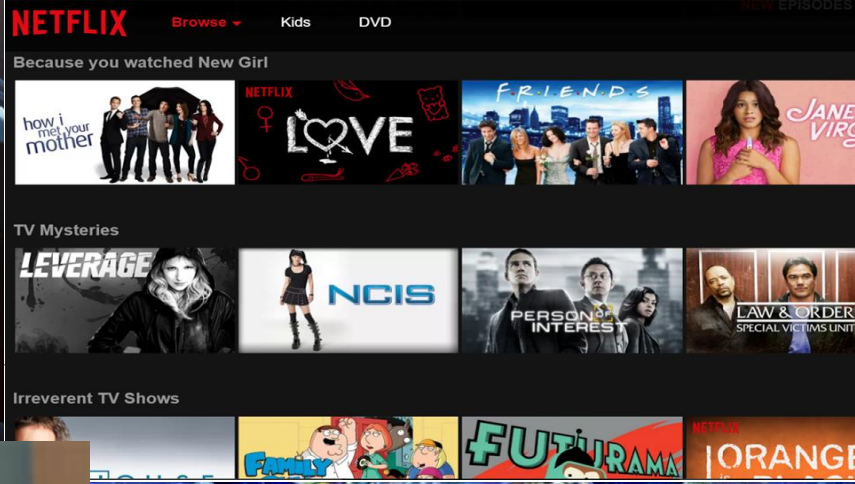
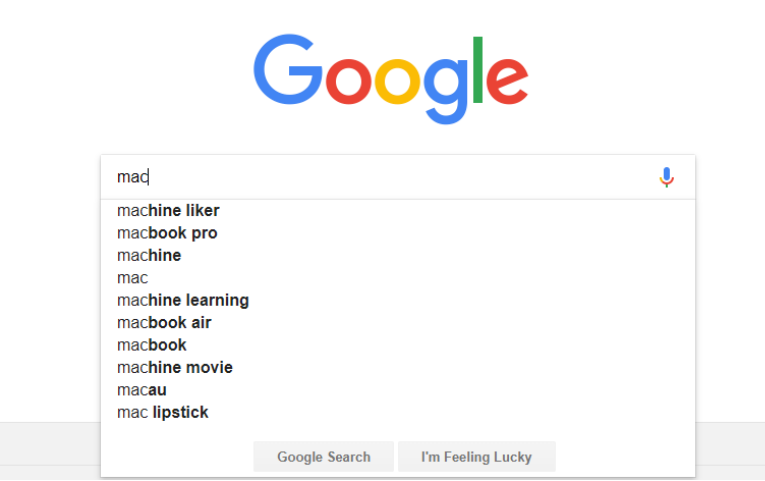
ML-101:
An Introduction to Machine Learning

