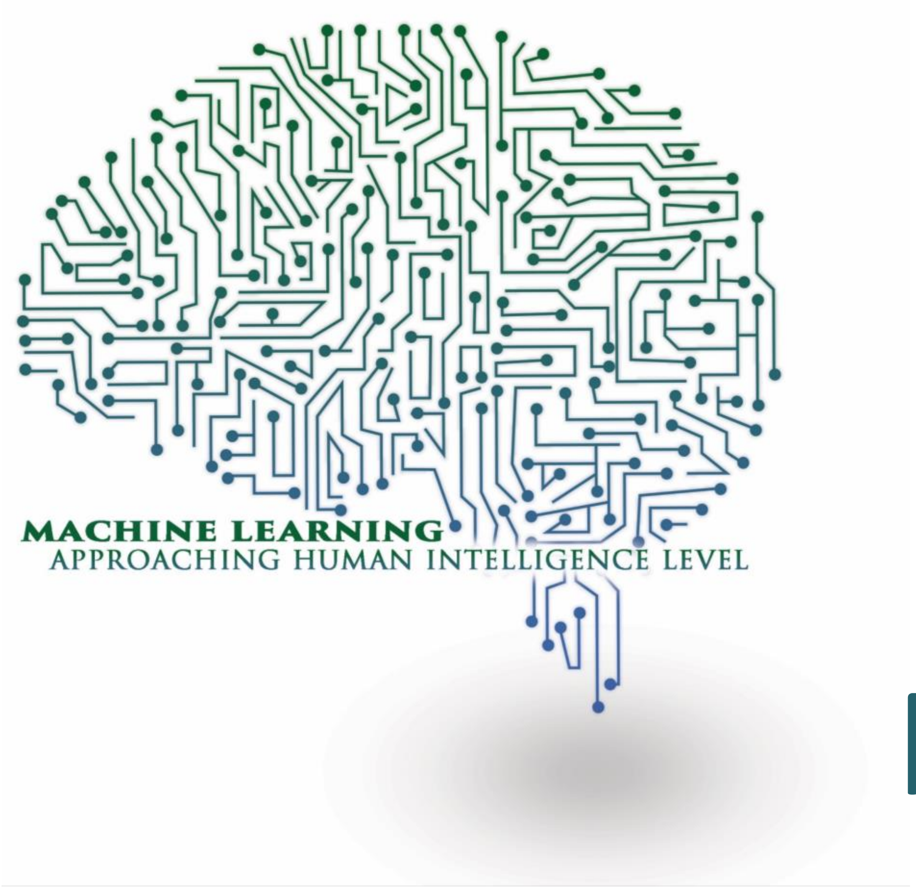


Machine learning

Introduction to ML and Linear Regression

- The goals of AI is to create a machine which can mimic a human mind and to do that it needs learning capabilities, but also more; knowledge representation , logical reasoning and event thighs like abstract thinking.

- Machine learning on the other hand focus on writing software or algorithms that can learn from past experience.

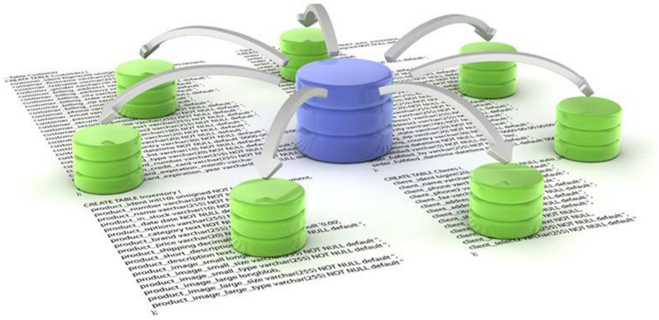
- Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.

- Tom Mitchell (1998) definition "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E".

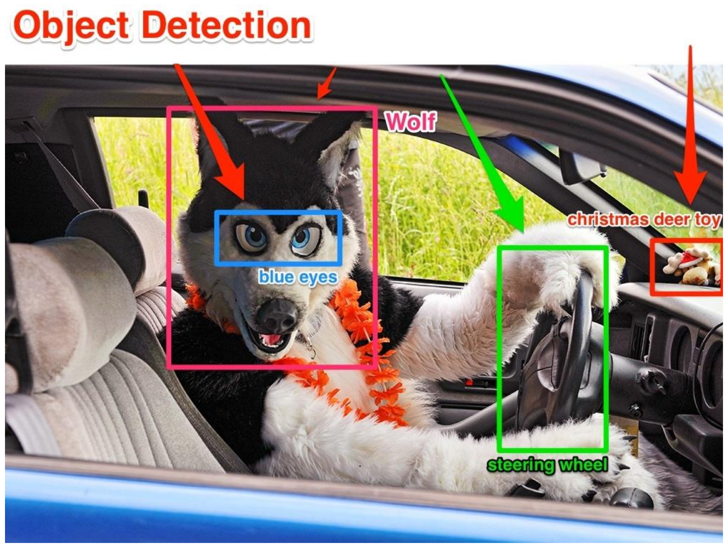
Outline

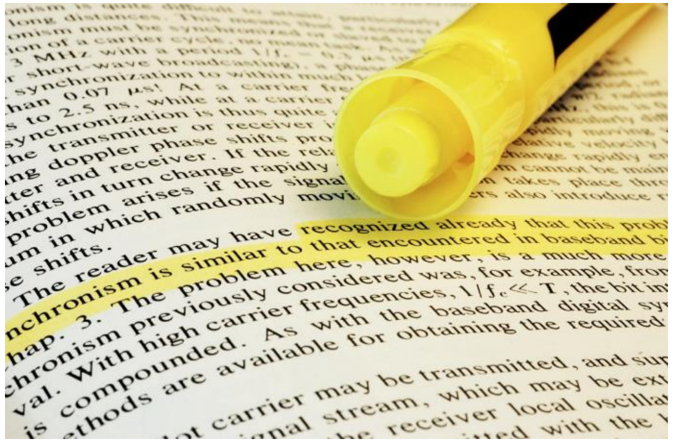
- Introduction.
- Machine learning paradigms.
  - Supervised learning
  - Unsupervised learning
  - Reinforcement learning
- Linear regression.
- Gradient decent.

