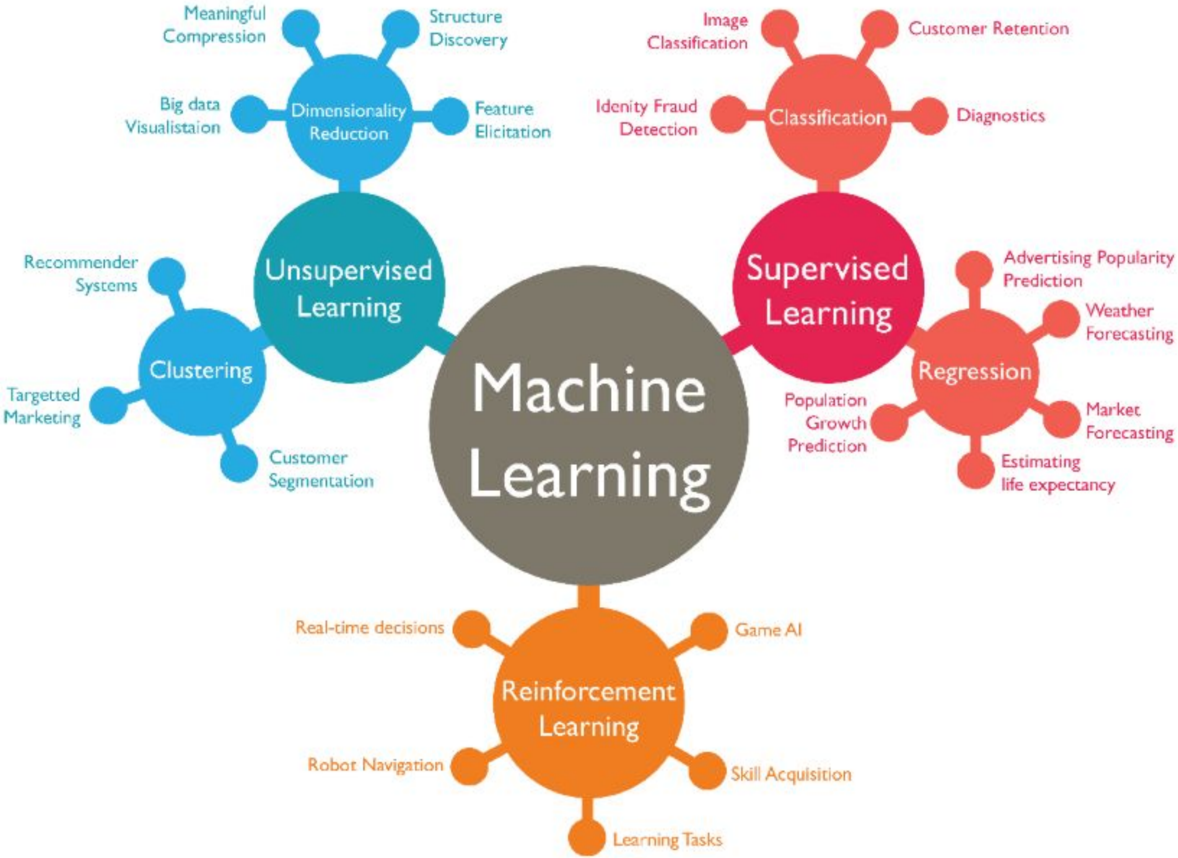
Logistic Regression

Machine Learning

Machine Learning

- Application of Artificial Intelligence that focuses of generating models using data and learning from it


Meaningful Compression
Structure Discovery
Big data Visualistaion
Dimensionality Reduction
Feature Elicitation
Unsupervised Learning
Recommender Systems
Clustering
Targetted Marketing
Customer Segmentation
Image Classification
Customer Retention
Idenity Fraud Detection
Classification
Diagnostics
Supervised Learning
Advertising Popularity Prediction
Weather Forecasting
Regression
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Machine Learning
Real-time decisions
Game AI
Reinforcement Learning
Robot Navigation
Skill Acquisition
Learning Tasks

Machine Learning

- Supervised Learning
    - You train the machine using data which is already labeled with the correct answer
    - The algorithm learns from labeled training data and predicts outcomes for unforeseen data

