

From Regression to Deep Learning

Practice **LESS** Deep Learning
Learn - Experiment - Share - Seek

Barathi Ganesh HB

Centre for Excellence in Computational Engineering and Networking (CEN)
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
email: barathiganesh.hb@gmail.com



Outline

ML Introduction

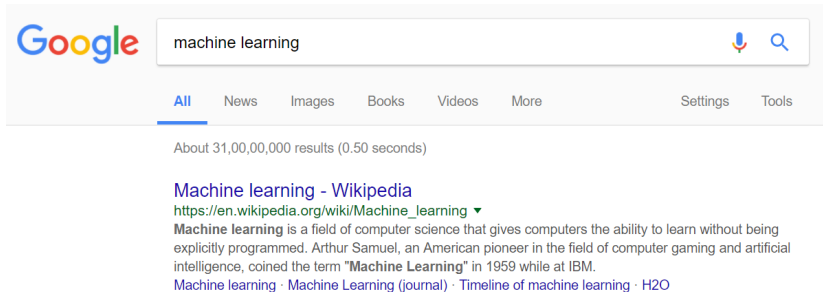
Regression to Deep Learning

Need of Deep Learning

Matrix Representation



Machine Learning Introduction



Google

machine learning

All News Images Books Videos More Settings Tools

About 31,00,00,000 results (0.50 seconds)

Machine learning - Wikipedia
https://en.wikipedia.org/wiki/Machine_learning ▼

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed. Arthur Samuel, an American pioneer in the field of computer gaming and artificial intelligence, coined the term "**Machine Learning**" in 1959 while at IBM.

[Machine learning](#) · [Machine Learning \(journal\)](#) · [Timeline of machine learning](#) · [H2O](#)

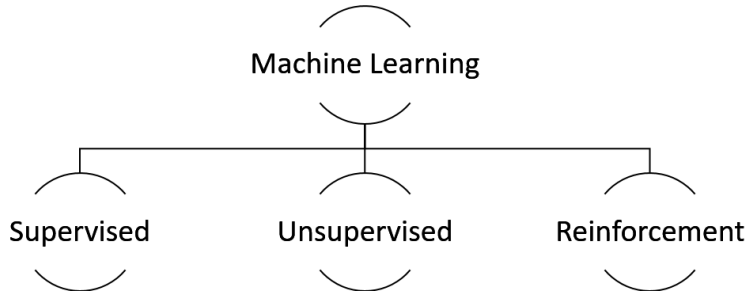


- Reducing human/machine efforts required to perform a task (time optimization).
- Increasing the performance of a task (efficiency optimization).

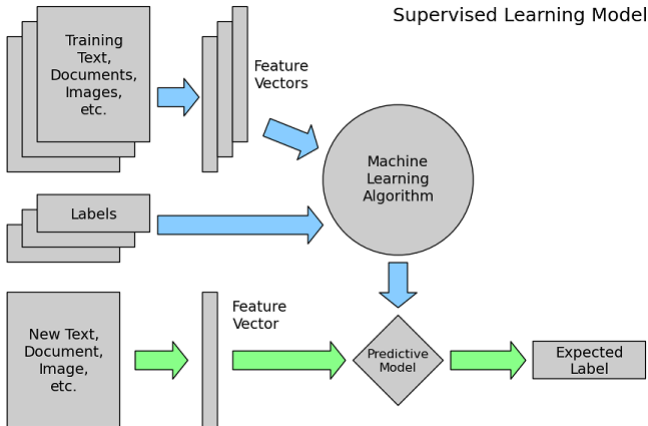


Steps in Machine Learning

- Collecting data
- Preparing the data
- Training a model
- Evaluating the model
- Improving the performance



Supervised Learning



source: www.allprogrammingtutorials.com/tutorials/introduction-to-machine-learning.php



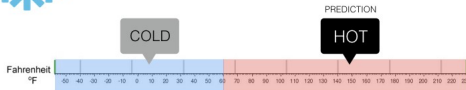
Regression

What is the temperature going to be tomorrow?



Classification

Will it be Cold or Hot tomorrow?