BY SARTHAK CONSUL

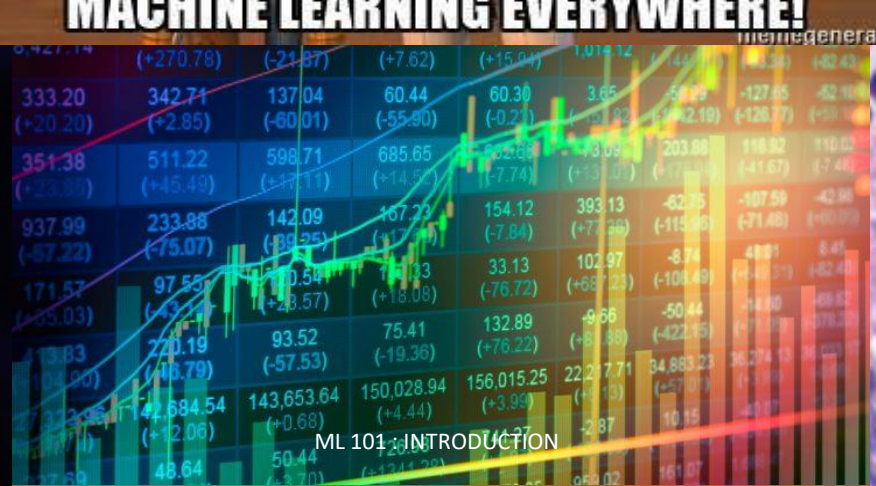

SPEED LIMIT 30
UBER
UBER

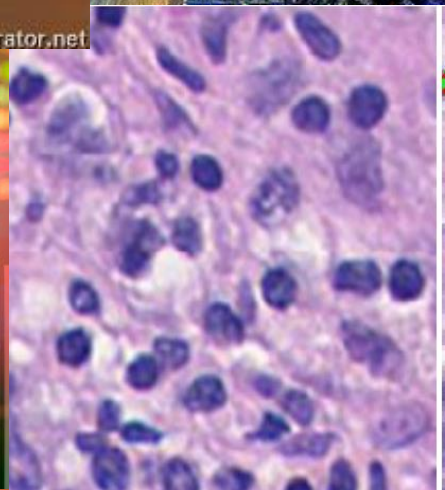

相似度
67.2%

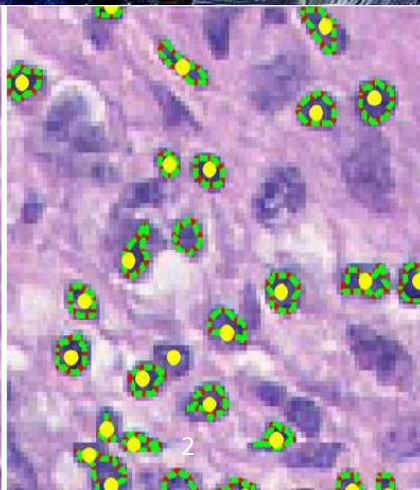

NETFLIX
Browse
Kids
DVD
Because you watched New Girl
how i met your mother
LOVE
FRIENDS
TV Mysteries
LEVERAGE
NCIS
PERSON OF INTEREST
LAW & ORDER
SPECIAL VICTIMS UNIT
Irreverent TV Shows
FAMILY GUY
FUTURAMA
ORANGE


Google
mac
machine liker
macbook pro
machine
mac
machine learning
macbook air
macbook
machine movie
macau
mac lipstick
Google Search
I'm Feeling Lucky


MACHINE LEARNING
MACHINE LEARNING EVERYWHERE!
memegenerator.net







"Machine learning is glorified statistics"

"Machine learning is statistics scaled up to big data"

"Machine learning is for Computer Science majors who couldn't pass a Statistics course." – A disgruntled statistician

ML is an HYBRID field

- (Convex) Optimization
- Bayesian Statistics
- Lots and lots of linear algebra
- Sampling Theory
- Inspiration from biology, psychology