Machine Learning

- Supervised Learning
    - You train the machine using data which is already labeled with the correct answer
    - The algorithm learns from labeled training data and predicts outcomes for unforeseen data
- Unsupervised Learning
    - Machine finds patterns or discovers information on its own. Data is not labeled
    - Ability to cluster data with hidden features

Supervised Learning

- Regression
  - Predicts a single output value using training data

Supervised Learning

- Regression
  - Predicts a single output value using training data
- Classification
  - Group the output inside a class or category

Logistic Regression

- Model the probability of an event occuring depending on the values of independent variables

Logistic Regression

- Model the probability of an event occuring depending on the values of independent variables
- Estimate the probability that an event occurs for some observations versus the probability that the event does not occur

Logistic Regression

- Model the probability of an event occuring depending on the values of independent variables
- Estimate the probability that an event occurs for some observations versus the probability that the event does not occur
- Predict the effect of variables on some binary or multiclass response

Logistic Regression

- Model the probability of an event occuring depending on the values of independent variables
- Estimate the probability that an event occurs for some observations versus the probability that the event does not occur
- Predict the effect of variables on some binary or multiclass response
- Classify observations by estimating the probability that an observation is in a particular category

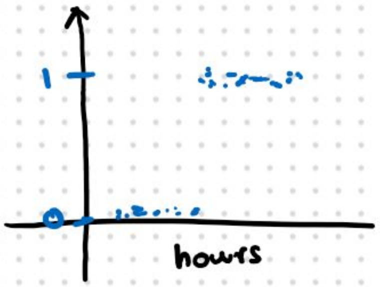
Logistic Regression

- Model the probability of an event occuring depending on the values of independent variables
- Estimate the probability that an event occurs for some observations versus the probability that the event does not occur
- **Predict** the effect of variables on some binary or multiclass response
- **Classify** observations by estimating the probability that an observation is in a particular category

Logistic Regression

Logistic Regression

hours	passed?
3	0
7	1
8	1
4	1
2	0



did I pass the class?

This becomes a classification problem.

$P(y_i = 1 \mid x ; \theta)$ - Probability that a given dependent value is 1 given some x, and some parameters θ.

Logistic Regression

$$P(y_i = 1 \mid x ; \theta) = \theta^T \bar{x}$$

$$\theta^T = [\theta_0, \theta_1, \ldots], \quad \bar{x} = [1, x_0, \ldots]$$

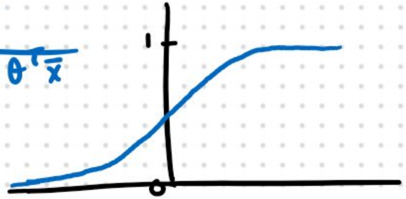
A linear model will not work here

A linear regression model could predict probabilities beyond 0 or 1 which is not possible.

So our hypothesis function becomes:

$$h(x) = \sigma(\theta^T \bar{x}) = \frac{1}{1 + e^{-\theta^T \bar{x}}}$$

where σ = sigmoid function



Logistic Regression

So, for 1 data point i;

$$P(y_i = 1 \mid x_i ; \theta) = \frac{1}{1 + e^{-\theta^T \bar{x}}}$$

and $P(y_i = 0 \mid x_i ; \theta) = 1 - h(x)$

Combining these together:

$$P(y_i \mid x_i ; \theta) = h(x_i)^{y_i} \left(1 - h(x_i)\right)^{1 - y_i}$$

Bernoulli Distribution.

For all data, X and Y:

$$L(\theta) = P(Y \mid X ; \theta) = \prod_{i=1}^{m} h(x_i)^{y_i} \left(1 - h(x_i)\right)^{(1 - y_i)}$$

$L(\theta)$ = likelihood function for θ.

Logistic Regression

$L(\theta)$ represents how plausible the model (θ parameters) is given all of my data points.

★ It is assumed that data points are independent

How to find optimal θ parameters? maximize $L(\theta)$.

But it is hard to maximize this function with lots of probabilities multiplied.

So do Log likelihood

$$\mathcal{L}(\theta) = \sum_{i=1}^{m} y_i \log(\sigma(\theta^T \bar{x}_i)) + (1 - y_i) \log(\sigma(\theta^T \bar{x}_i))$$

Logistic Regression

$$\mathcal{L}(\theta) = \sum_{i=1}^{m} y_i \log\left(\sigma(\theta^T \bar{x}_i)\right) + (1-y_i)\log\left(\sigma(\theta^T \bar{x}_i)\right)$$

Since we want to maximize this function, and not minimize, we will do gradient <u>ascent</u>.

