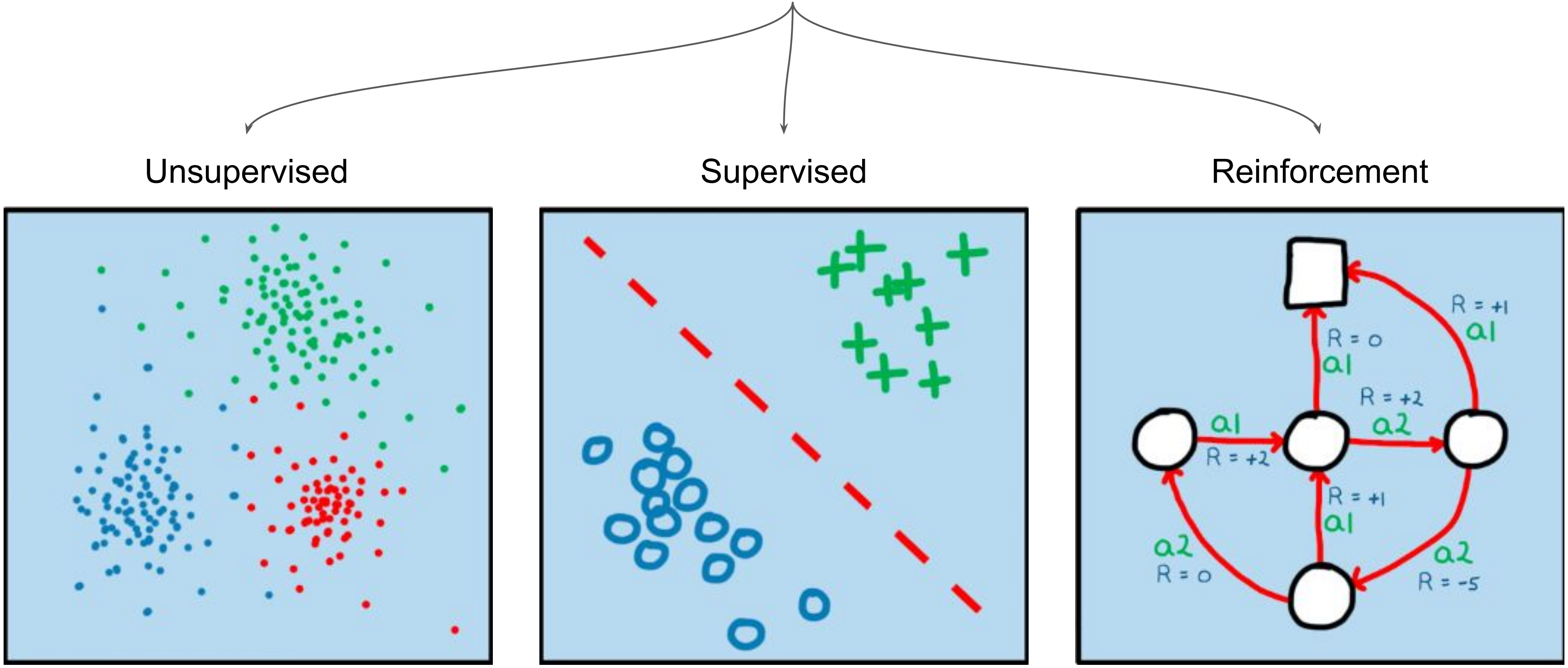
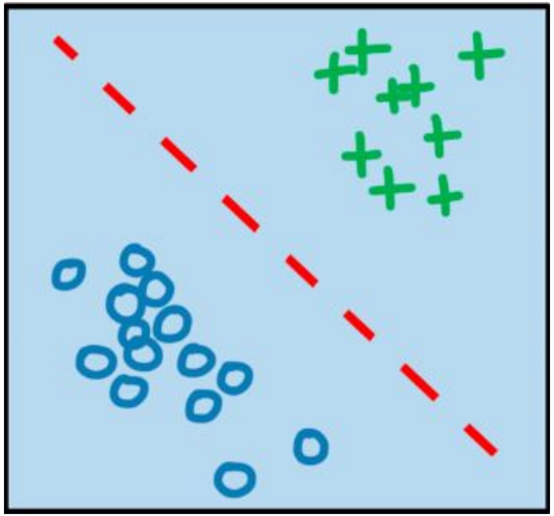
Machine Learning

