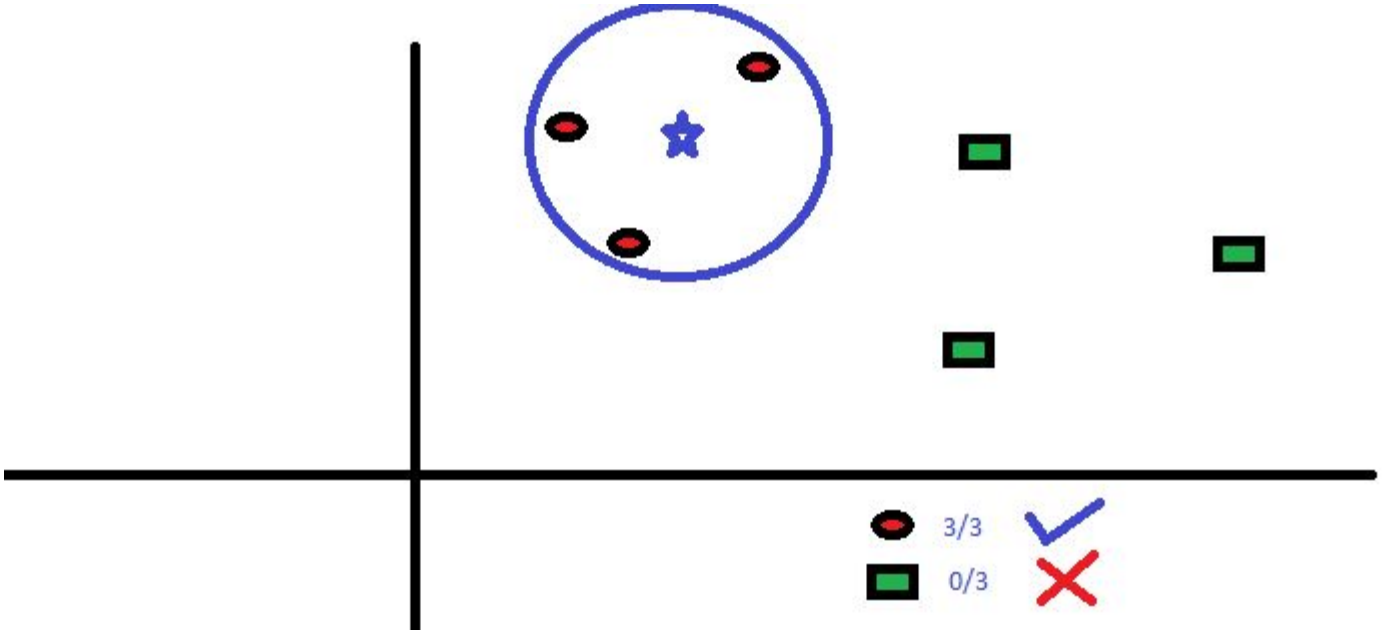
Components of Learning Process

Representation	Evaluation	Optimization
Representation/format of data i.e., its features. Also involves operations done on data as pre-processing.	Standard metrics to compare results of different algorithms. E.g., L1 loss.	Actual learning to fit and generalize on given data with maximum efficiency.