Applications of machine learning



Database mining

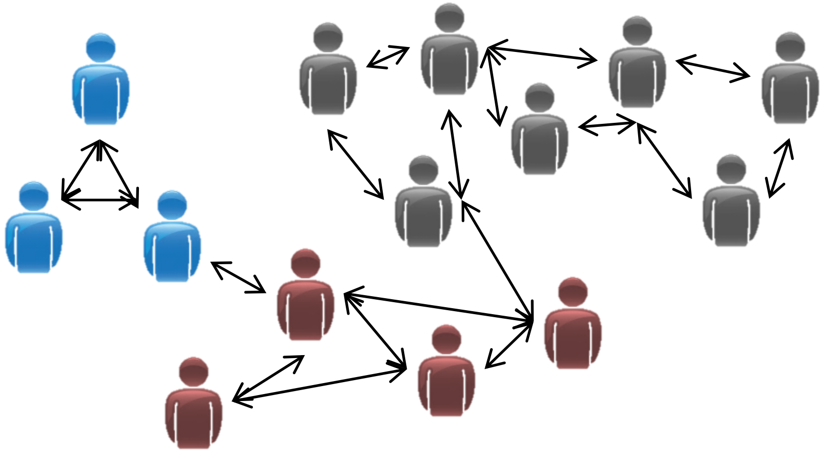




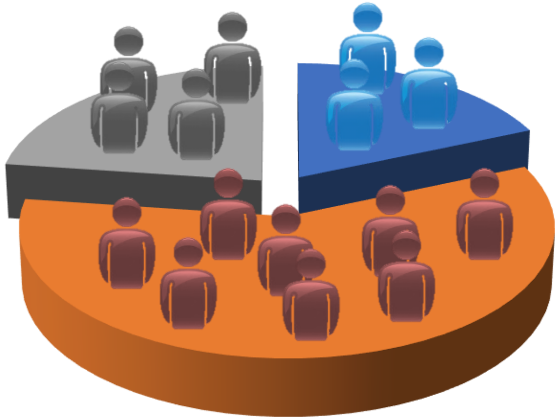
Text analysis

Speech Recognition

Astronomical data analysis



Social network analysis



Market segmentation

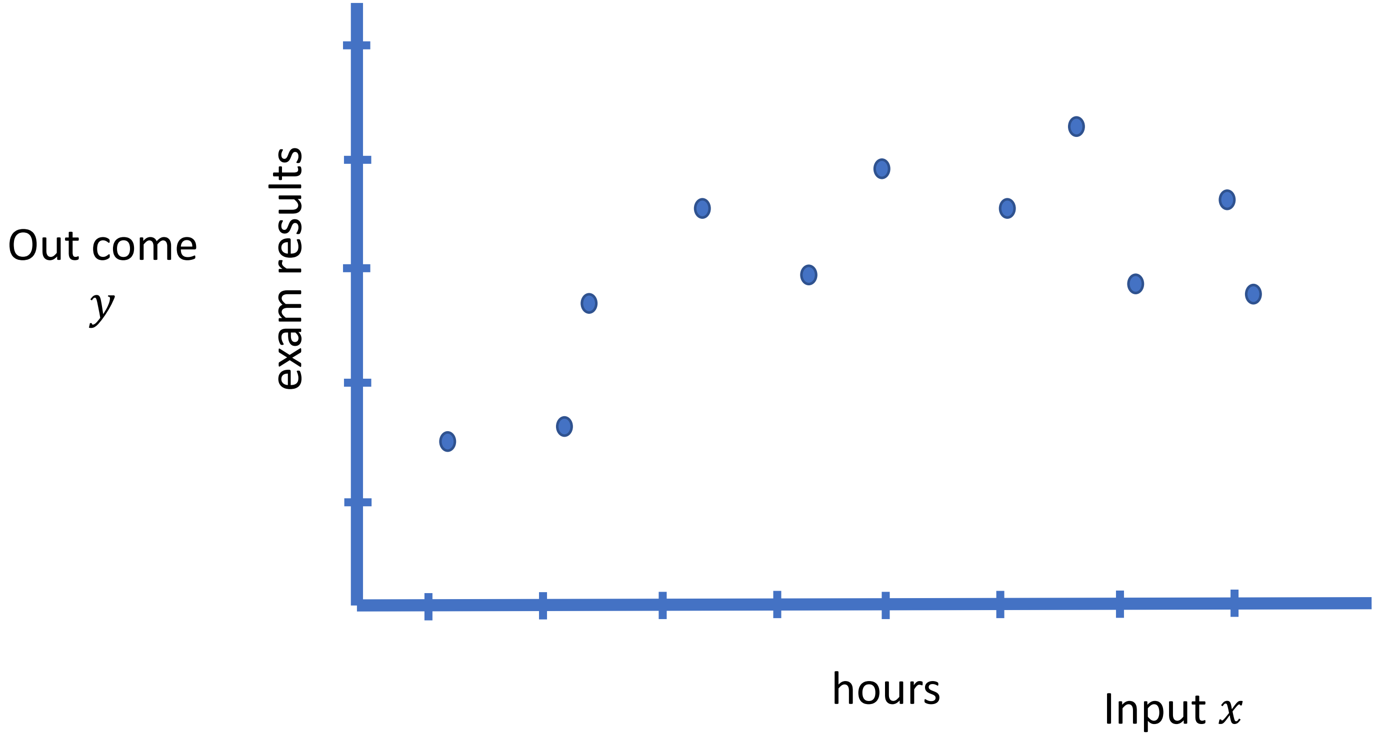
Genome classification

Machine learning algorithms

- Supervised learning
- Unsupervised learning
- Reinforcement learning

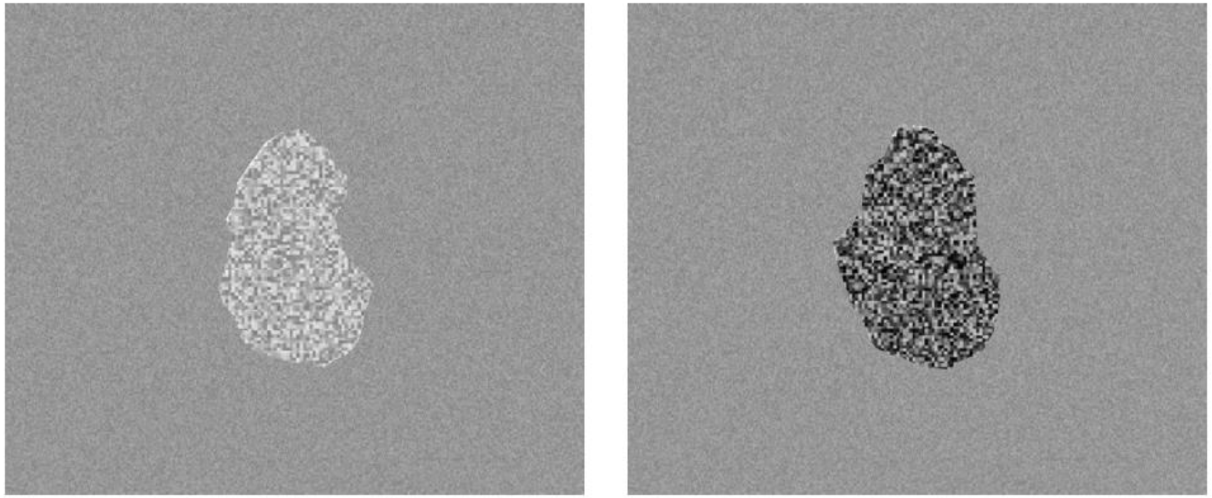
Supervised learning

- Supervised learning is the machine learning task of inferring a function from ***labelled*** training data.
- Predicting exam results