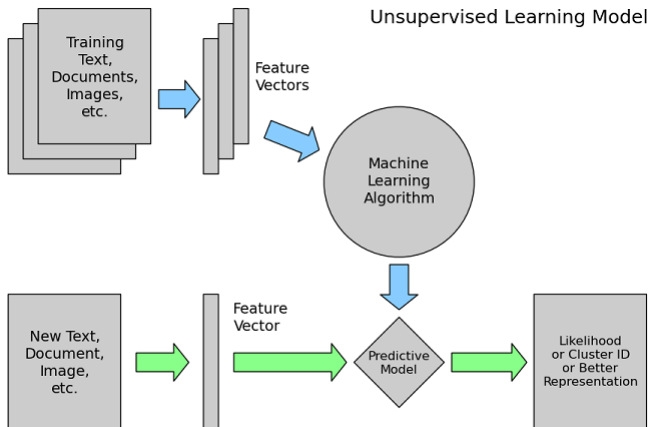
source: https://medium.com/@ali_88273/regression-vs-classification-87c224350d69



Common Supervised Learning Algorithms

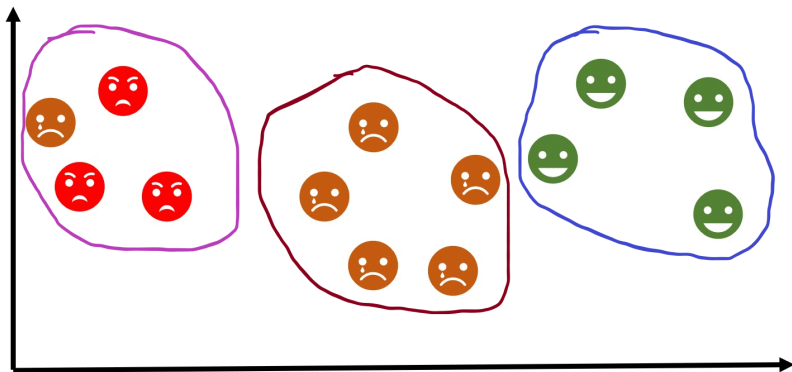
- Linear Regression
- Logistic Regression
- Support Vector Machines
- Support Vector Regression
- Decision Trees
- Random Forest Tree
- Naive Bayes

Unsupervised Learning



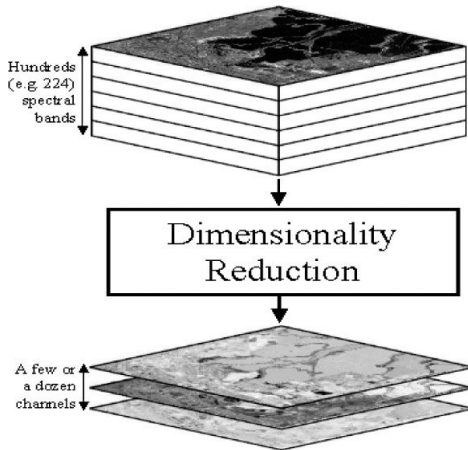
source: www.allprogrammingtutorials.com/tutorials/introduction-to-machine-learning.php

Clustering



source: <https://towardsdatascience.com/clustering-unsupervised-learning-788b215b074b>

Dimensionality Reduction



source: <http://spie.org/newsroom/3560-dimensionality-reduction-of-multidimensional-satellite-imagery?SSO=1>



Common Unsupervised Learning Algorithms

- K-means
- Affinity Propagation
- Singular Value Decomposition
- Non-negative matrix factorization



You have already learned the Machine Learning. When?

?



You have already learned the Machine Learning. When?

$$2x = 6 \quad (1)$$

$$(2x - 6) = 0 \quad (2)$$

$$x = ? \quad (3)$$

You have already learned the Machine Learning. When?

$$2x = 6 \quad (4)$$

$$(2x - 6) = 0 \quad (5)$$

$$x = ? \quad (6)$$

$$x = 6/2 = 3 \quad (7)$$

$$2(3) - 6 = 0 \quad (8)$$



You have already learned the Machine Learning. When?

$$2a + b + c = 4 \quad (9)$$

$$a + 3b + 2c = 5 \quad (10)$$

$$a = 6 \quad (11)$$

You have already learned the Machine Learning. When?

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \quad (12)$$

You have already learned the Machine Learning. When?

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 0 & 0 \end{bmatrix}, x = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \quad (13)$$

You have already learned the Machine Learning. When?