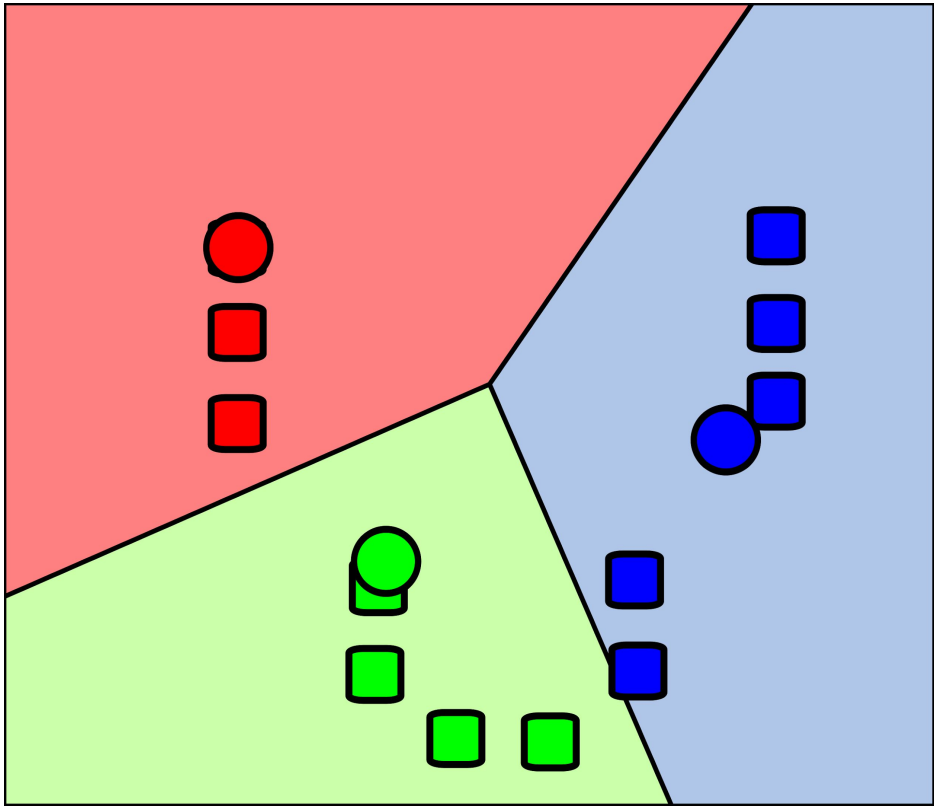
Data Representation

- k-nearest neighbors
- k-means clustering
- Decision Tree

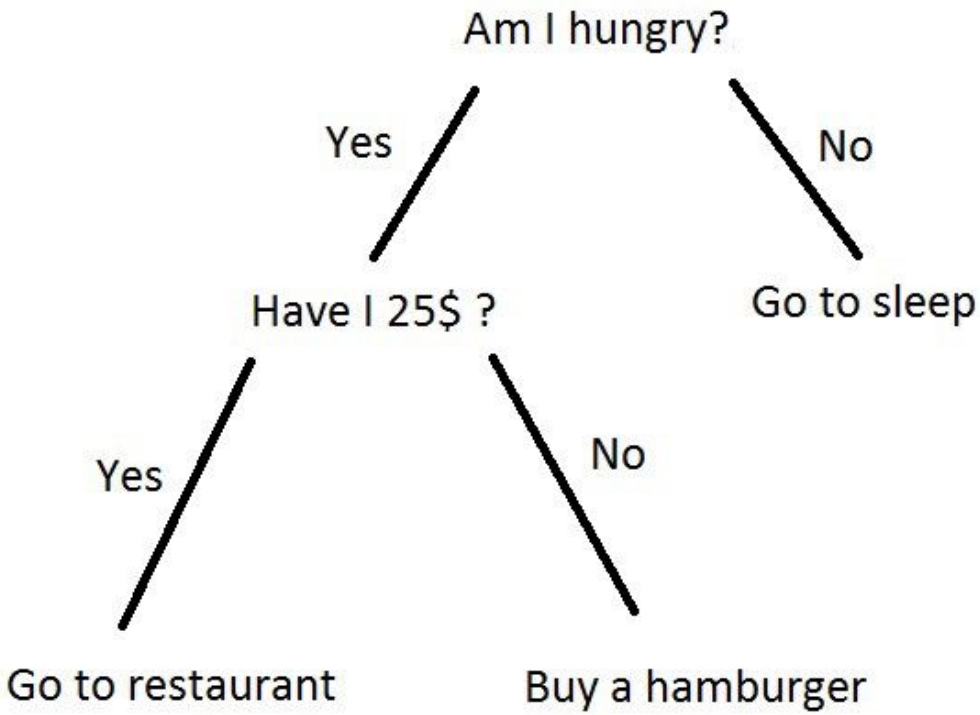
k-Nearest Neighbors



k-Means Clustering



Decision Tree



Evaluation

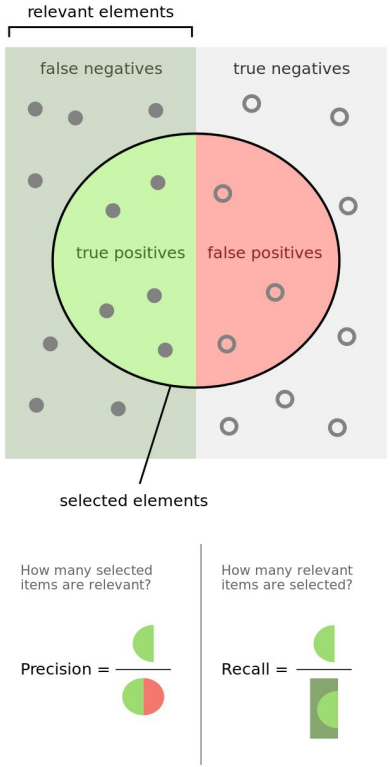
- L1 & L2 Loss
- Cross Entropy Loss
- K-L Divergence
- Git Loss
- Precision and Recall => F1