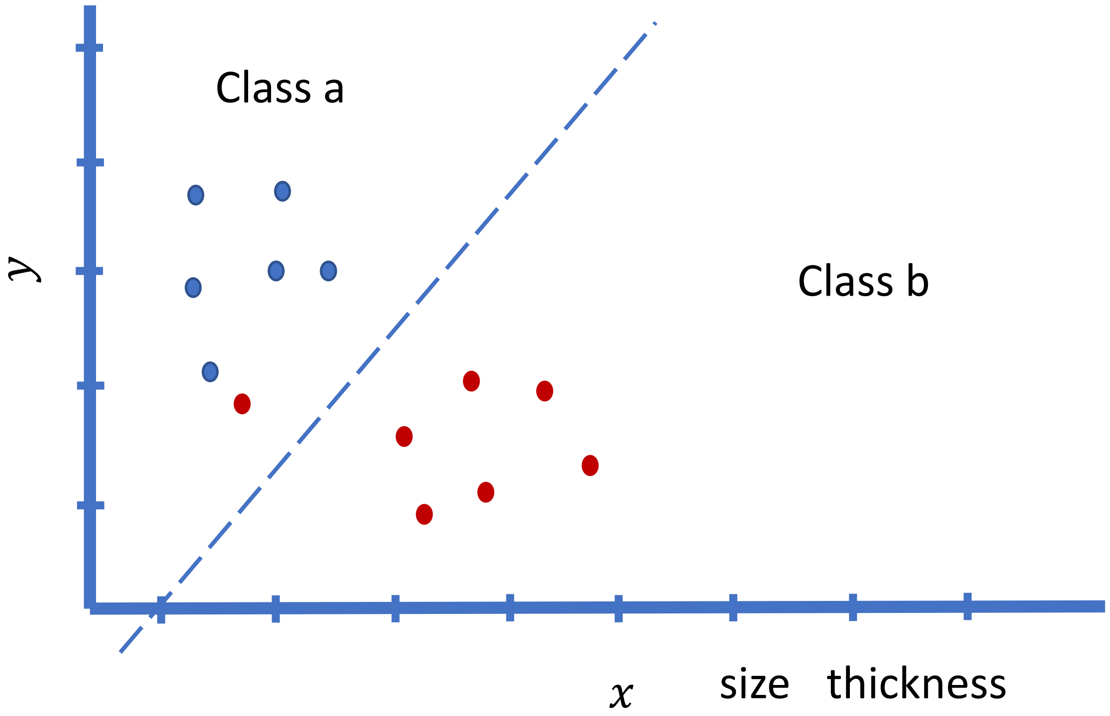


y

Real value



(a) (b)