Lesson 1

Machine Learning



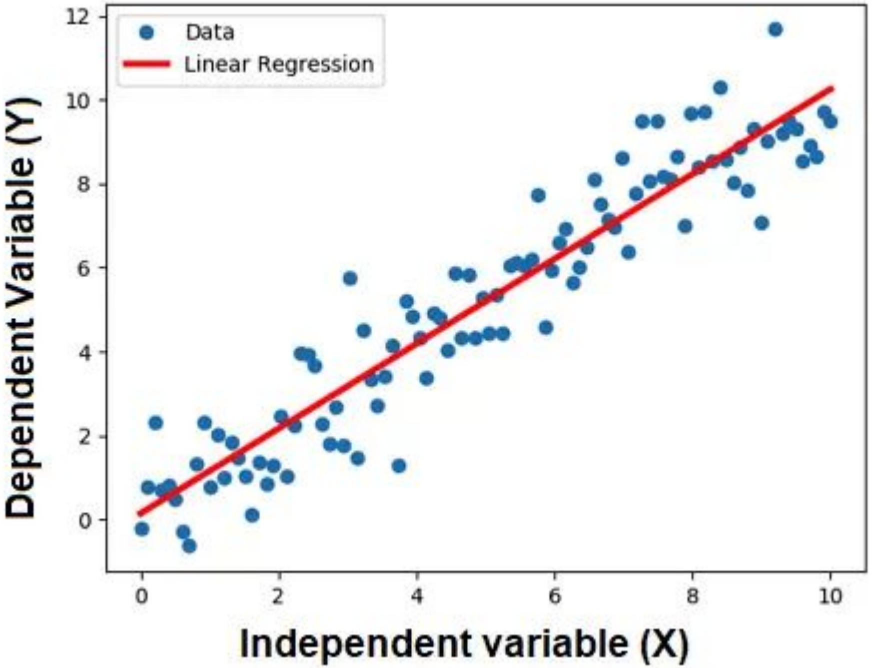
Supervised Learning



Regression

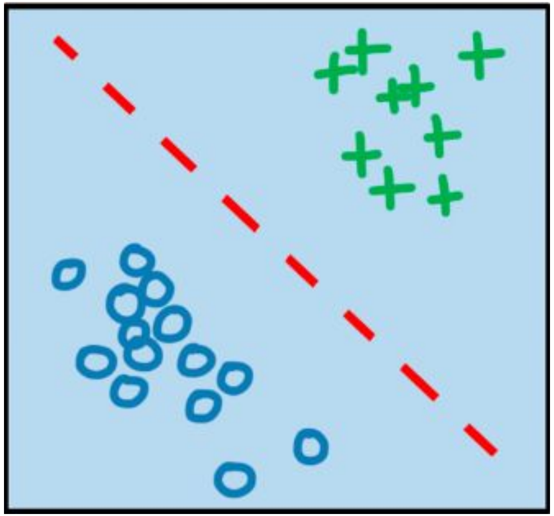
Classification

Supervised Learning



X	Y
1	2
3	4
3	5
6	7
7	7
3	4
⋮	⋮

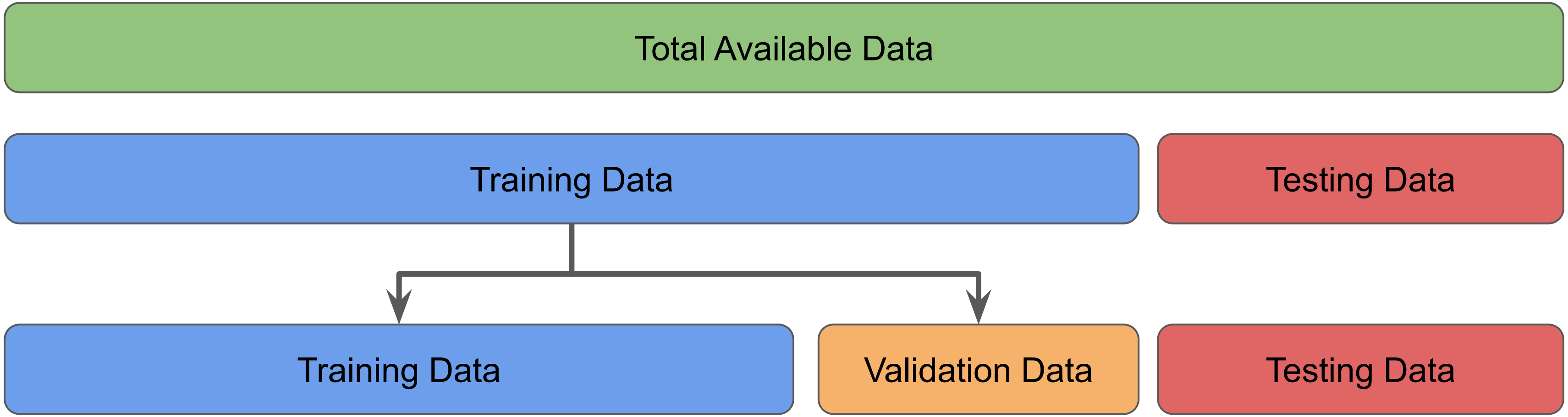
Supervised Learning



x1	x2
1	2
3	4
3	2
6	3
7	4
3	1
. . .	. . .