L1 & L2 Loss

$$L1LossFunction = \sum_{i=1}^{n} |y_{true} - y_{predicted}|$$

$$L2LossFunction = \sum_{i=1}^{n} (y_{true} - y_{predicted})^2$$

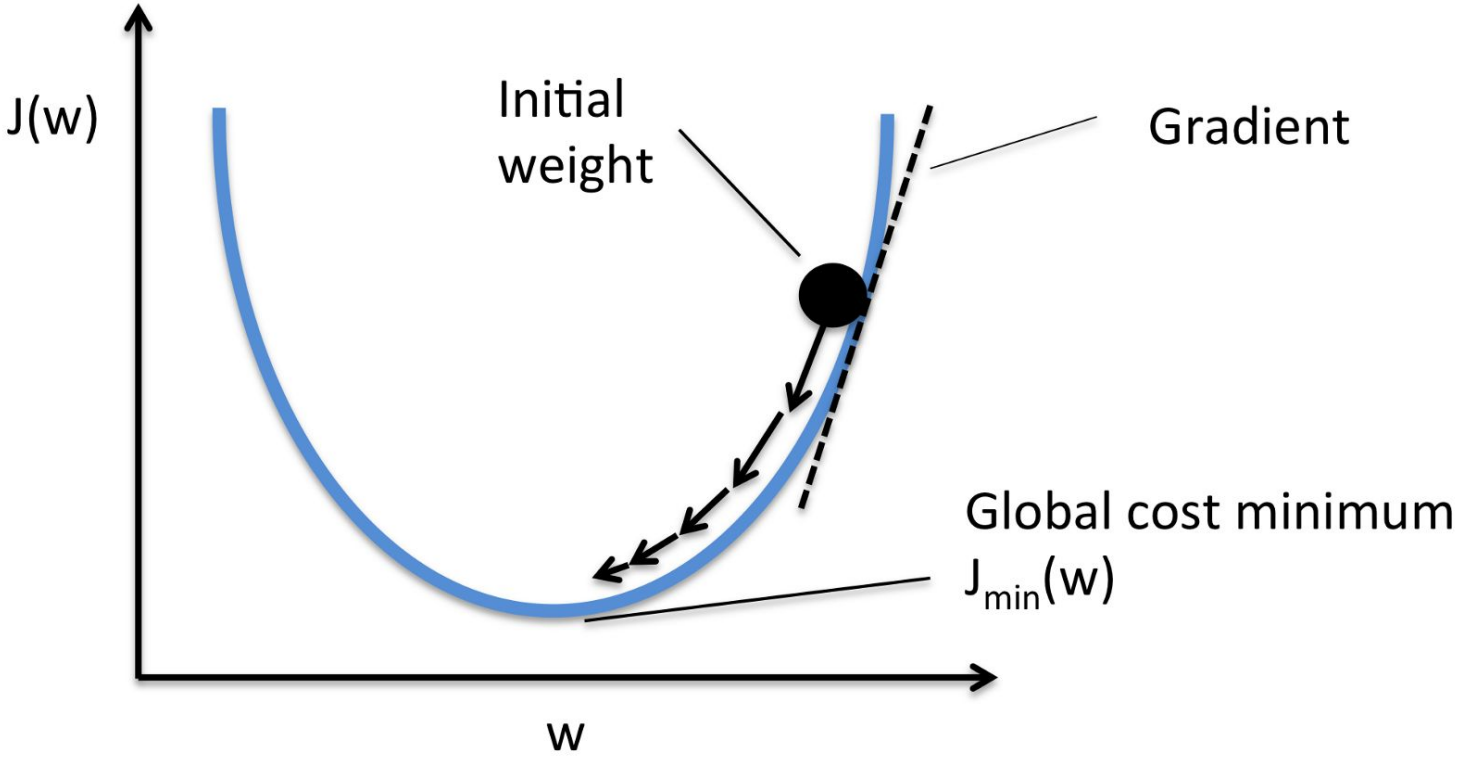
Precision and Recall



Optimization

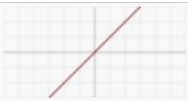
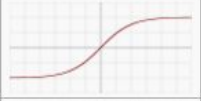
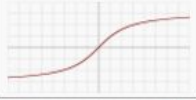
- Searching
- Gradient Descent
- Linear/Quadratic Programming

Gradient Descent



Things to keep in mind while OPTIMIZING

Activation Functions

Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$
Logistic (a.k.a Soft step)		$f(x) = \frac{1}{1+e^{-x}}$	$f'(x) = f(x)(1-f(x))$
TanH		$f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1$	$f'(x) = 1 - f(x)^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2+1}$
Rectified Linear Unit (ReLU)		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$

Bias vs Variance and Regularization