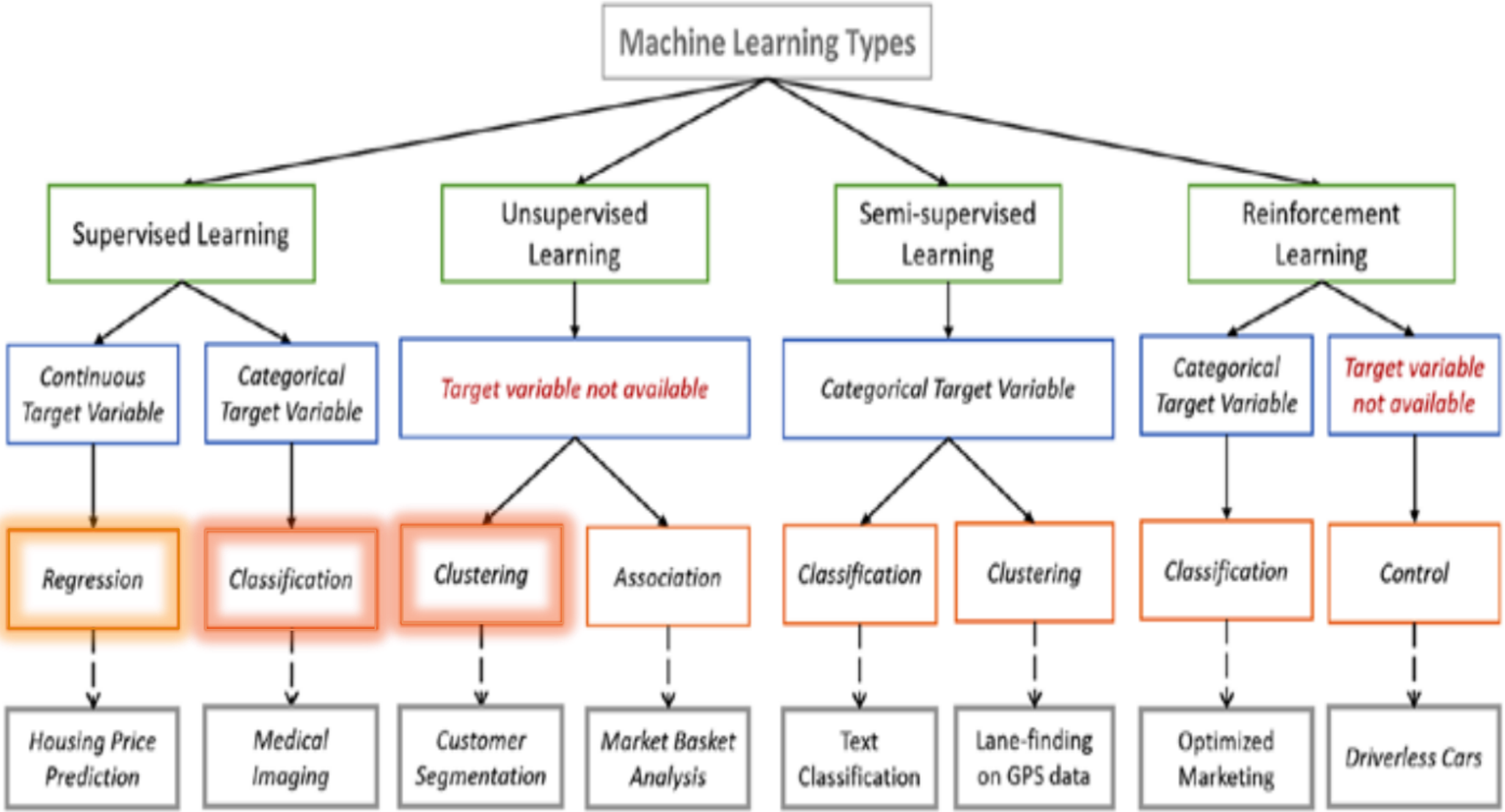
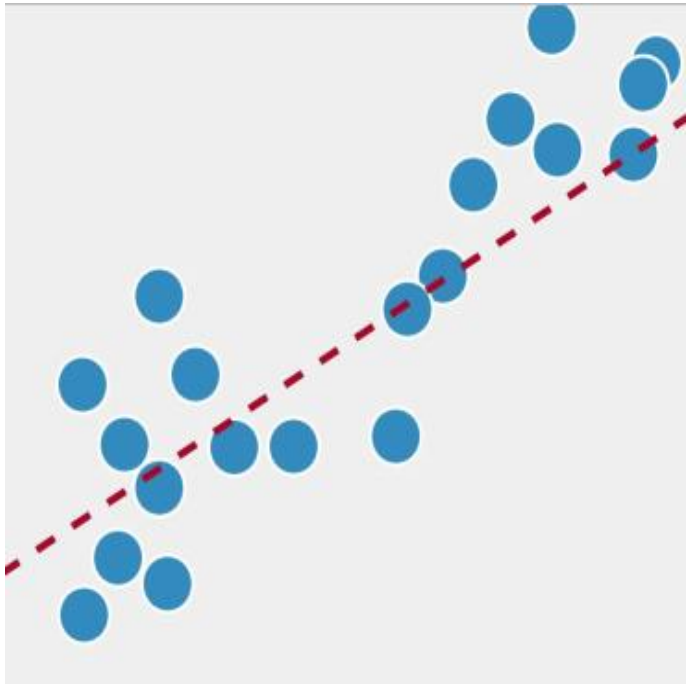
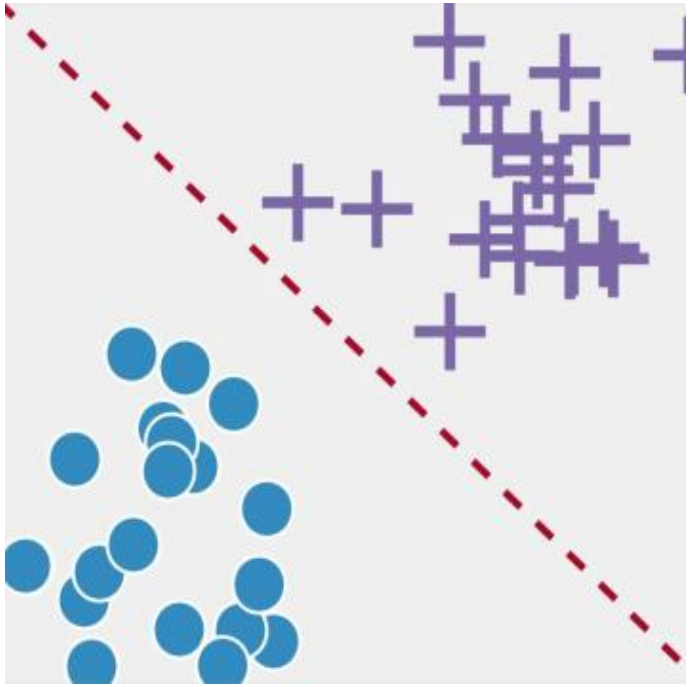

