# Applied Artificial Intelligence

## Session 7: Linear Algebra for AI and Machine Learning I

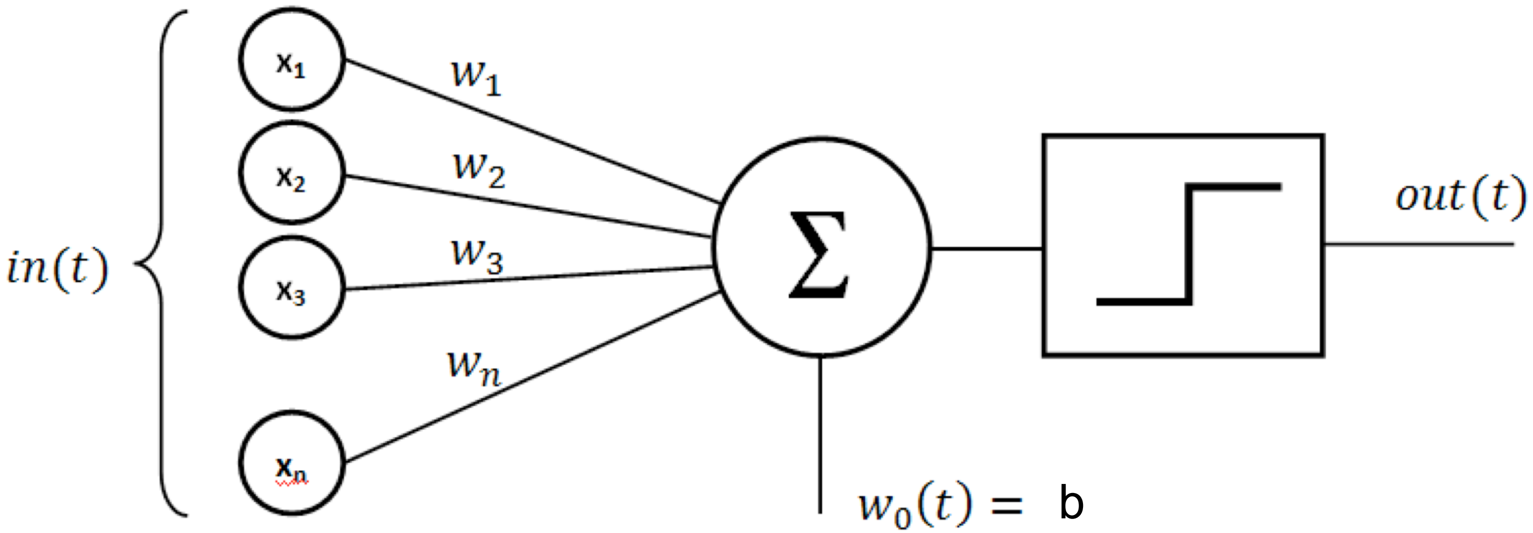
**Fall 2018**

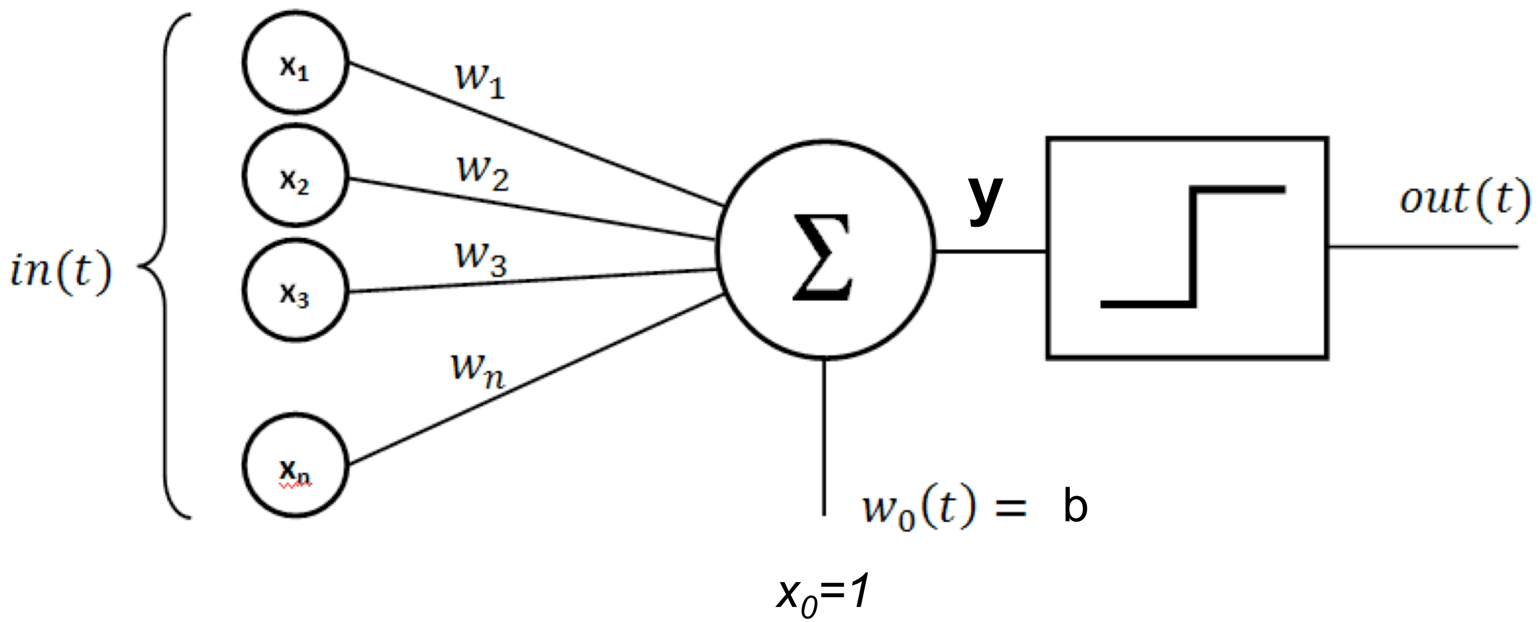