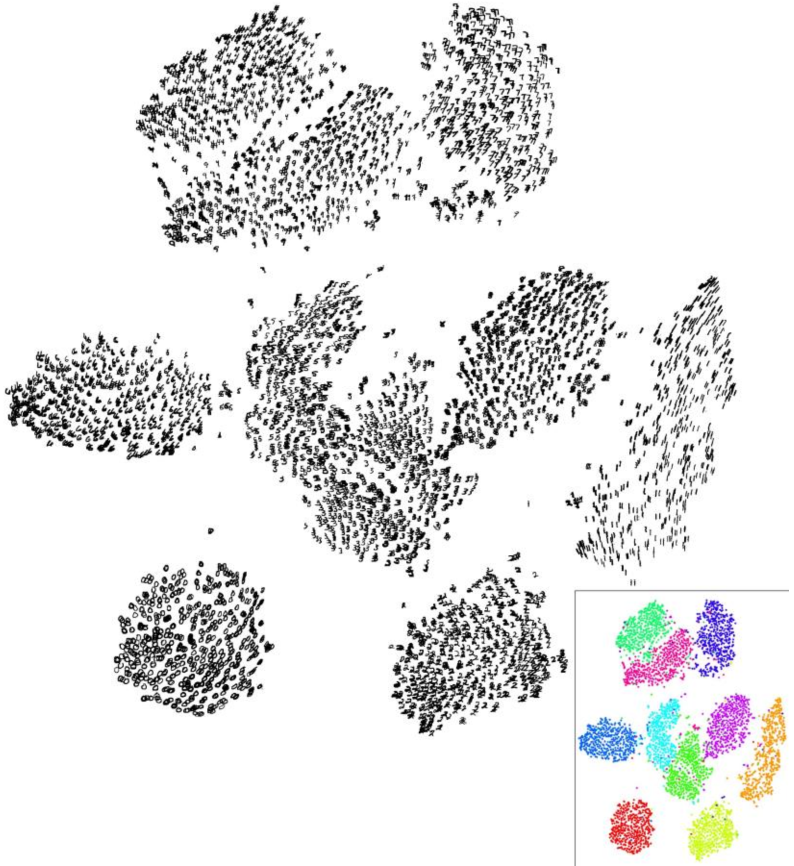
y

$\{0,1\}$

Binary value

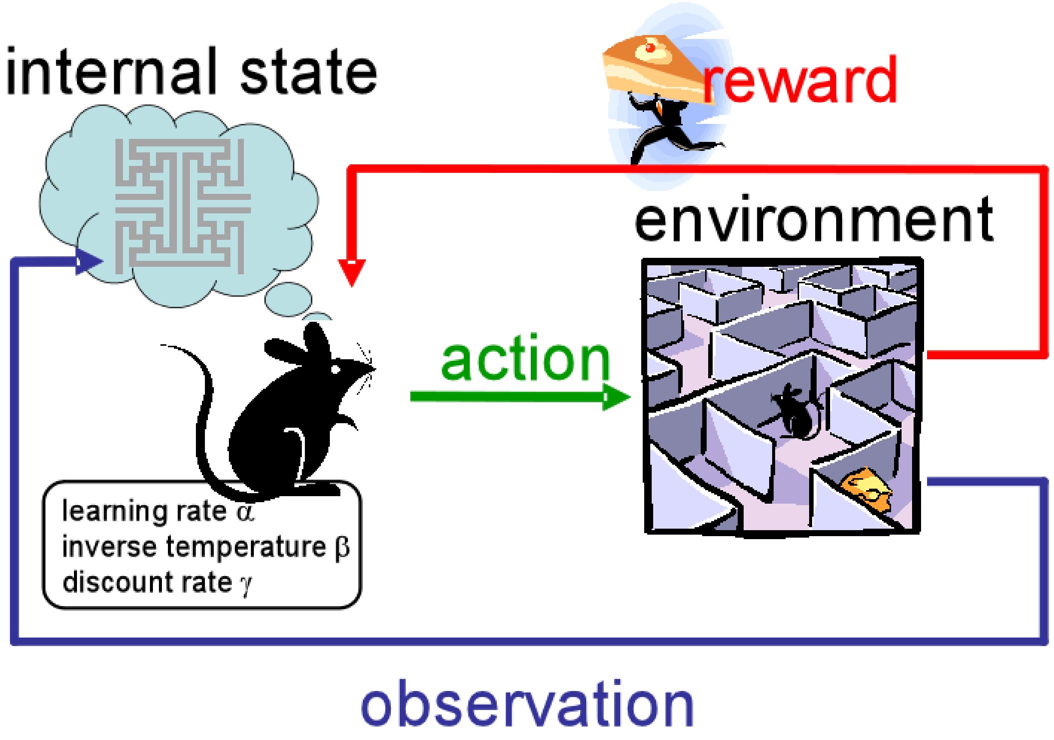
Unsupervised learning

- Unsupervised learning is the machine learning task of inferring a function to describe hidden structure from "unlabelled" data.



Reinforcement learning

- Is an area of machine learning inspired by behaviourist psychology, concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.



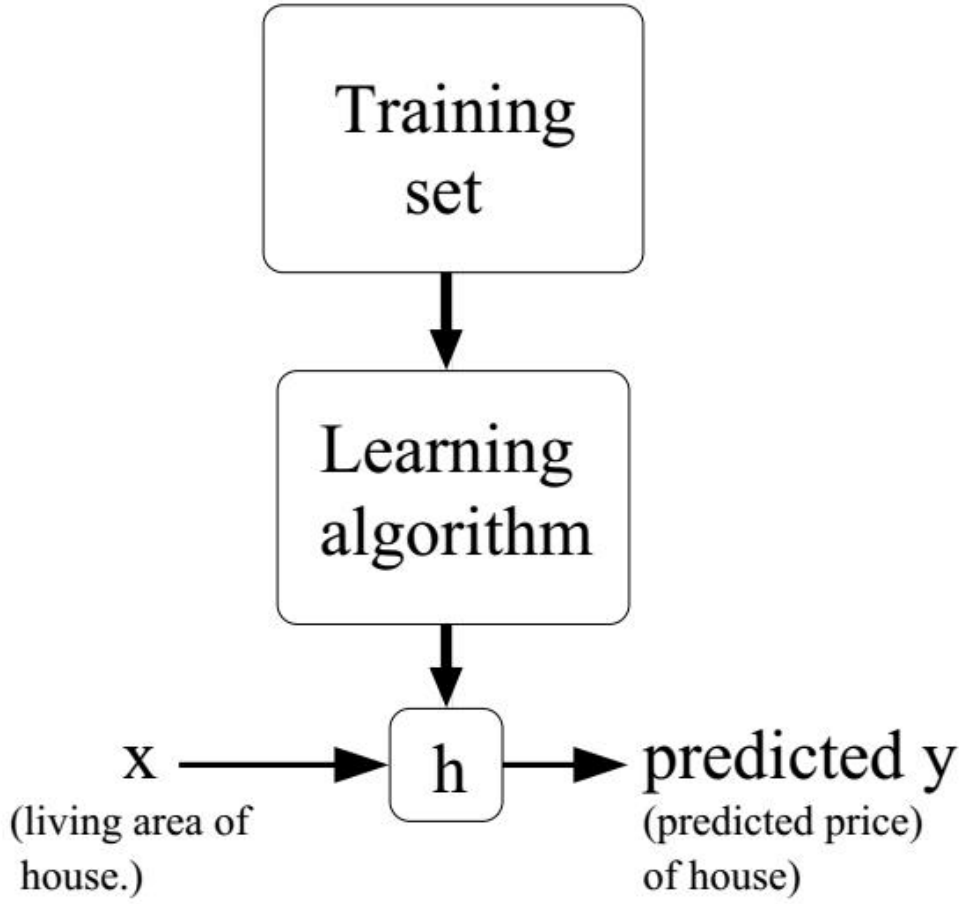
Machine learning basic terminology

$Xs\ variables\ \equiv features$

$\theta s\ \equiv model\ prameter, the\ constant\ we\ are\ learning\ them$

$h_\theta(x)\ model\ \equiv hypothesis$

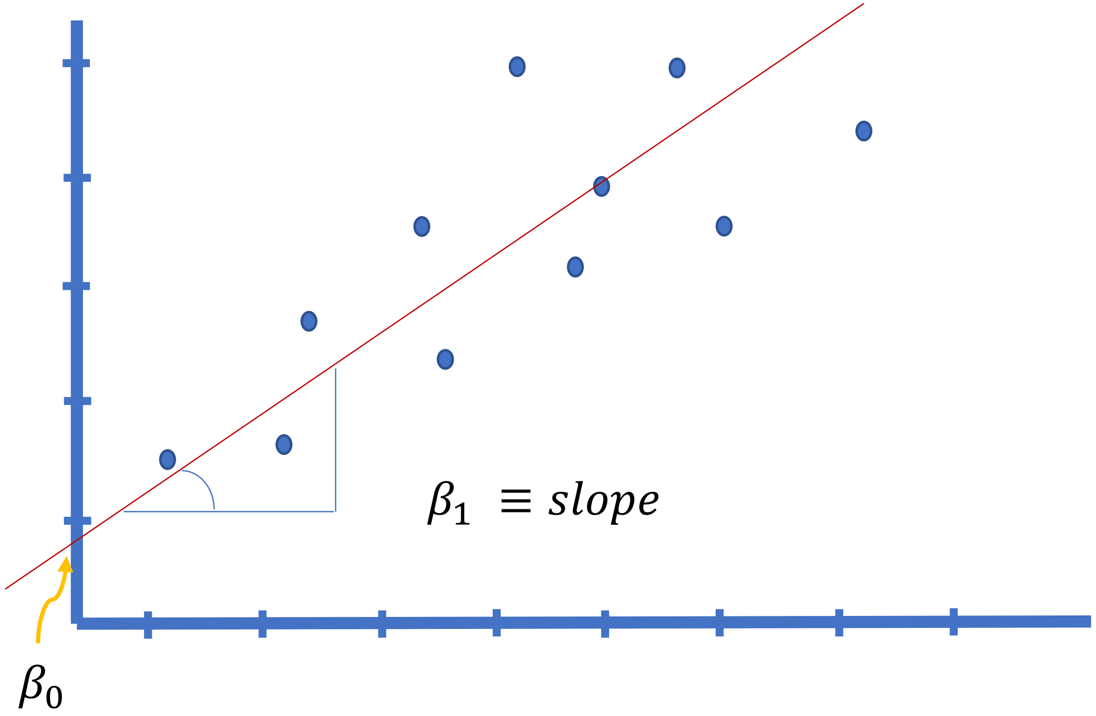
$Cost\ function\ J(\theta) \equiv the\ evaluation\ criteria$