Machine Learning Types
Supervised Learning
Unsupervised Learning
Semi-supervised Learning
Reinforcement Learning
Continuous Target Variable
Categorical Target Variable
Target variable not available
Categorical Target Variable
Categorical Target Variable
Target variable not available
Regression
Classification
Clustering
Association
Classification
Clustering
Classification
Control
Housing Price Prediction
Medical Imaging
Customer Segmentation
Market Basket Analysis
Text Classification
Lane-finding on GPS data
Optimized Marketing
Driverless Cars

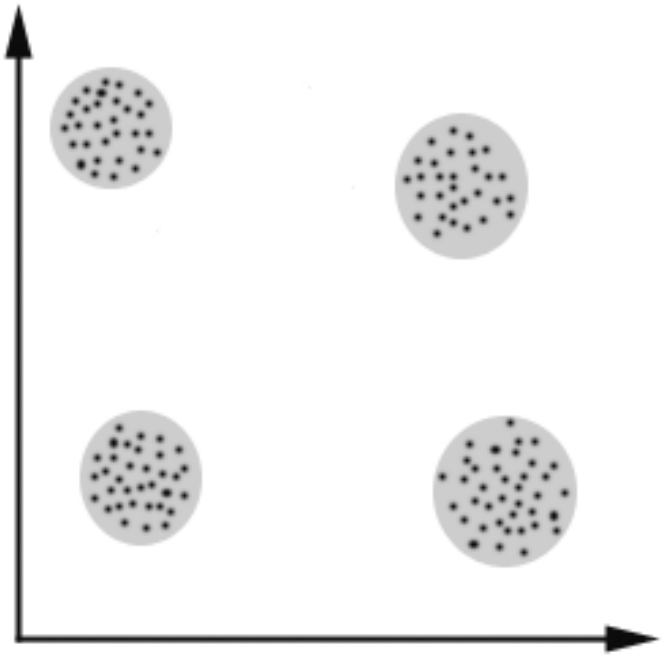
Image Source:
https://medium.com/@heyozramos/regression-vs-classification-86d73c281c5e
home.deib.polimi.it