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \quad (14)$$

$$Ax = b \quad (15)$$

$$(Ax - b) = ? \quad (16)$$

You have already learned the Machine Learning. When?

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \quad (17)$$

$$Ax = b \quad (18)$$

$$(Ax - b) = 0 \quad (19)$$

$$x = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = ? \quad (20)$$



What is Regression

Regression?



What is Regression

$$x + y = z$$

Solving $Ax=b$

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \quad (21)$$

$$X W = Y \quad (22)$$

$$(X W - Y) = 0 \quad (23)$$

$$W = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = ? \quad (24)$$

$$X^\dagger X W = X^\dagger Y \quad (25)$$

$$I W = X^\dagger Y \quad (26)$$

$$W = X^\dagger Y \quad (27)$$

Decimal Value Prediction

ID	digit1	digit2	digit3	value
1	0	0	0	0
2	0	0	1	1
3	0	1	0	2
4	0	1	1	3
5	1	0	0	4
6	1	0	1	5
7	1	1	0	6
8	1	1	1	7

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} w1 \\ w2 \\ w3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{bmatrix} \quad (28)$$

$$X \mathbf{w} = \mathbf{y} \quad (29)$$

$$X^\dagger X \mathbf{w} = X^\dagger \mathbf{y} \quad (30)$$

$$\mathbf{w} = X^\dagger \mathbf{y} \quad (31)$$

$$\mathbf{w} = X^\dagger \mathbf{y} = X^\dagger \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \quad (32)$$

$$X \mathbf{w} = \mathbf{y} \quad (33)$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{bmatrix} \quad (34)$$

$$X \mathbf{w} = \mathbf{y}^{pre} \quad (35)$$

$$training \text{ error} = abs(\mathbf{y} - \mathbf{y}^{pre}) \quad (36)$$

$$\text{training error} = \text{sum}(\text{abs}(\mathbf{y} - \mathbf{y}^{pre})) \quad (37)$$

$$\mathbf{y} - \mathbf{y}^{pre} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{bmatrix} = \text{sum} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 0 \quad (38)$$

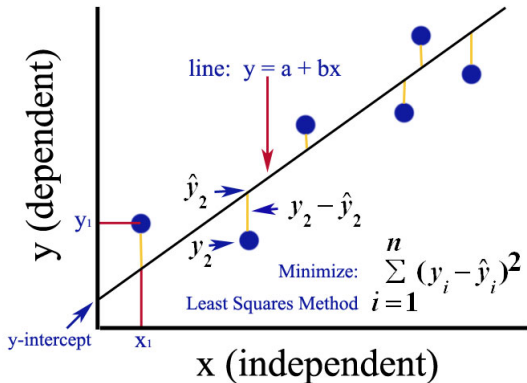
$$\begin{bmatrix} digit1 & digit2 & digit3 \end{bmatrix} \begin{bmatrix} w1 \\ w2 \\ w3 \end{bmatrix} = [value] \quad (39)$$

$$digit1 * w1 + digit2 * w2 + digit3 * w3 = value \quad (40)$$

$$\begin{bmatrix} digit1 & digit2 & digit3 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} = [value] \quad (41)$$

$$digit1 * 4 + digit2 * 2 + digit3 * 1 = value \quad (42)$$

Linear Regression



source: solutions4statistics.com

Decimal Value Prediction

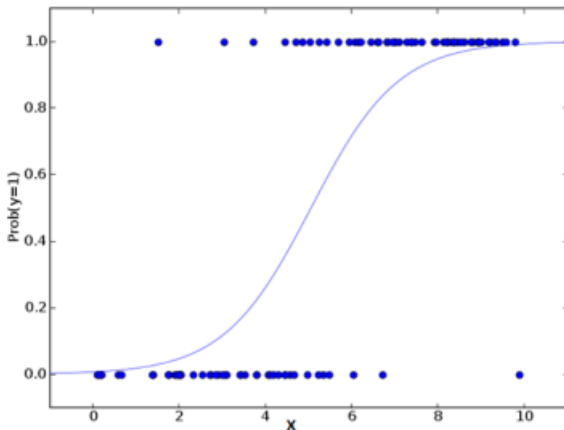
ID	digit1	digit2	digit3	value	decision
1	0	0	0	0	0
2	0	0	1	1	0
3	0	1	0	2	0
4	0	1	1	3	0
5	1	0	0	4	1
6	1	0	1	5	1
7	1	1	0	6	1
8	1	1	1	7	1