Linear Regression

- The simplest version of regression is when x_i is simple (one-dimensional) and $h(x)$ is assumed to be linear.

$$h(x) = \beta_0 + \beta_1 x$$



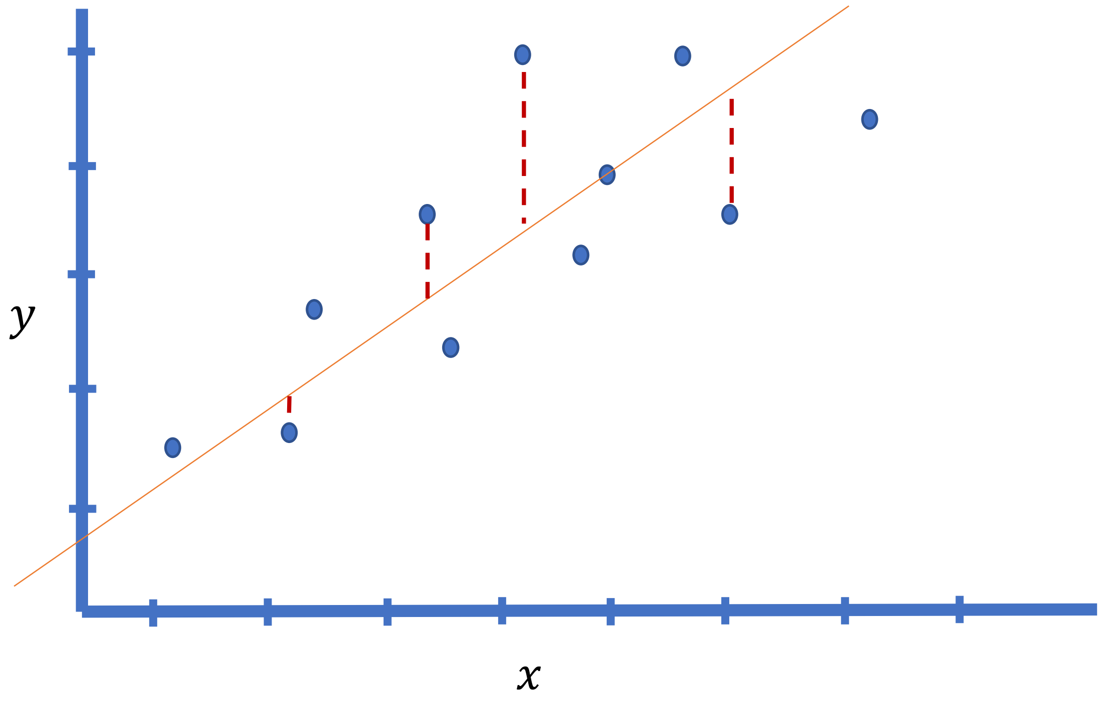
$$h_\theta(x) = \theta_0 + \theta_1 x$$

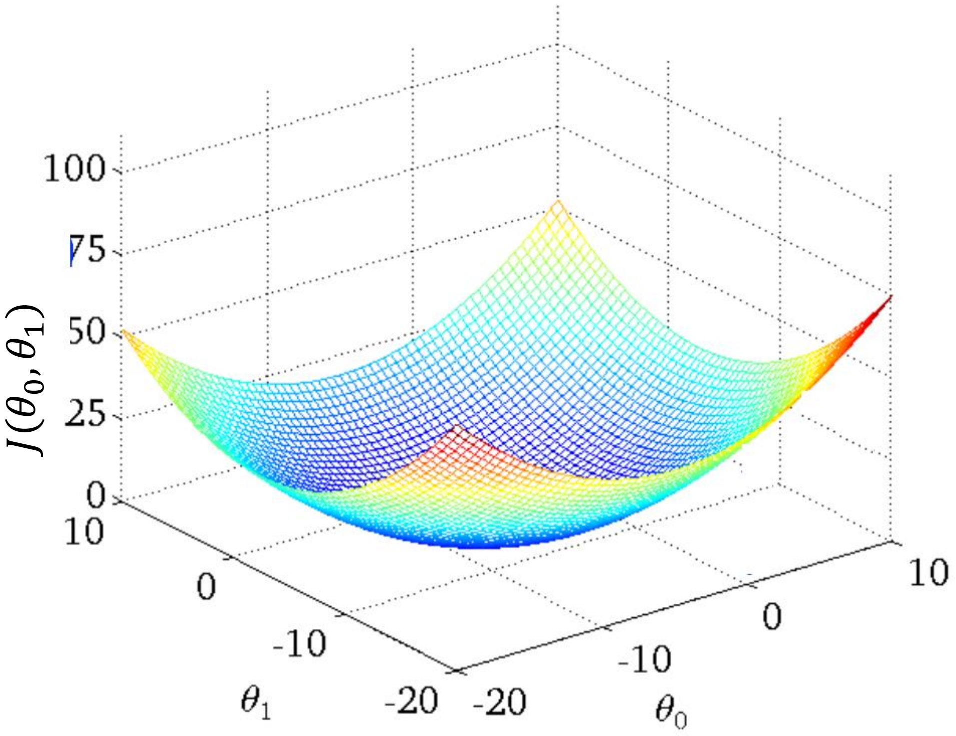
Cost function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$$