Label
O
O
X
X
X
O
. . .

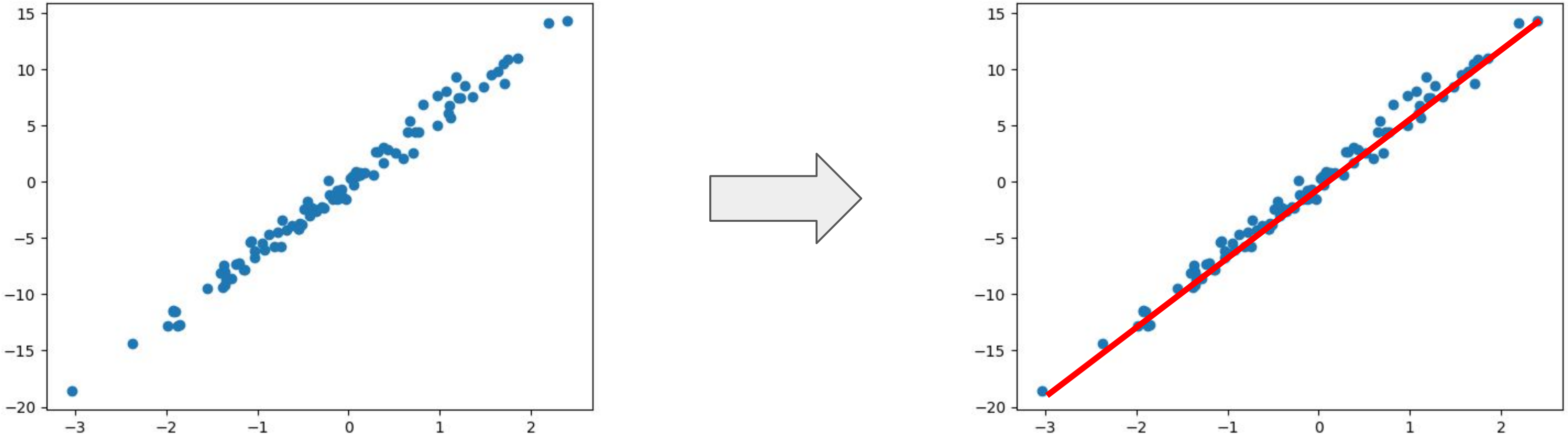
Dataset


Total Available Data
Training Data
Testing Data
Training Data
Validation Data
Testing Data

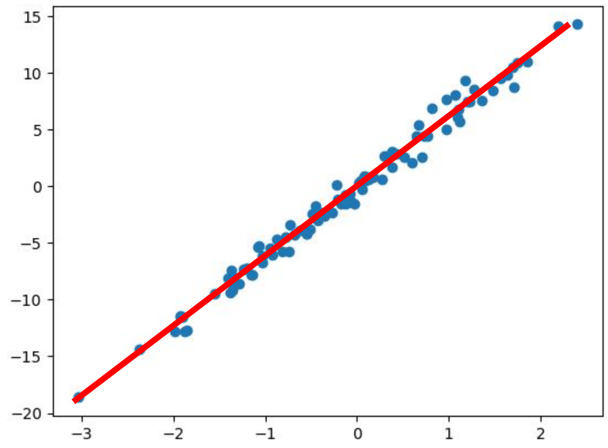
Supervised Machine Learning Algorithms

- Nearest Neighbor
- Naive Bayes
- Decision Trees
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVM)
- Neural Networks

Linear Regression



Linear Regression



$$y = \theta x + b$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y_i} - y)^2$$