$$\begin{bmatrix} digit1 & digit2 & digit3 \end{bmatrix} \begin{bmatrix} w1 \\ w2 \\ w3 \end{bmatrix} = [value] \quad (43)$$

$$digit1 * w1 + digit2 * w2 + digit3 * w3 = value \quad (44)$$

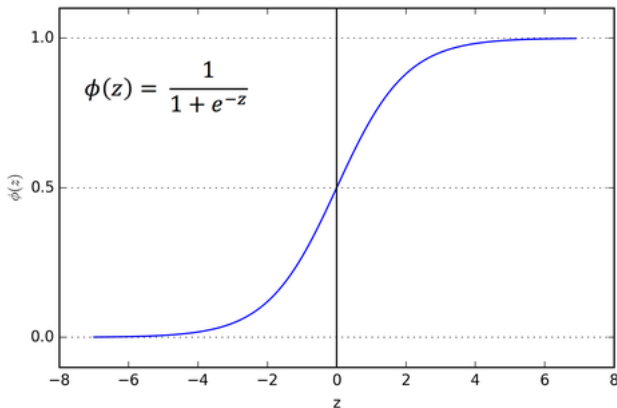
$$Prediction = \begin{cases} 1 & \text{if } 4 \geq value \\ 0 & \text{else} \end{cases} \quad (45)$$

Logistic Regression



source: solutions4statistics.com

Logistic - Sigmoid Function



<https://sebastianraschka.com/images/faq/logisticregr-neuralnet/sigmoid.png>

Logistic - Sigmoid

$$\Phi(z) = \frac{1}{1 + \exp^{-z}} \quad (46)$$

$$\Phi(-6) = \frac{1}{1 + \exp^{-(-6)}} = \frac{1}{1 + 403.42} = 0.0024 \quad (47)$$

$$\Phi(0) = \frac{1}{1 + \exp^0} = \frac{1}{1 + 1} = 0.5 \quad (48)$$

$$\Phi(6) = \frac{1}{1 + \exp^{-(6)}} = \frac{1}{1 + 0.0024} = 0.997 \quad (49)$$



Logistic Regression

$$\Phi(z) = \frac{1}{1 + \exp^{-z}} \quad (50)$$

$$\mathbf{y} = \Phi(\mathbf{X} \mathbf{w}) = \frac{1}{1 + \exp^{-(\mathbf{X} \mathbf{w})}} \quad (51)$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} w1 \\ w2 \\ w3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad (52)$$

$$X \mathbf{w} = \mathbf{y} \quad (53)$$

$$X^\dagger X \mathbf{w} = X^\dagger \mathbf{y} \quad (54)$$

$$\mathbf{w} = X^\dagger \mathbf{y} \quad (55)$$

$$\mathbf{w} = X^\dagger \mathbf{y} = X^\dagger \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.24054754 \\ -0.11269202 \\ -0.11269202 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \quad (56)$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1.24054754 \\ -0.11269202 \\ -0.11269202 \end{bmatrix} = \text{sigmoid} \left(\begin{bmatrix} 0.0 \\ -0.11269202 \\ -0.11269202 \\ -0.22538404 \\ 1.24054754 \\ 1.12785552 \\ 1.12785552 \\ 1.0151635 \end{bmatrix} \right) \quad (57)$$

$$X \mathbf{w} = \text{sigmoid}(\mathbf{y}^{pre}) \quad (58)$$

$$\text{sigmoid} \left(\begin{bmatrix} 0.0 \\ -0.11269202 \\ -0.11269202 \\ -0.22538404 \\ 1.24054754 \\ 1.12785552 \\ 1.12785552 \\ 1.0151635 \end{bmatrix} \right) = \begin{bmatrix} 0.5 \\ 0.47185 \\ 0.47185 \\ 0.44389 \\ 0.77565 \\ 0.75544 \\ 0.75544 \\ 0.73402 \end{bmatrix} \quad (59)$$