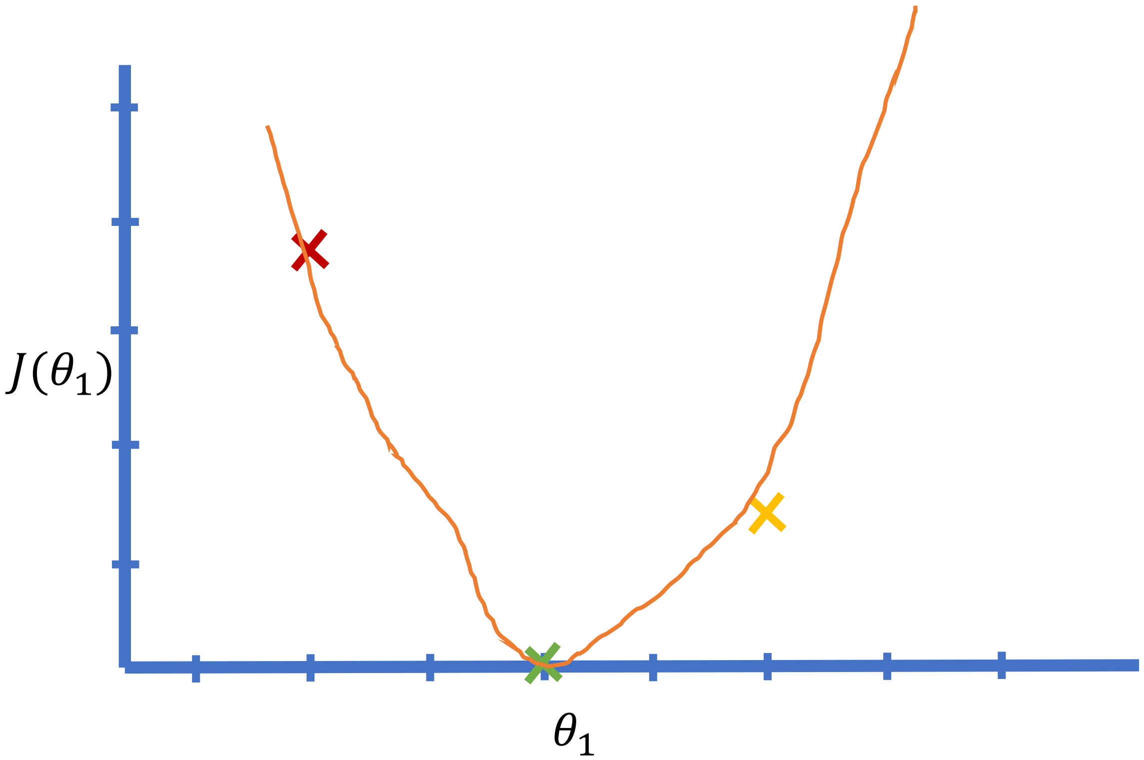
Goal:

$$Min_{\theta_0, \theta_1}: J(\theta_0, \theta_1)$$





100
75
50
25
0
$J(\theta_0, \theta_1)$
10
0
-10
-20
θ_1
-20
-10
0
10
θ_0



$J(\theta_1)$
θ_1

Summary

$Hypoytesies$: $h_\theta(x) = \theta_0 + \theta_1 x$

$Parameters$: θ_0, θ_1

Cost function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$$

Goal:

$$Min_{\theta_0, \theta_1}: J(\theta_0, \theta_1)$$

Gradient decent

- We have

$We\ have\ a\ function \quad J(\theta_0, \theta_1)$

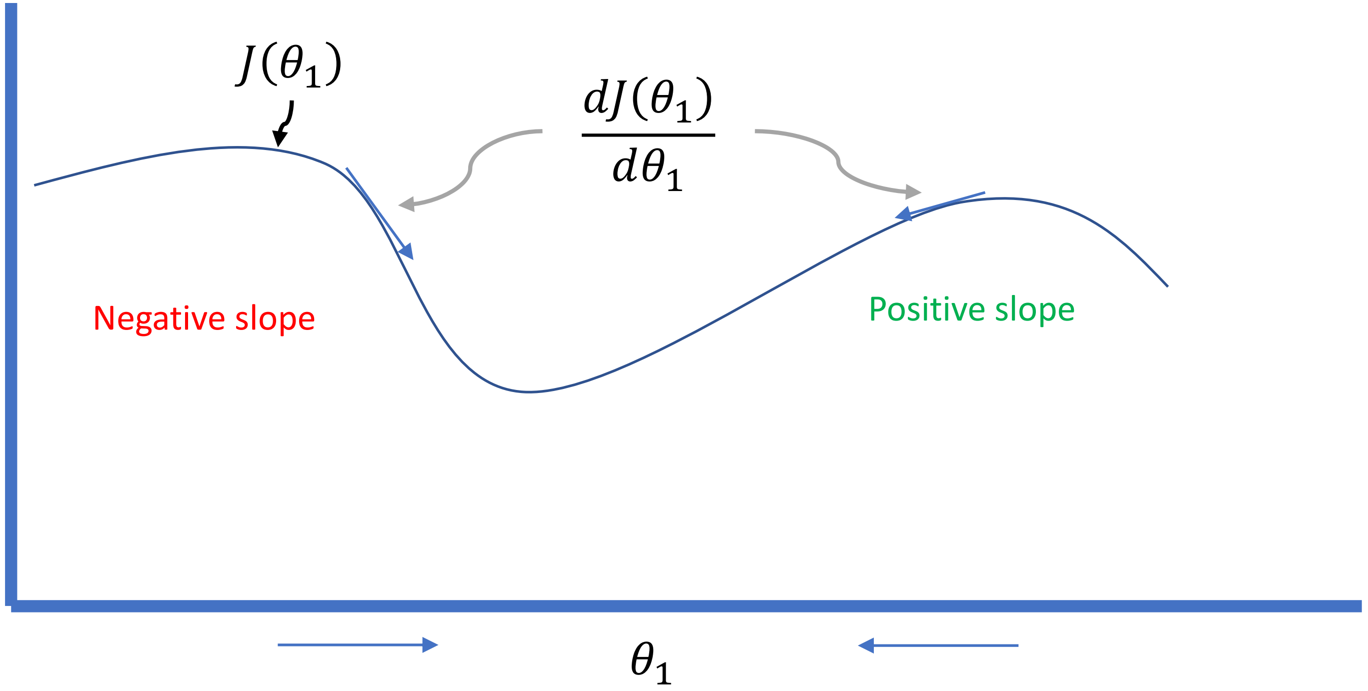
$We\ want \quad Min_{\theta_0,\theta_1}: J(\theta_0, \theta_1)$

- Outline:
  - Start with initial θ_0, θ_1
  - Keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$

Recall form calculus

$$\theta_i := \theta_i - \alpha \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_i}$$
$$(for\ i = 1{,}0)$$

As we approach a local minimum, gradient descent will automatically take smaller steps. So, no need to decrease α over time.