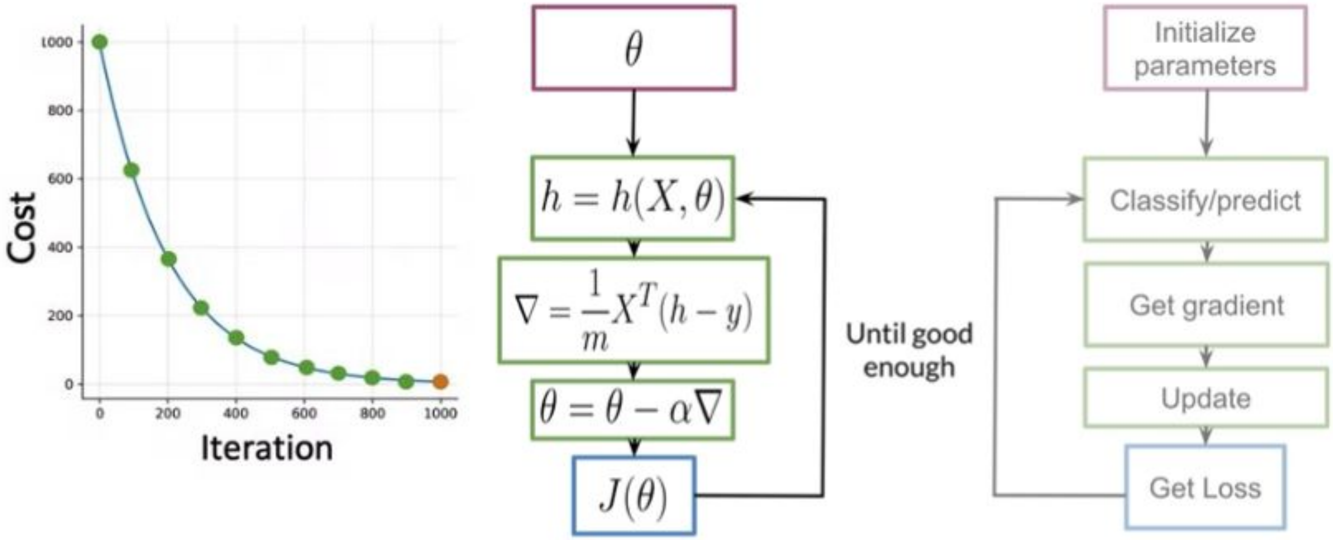
$$L(\theta, b) = \frac{1}{N}((x\theta + b) - y)^2$$

$$\frac{\mathrm{d}L}{\mathrm{d}\theta} = \frac{1}{N}x^T((x\theta + b) - y)$$

$$\frac{\mathrm{d}L}{\mathrm{d}b} = \frac{1}{N}((x\theta + b) - y)$$

Linear Regression: Training

Usually you keep training until the cost converges. If you were to plot the number of iterations versus the cost, you should see something like this:



You initialize your parameter **θ**, that you can use in your equation, you then compute the gradient that you will use to update **θ**, and then calculate the cost. You keep doing so until good enough.

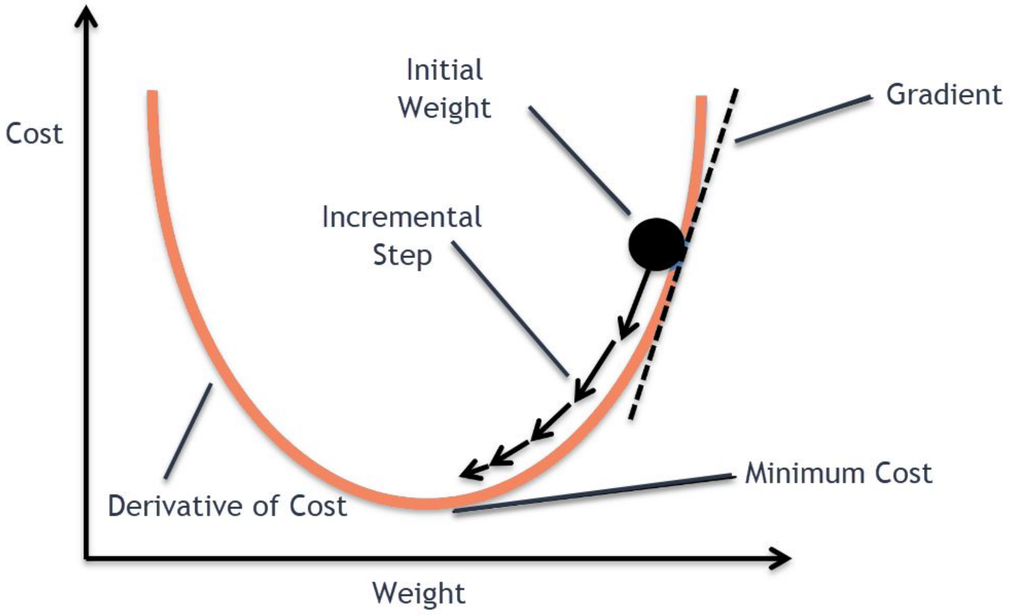
Linear Regression : Gradient Descent

$$\min_{\theta} J(\theta)$$

$$\theta \leftarrow \theta - \alpha \frac{\mathrm{d}}{\mathrm{d}\theta} J(\theta)$$

$$\frac{\mathrm{d}}{\mathrm{d}\theta} J(\theta) = (h(x, \theta) - y)\, x$$

$$\theta \leftarrow \theta - \alpha(h(x, \theta) - y)x$$



α → learning rate