$$X \mathbf{w} = \text{sigmoid}(\mathbf{y}^{pre}) \quad (60)$$

$$\text{training error} = \text{sum}(\text{abs}(\mathbf{y} - \text{sigmoid}(\mathbf{y}^{pre}))) \quad (61)$$

$$\text{training error} = \text{sum}(\text{abs}(\mathbf{y} - \text{sigmoid}(\mathbf{y}^{\text{pre}}))) \quad (62)$$

$$\text{abs}(\mathbf{y} - \text{sigmoid}(\mathbf{y}^{\text{pre}})) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (63)$$

$$\text{training error} = 0 \quad (64)$$

$$\begin{bmatrix} 0.5 \\ 0.47185 \\ 0.47185 \\ 0.44389 \\ 0.77565 \\ 0.75544 \\ 0.75544 \\ 0.73402 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad (65)$$

$$Prediction = \begin{cases} 1 & \text{if } \text{sigmoid}(\mathbf{y}^{pre}) \geq 0.5 \\ 0 & \text{else} \end{cases} \quad (66)$$

$$\begin{bmatrix} \text{digit1} & \text{digit2} & \text{digit3} \end{bmatrix} \begin{bmatrix} 1.24054754 \\ -0.11269202 \\ -0.11269202 \end{bmatrix} = [\text{value}] \quad (67)$$

$$\text{digit1} * w1 + \text{digit2} * w2 + \text{digit3} * w3 = \text{value} \quad (68)$$

$$\text{sigmoid}(\mathbf{y}^{pre}) = \frac{1}{1 + \exp^{-(\text{digit1} * w1 + \text{digit2} * w2 + \text{digit3} * w3)}} \quad (69)$$

$$\text{Prediction} = \begin{cases} 1 & \text{if } \text{sigmoid}(\mathbf{y}^{pre}) \geq 0.5 \\ 0 & \text{else} \end{cases} \quad (70)$$

Evaluating the model

- Accuracy

$$\text{Accuracy} = \frac{\# \text{ correctly classified instances}}{\text{total} \# \text{ instances}} \quad (71)$$

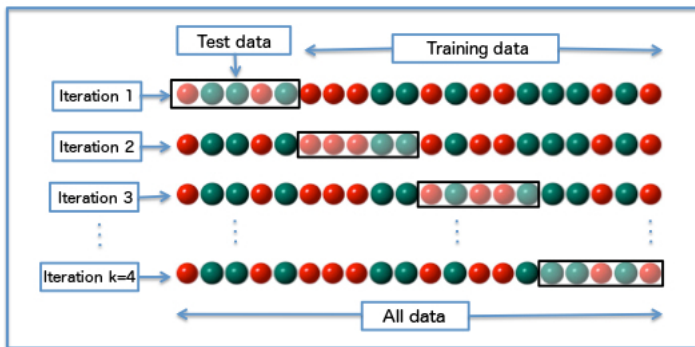
Evaluating the model

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} == \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \textit{False} \\ \textit{True} \\ \textit{True} \\ \textit{True} \\ \textit{True} \\ \textit{True} \\ \textit{True} \\ \textit{False} \end{bmatrix} \quad (72)$$