Gradient descent algorithm

$$repeat\ until\ convergence\ \{$$

$$\theta_i \coloneqq \theta_i \ - \ \alpha \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_i}$$

$$(for\ i = 1{,}0)$$

$$\}$$

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_i} =$$

$$i = 0 : \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_i} =$$

$$i = 1 : \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_i} =$$

Linear Regression Model

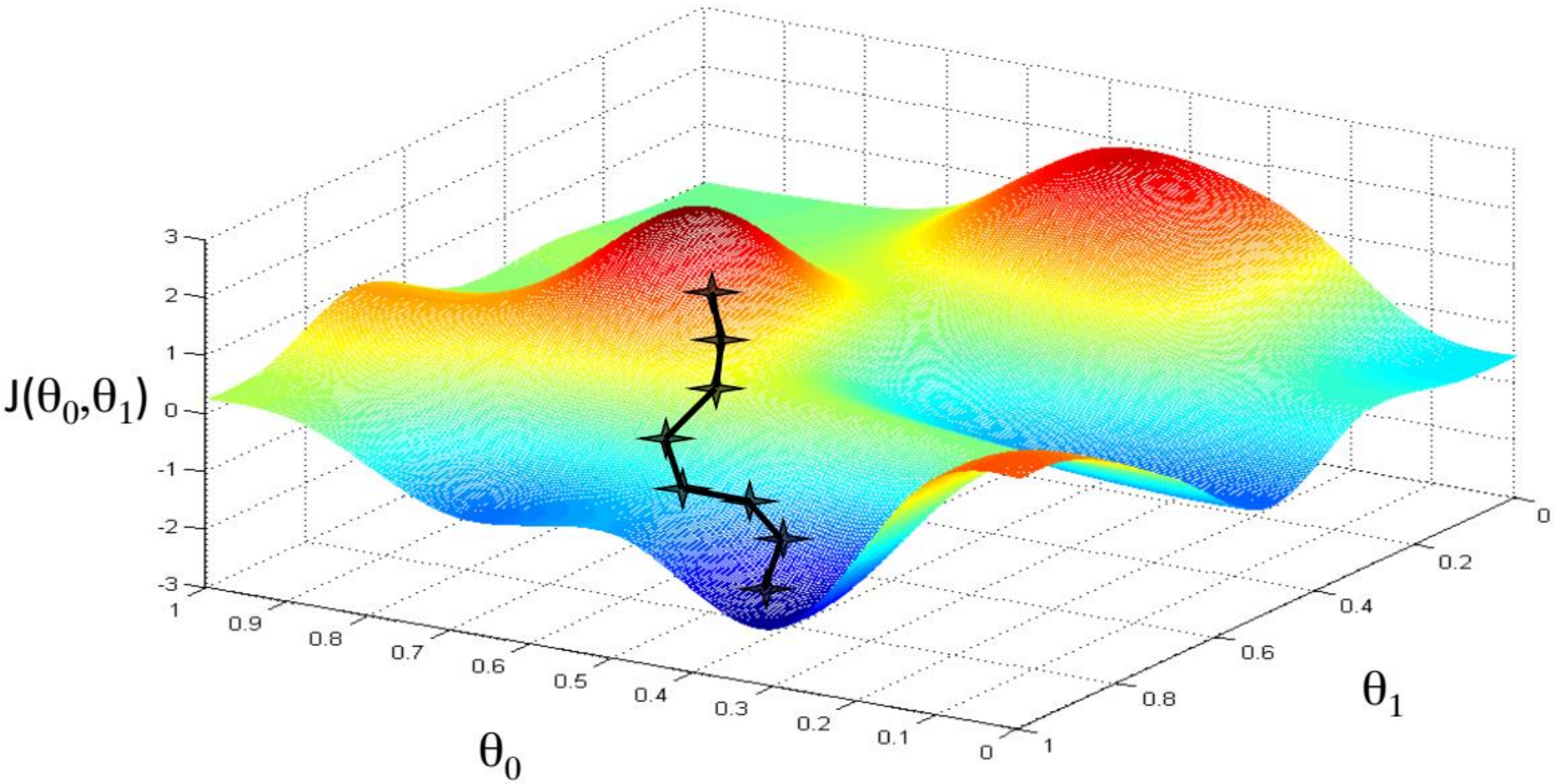
$$h_\theta(x) = \ \theta_0 + \ \theta_1 x$$

$$J(\theta_0, \theta_1) = \ \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) \ - y_i)^2$$