Regression



Classification



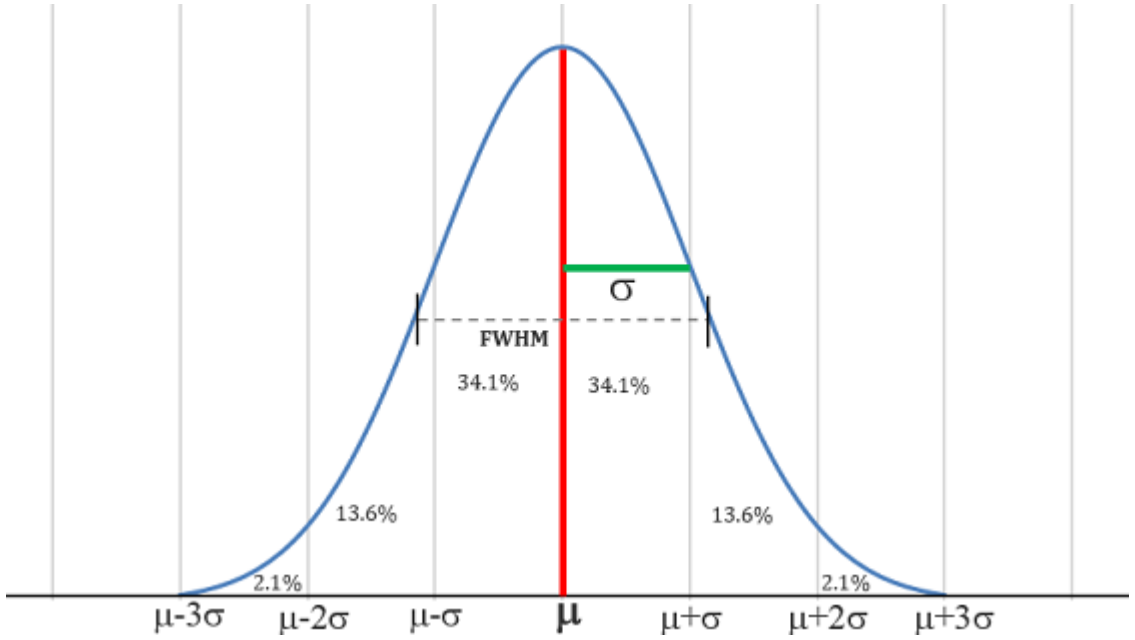
Clustering

Image Source:
https://sites.google.com/site/nuclearremotelaboratory/prepare-the-samples/lesson-1/exp-1-roadsite-plants/analyze-data/gaussian-distribution

Some Basic Statistics

- Probability Distributions

The probability of a real-valued variable x falling in the interval $(x, x + \delta x)$ is given by $p(x)\delta x$ for $\delta x \to 0$, then $p(x)$ is called the probability density over x



The Gaussian Distribution

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

Some Basic Statistics contd.

- Mean

$$\mathbb{E}[f] = \sum_x p(x) f(x) \qquad \mathbb{E}[f] = \int p(x) f(x)\, \mathrm{d}x.$$

- Variance

$$\mathrm{var}[f] = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right] \qquad \mathrm{var}[f] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2.$$

- Covariance

$$\begin{aligned} \mathrm{cov}[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y] \end{aligned}$$

- Standard Deviation

Pre-processing of Data

- Remove noise
- Filtering and Sampling
- Feature Scaling and Zero-Mean [Normalization]
- Building dictionaries
- Encoding
- Handling missing data
- PCA*
- Whitening

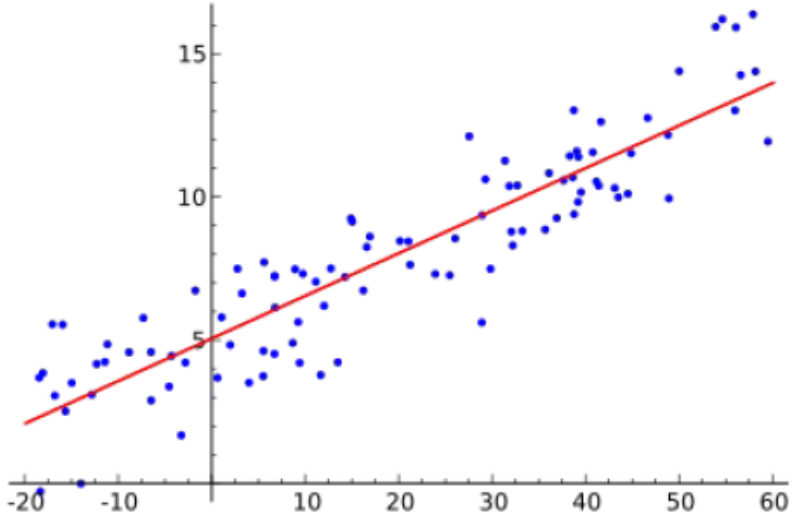
Image Source:
https://en.wikipedia.org/wiki/Linear_regression