$$\text{Accuracy} = 6 / 8 * 100 = 75 \%$$

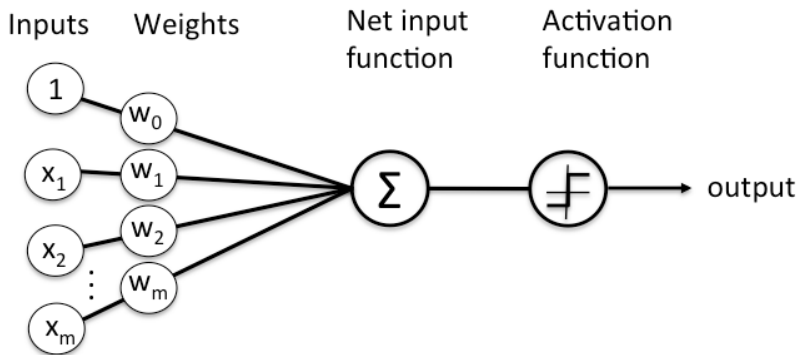
Improving the performance

10 - fold 10-cross validation



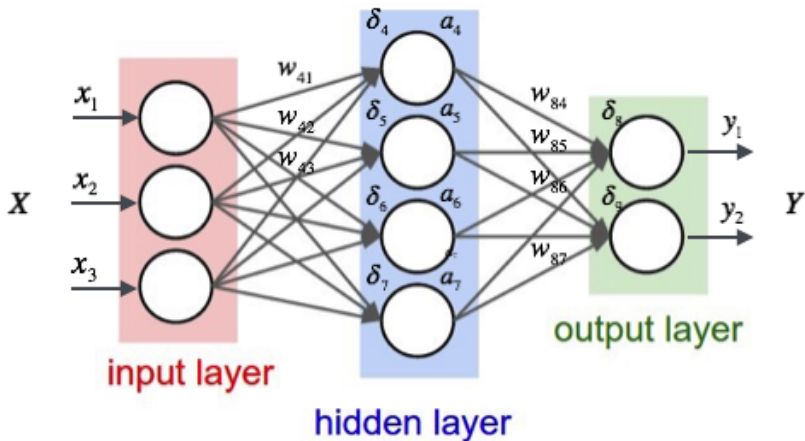
Source: wikipedia

Logistic Regression as a Neuron



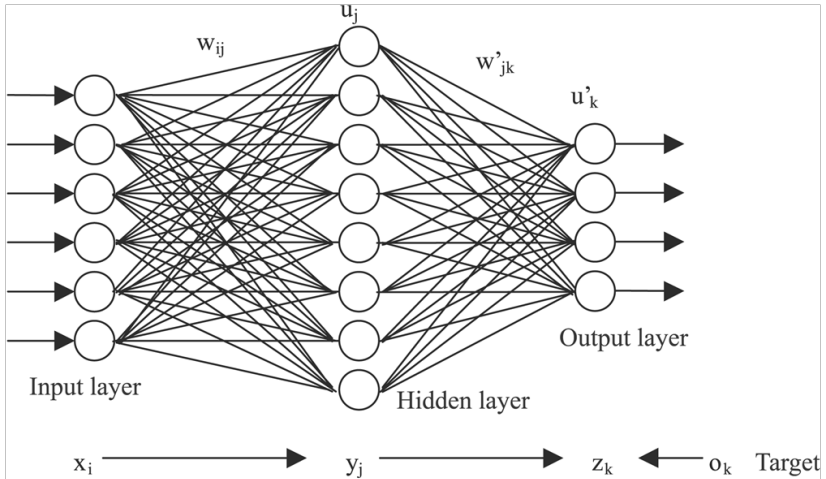
www.techmaru.com/technology/artificial-neural-networks/neural-network-elements

Neuron to Neurons



medium.com/@curiously/tensorflow-for-hackers-part-iv-neural-network-from-scratch-1a4f504dfa8

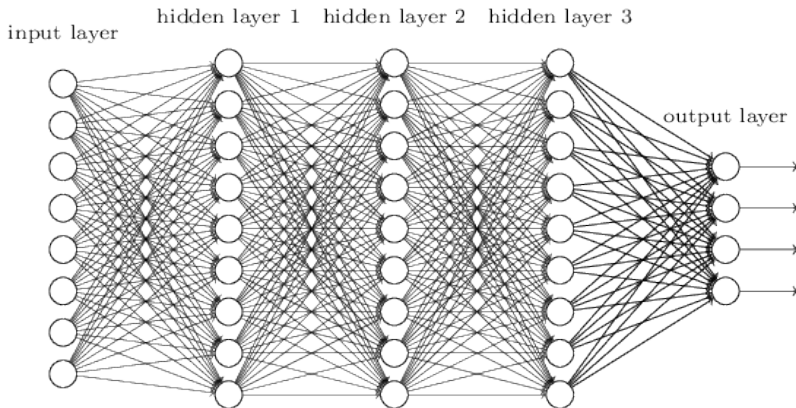
Single Layer Network



www.extremetech.com/extreme/

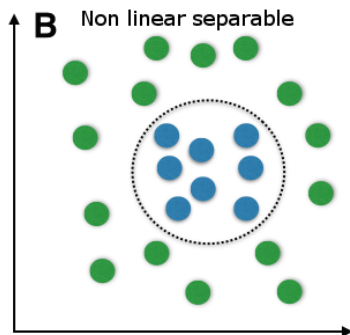
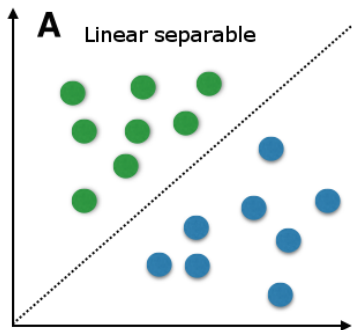
215170-artificial-neural-networks-are-changing-the-world-what-are-they

Multi Layer Network



in.mathworks.com/matlabcentral/fileexchange/64247-simple-neural-network

Why Deep Learning?