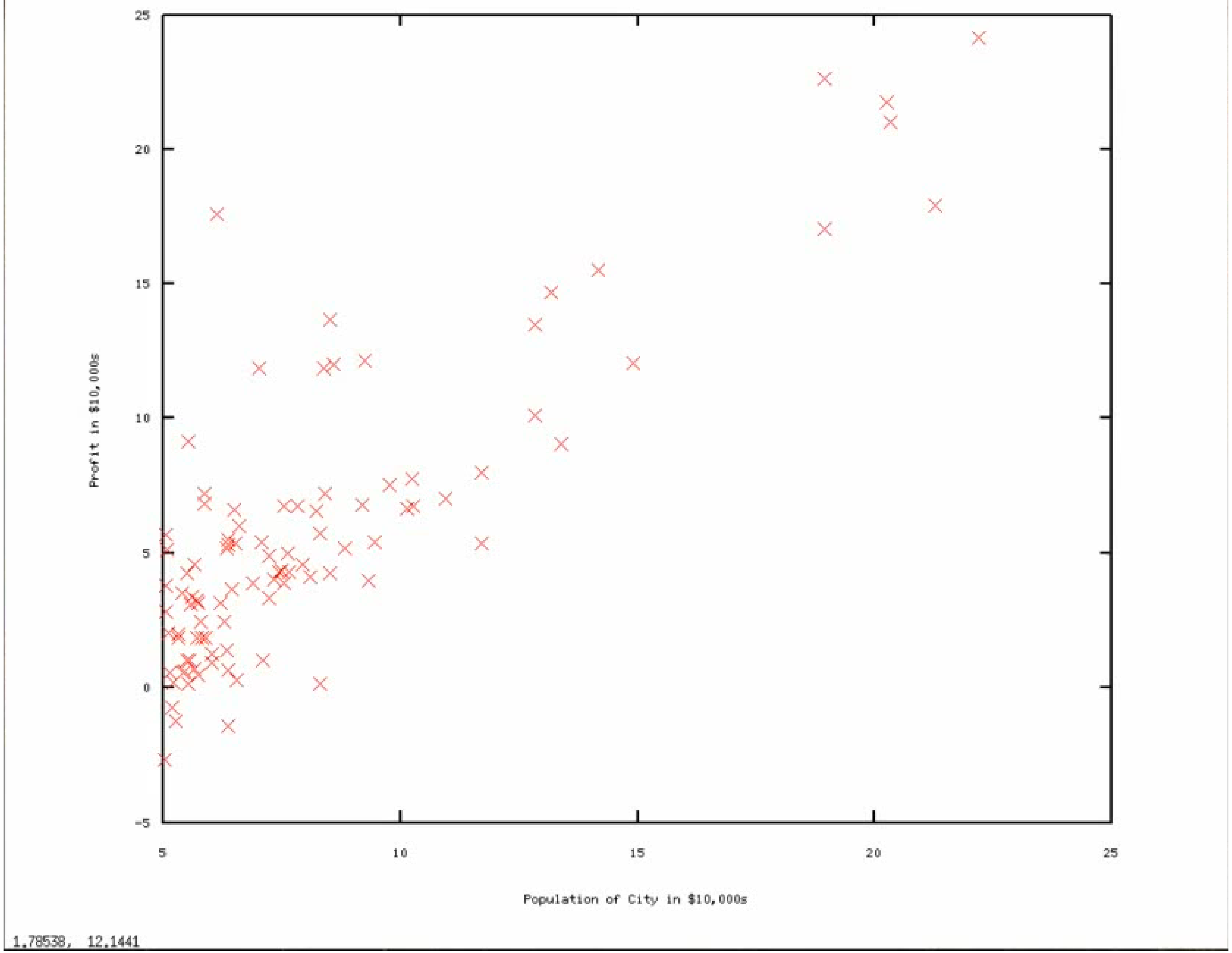
$$
\left.
\begin{aligned}
&\mathit{repeat\ until\ convergence}\ \{\\
&\theta_0 \coloneqq \theta_0 - \alpha \frac{1}{2m} \sum_{i=1}^{m} \big((\theta_0 + \theta_1 x) - y_i\big)\\
&\theta_1 \coloneqq \theta_1 - \alpha \frac{1}{2m} \sum_{i=1}^{m} \big((\theta_0 + \theta_1 x) - y_i\big)\, x_i\\
&\}
\end{aligned}
\right\}
\begin{array}{c}
\mathit{Update}\ \theta_0, \theta_1\\
\mathit{simultaneously}
\end{array}
$$

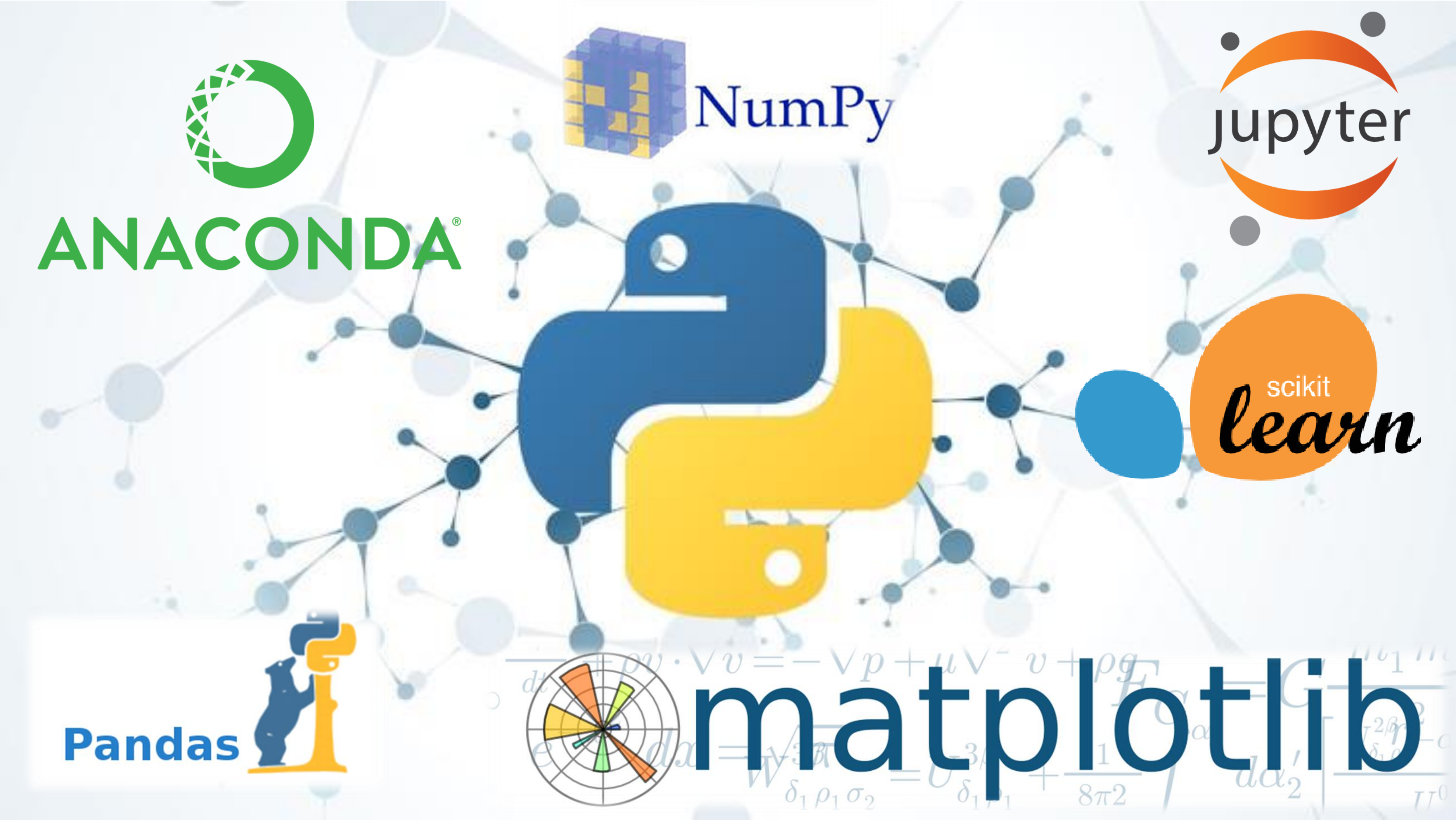
If α is too small, gradient descent can be slow.

If α is too large, gradient descent can overshoot the minimum. It may fail to converge, or even diverge.



J(θ0,θ1)
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ0
0
0.2
0.4
0.6
0.8
1
θ1


25
20
15
10
5
0
-5
Profit in $10,000s
5
10
15
20
25
Population of City in $10,000s
1.78538, 12.1441


ANACONDA
NumPy
jupyter
scikit
learn
Pandas
matplotlib