Linear Regression

- Hypothesis Function: $h_\theta(\mathbf{x}) = \theta_0 + \theta_1.x_1 + \ldots + \theta_n.x_n$
- Cost Function, $J(\boldsymbol{\theta})$ – Mean Squared Error

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

- Optimizing the cost Function via Gradient Descent



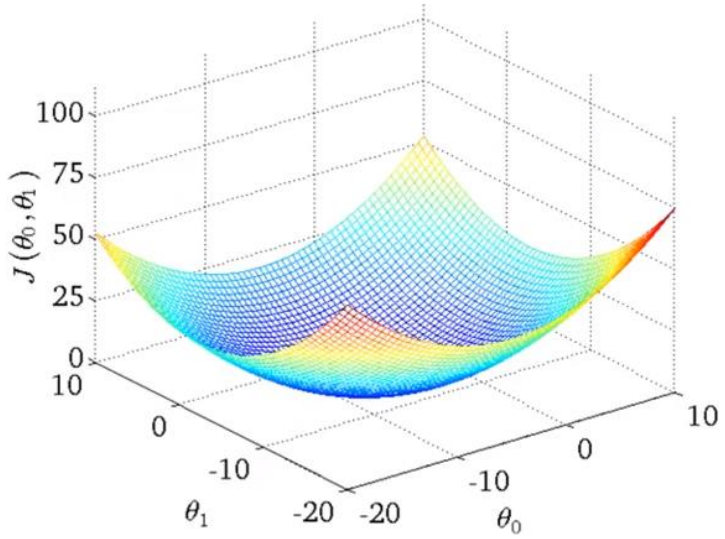
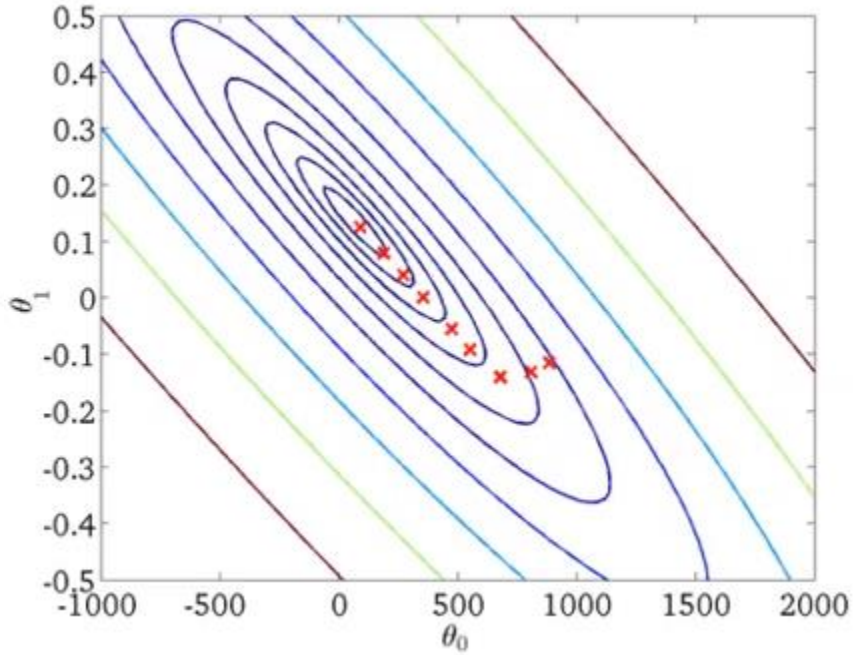