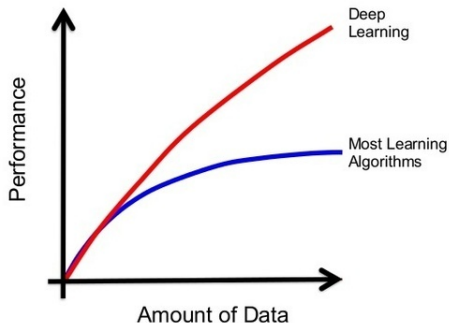


Source: https://leonardoraujosantos.gitbooks.io/artificial-intelligence/content/linear_classification.html

[//leonardoraujosantos.gitbooks.io/artificial-intelligence/content/linear_classification.html](https://leonardoraujosantos.gitbooks.io/artificial-intelligence/content/linear_classification.html)

Why Deep Learning?

BIG DATA & DEEP LEARNING



Source: <https://qph.ec.quoracdn.net/main-qimg-bf69c291005e68620a1bef39ae8f029e-c>

Why now Deep Learning?

WHY IS DEEP LEARNING HOT NOW?

Three Driving Factors...

Big Data Availability	New ML Techniques	Compute Density
<p>facebook 350 millions images uploaded per day</p> <p>Walmart 2.5 Petabytes of customer data hourly</p> <p>You Tube 100 hours of video uploaded every minute</p>	Deep Neural Networks	GPUs

ML systems extract value from Big Data

<https://www.slideshare.net/DataScienceMD/deep-learning-with-gpus>



Common Deep Learning Algorithms

- Convolutional Neural Network
- Recurrent Neural Network
- Long-Short Term Memory Network
- Deep Neural Network
- Auto Encoders



Matrix Representation



Linear Equations to Matrix

$$2a + b + c = 4 \quad (73)$$

$$a + 3b + 2c = 5 \quad (74)$$

$$a = 6 \quad (75)$$

Linear Equations to Matrix

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \quad (76)$$

Linear Equations to Matrix

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 0 & 0 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \quad (77)$$



Text to Matrix

- **S1:** We are in CEN.
- **S2:** CEN is in Amrita.
- **S3:** Amrita is in CBE.



Text to Matrix

- **S1:** We are in CEN.
- **S2:** CEN is in Amrita.
- **S3:** Amrita is in CBE.