First find derivative :

for 1 data point ;

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = \frac{y}{\sigma(\theta^T \bar{x})}\left(\frac{\partial \sigma(\theta^T \bar{x})}{\partial \theta}\right) + \frac{1-y_i}{\sigma(\theta^T \bar{x})} \cdot -\left(\frac{\partial \sigma(\theta^T \bar{x})}{\partial \theta}\right)$$

$[\log(x)' = 1/x]$ [chain rule]

Note that : $$\frac{\partial \sigma(\theta^T \bar{x})}{\partial \theta} = \underbrace{\frac{\partial \sigma(\theta^T \bar{x})}{\partial(\theta^T \bar{x})}}_{\sigma(\theta^T \bar{x})(1-\sigma(\theta^T \bar{x}))} \cdot \frac{\partial(\cancel{\theta^T \bar{x}})}{\cancel{\partial \theta}} \to \bar{x}$$

Logistic Regression

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = y - \sigma(\theta^T \bar{x})\,\bar{x}$$

$$= (y - h(x))\,\bar{x}$$

$$\theta^+ = \theta^- \pm \alpha \frac{\partial \mathcal{L}(\theta)}{\partial \theta}$$

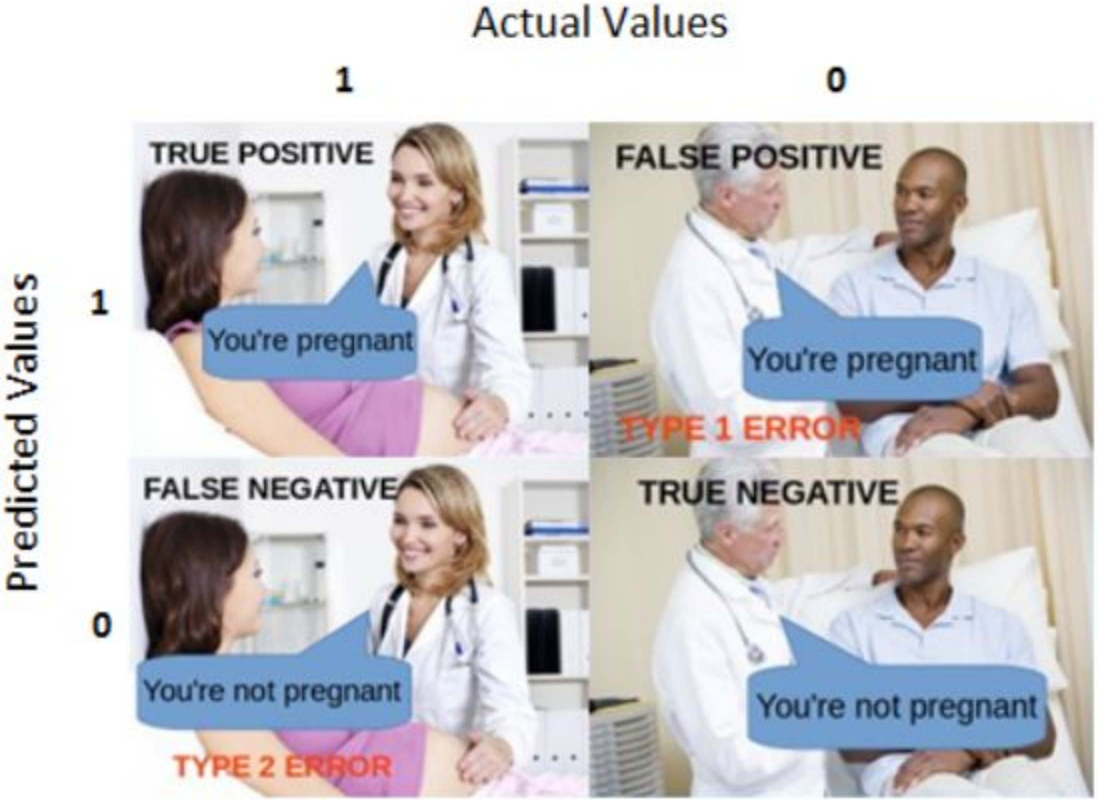
θ^+ = new θ values

θ^- = old θ values

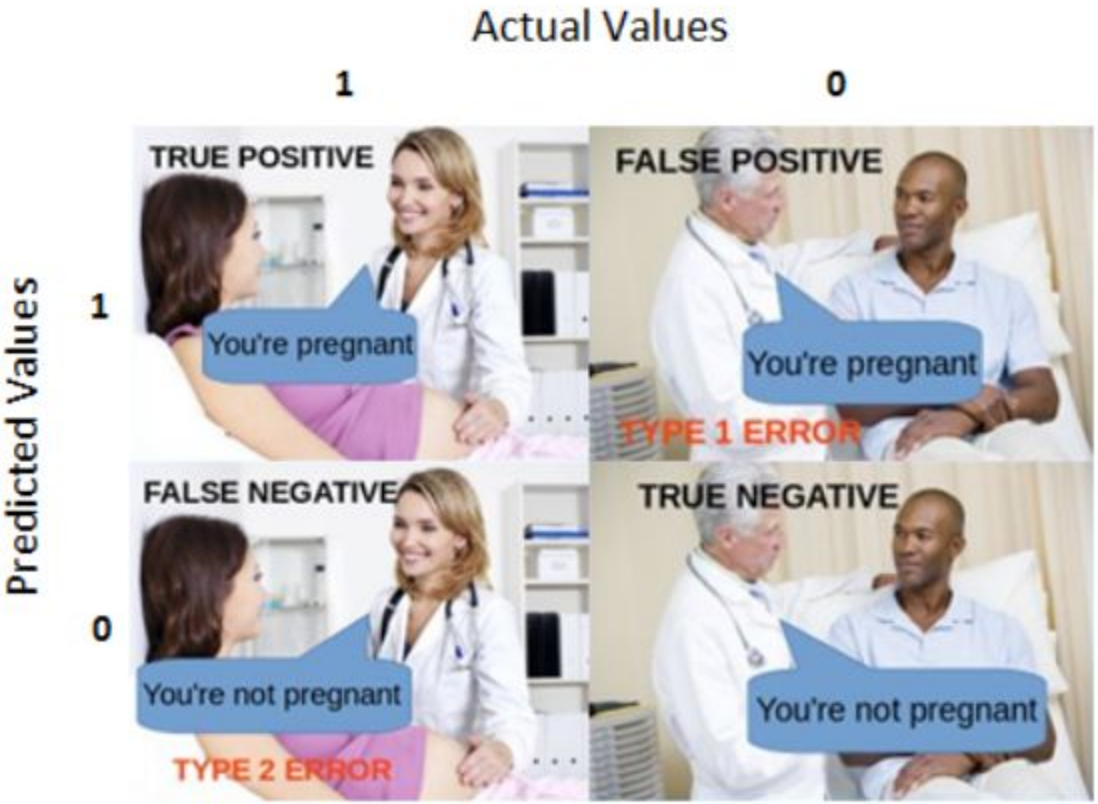
α = step size / learning rate

Confusion Matrix

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN



Actual Values
1
0
Predicted Values
1
0
TRUE POSITIVE
You're pregnant
FALSE POSITIVE
You're pregnant
TYPE 1 ERROR
FALSE NEGATIVE
You're not pregnant
TYPE 2 ERROR
TRUE NEGATIVE
You're not pregnant



TRUE POSITIVE (TP)
You predicted positive and it's true

TRUE NEGATIVE (TN)
You predicted negative and it's true

FALSE POSITIVE (FP)
You predicted positive and it's false

FALSE NEGATIVE(FN)
You predicted negative and it's false

Confusion Matrix (Measures)

1. Recall = TP / (TP+FN)
    a. Out of all the positive classes, how much we predicted correctly. Should be as high as possible

Confusion Matrix (Measures)

1. Recall = TP / (TP+FN)
    a. Out of all the positive classes, how much we predicted correctly. Should be as high as possible
2. Precision = TP / (TP + FP)
    a. Out of all the positive classes we have predicted correctly, how many are actually positive

Confusion Matrix (Measures)

1. Recall = TP / (TP+FN)
    a. Out of all the positive classes, how much we predicted correctly. Should be as high as possible
2. Precision = TP / (TP + FP)
    a. Out of all the positive classes we have predicted correctly, how many are actually positive
3. Accuracy = (TP+TN) / Total
    a. Out of all classes, how much did we predict correctly