Image Source: Coursera - Machine Learning by Andrew Ng

Gradient Descent

- Cost Function

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$



- Gradient Descent

Repeat till convergence{

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)}$$

}

Image Source: Coursera - Machine Learning by Andrew Ng

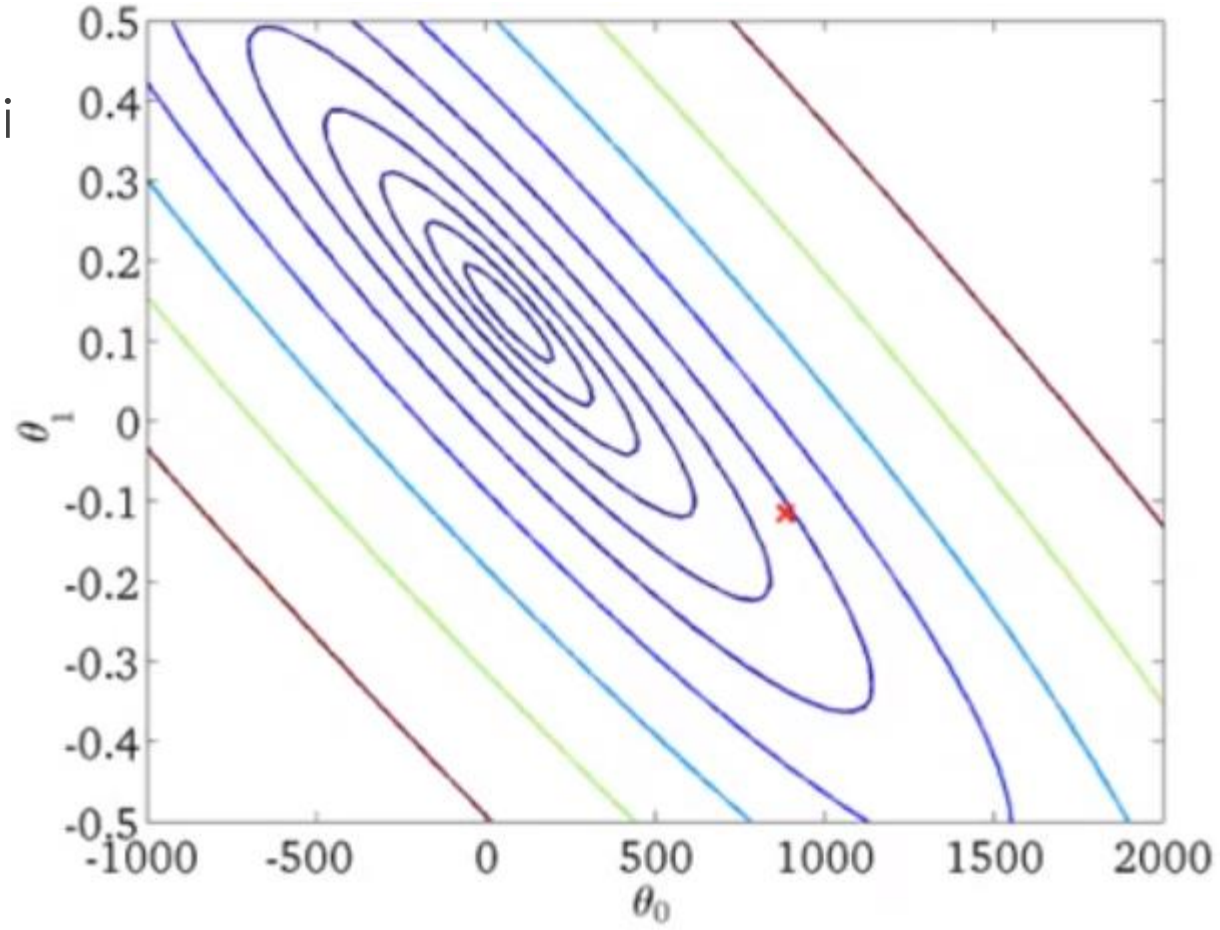
Gradient Descent contd.

- To improve convergence
  - Feature Scaling and Mean Normali

$$x' = \frac{x - \bar{x}}{\sigma}$$

- Choosing the learning rate, α

- Batch Gradient Descent
- Stochastic Gradient Descent
- Mini-Batch Gradient Descent



Modifications to SGD

- SGD + Momentum
- Nesterov
- AdaGrad
- RMSProp
- Adam

...

```
dx = compute_gradient(x)
x += learning_rate *dx
```

```
dx = compute_gradient(x)
vx = rho*vx + dx
x += learning_rate * vx
```

Cross Validation

- While training model, test data should NOT be touched
- Limited Data
- Split Data into train, **validation**, (test)

- Holdout Method, K-fold CV, Leave-One-Out CV