Vocabulary = amrita, are, cen, cbe, in, is, we

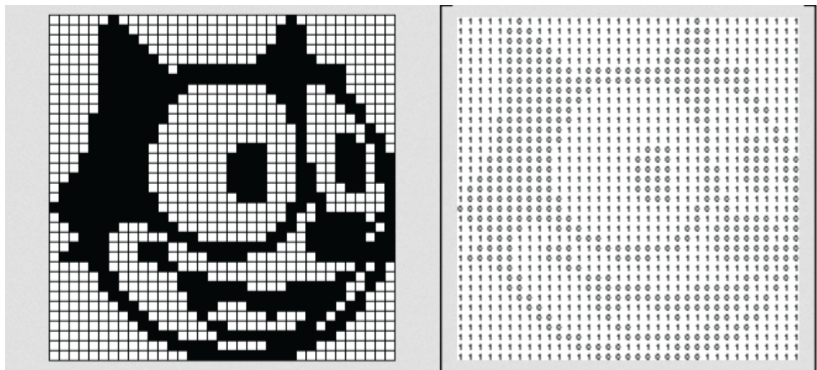


Text to Matrix

Table: Text Representation

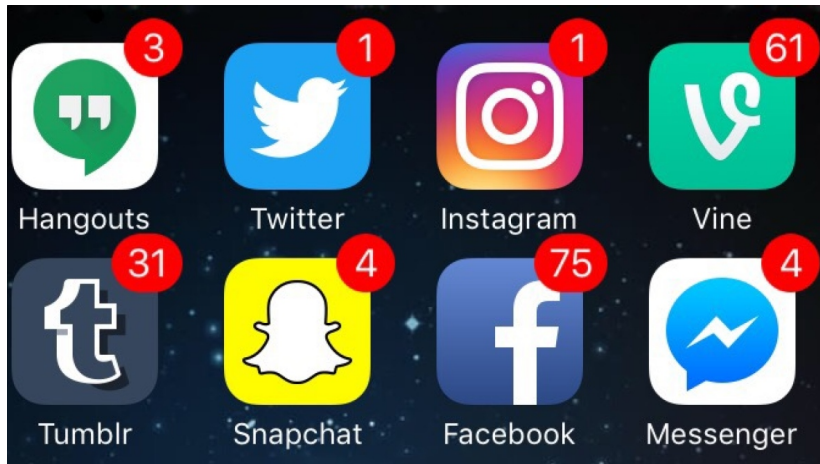
	amrita	are	cen	cbe	in	is	we
S1	0	1	1	0	1	0	1
S2	1	0	1	0	1	1	0
S3	1	0	0	1	1	1	0

Image to Matrix



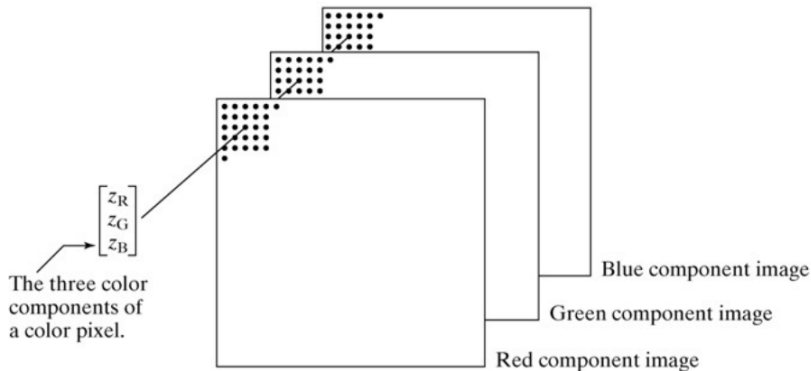
Source: blog.kleinproject.org/?p=588

Image to Matrix



Source: www.cbc.ca/news/trending

Image to Matrix



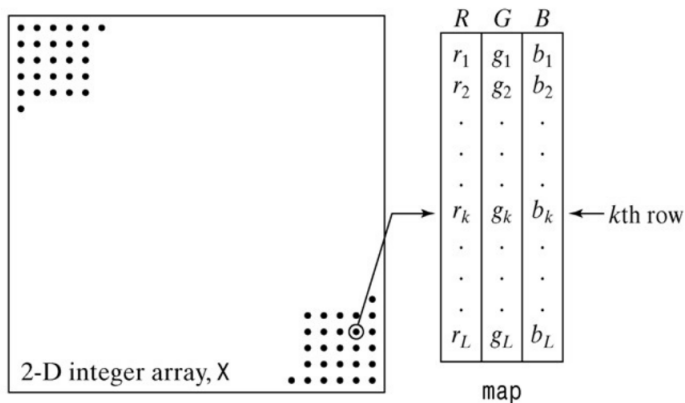
Source: <http://slideplayer.com/slide/8752313/>

Image to Matrix

88	82	84	88	85	83	80	93	102
88	80	78	80	80	78	73	94	100
85	79	80	78	77	74	65	91	99
38	35	40	35	39	74	77	70	65
20	25	23	28	37	69	64	60	57
22	26	22	28	40	65	64	59	34
24	28	24	30	37	60	58	56	66
21	22	23	27	38	60	67	65	67
23	22	22	25	38	59	64	67	66

Source: www1.adept.com/main/KE/DATA/ACE/AdeptSight_User/Vision_Basics_Mode.html

Image to Matrix



Value of circled element = k

Source: slideplayer.com/slide/8752313/

Thank You.

you can follow me through:

www.linkedin.com/in/barathiganeshhb

<https://barathiganesh-hb.github.io/>

<https://github.com/BarathiGanesh-HB/>