**NC State University**

**Instructor: Dr. Behnam Kia**

**Course Website: https://appliedai.wordpress.ncsu.edu/**

# Perceptron: A Computational Neuron Model

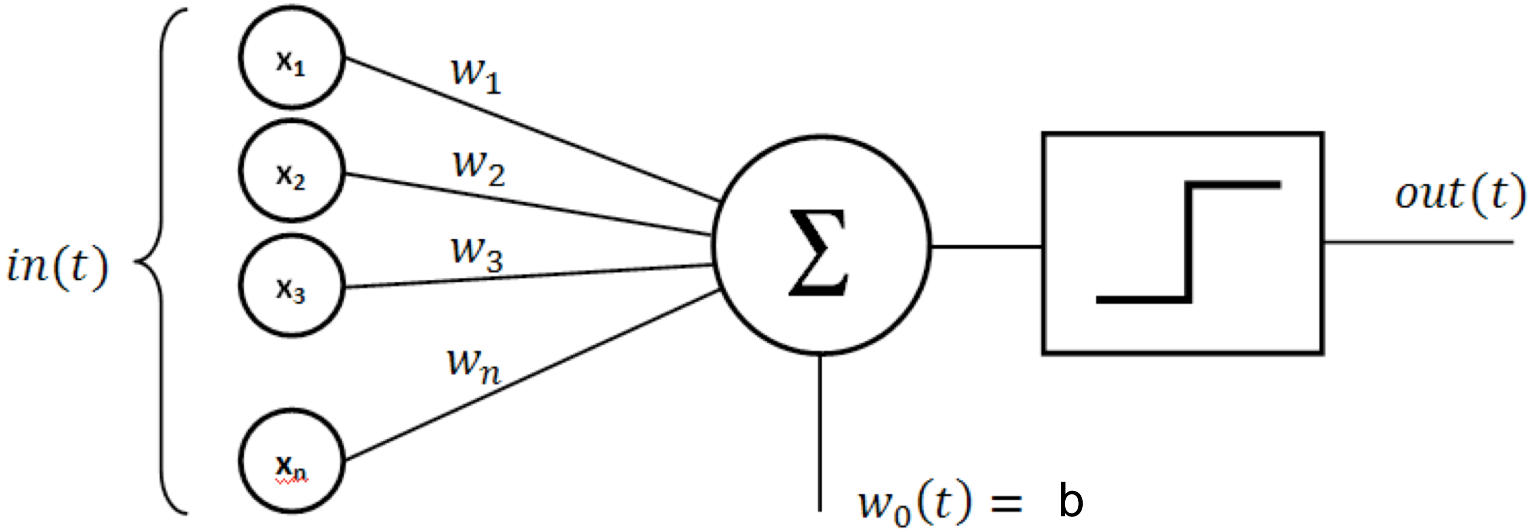





$$y = b + w_1x_1 + w_2x_2 + \ldots + w_nx_n$$

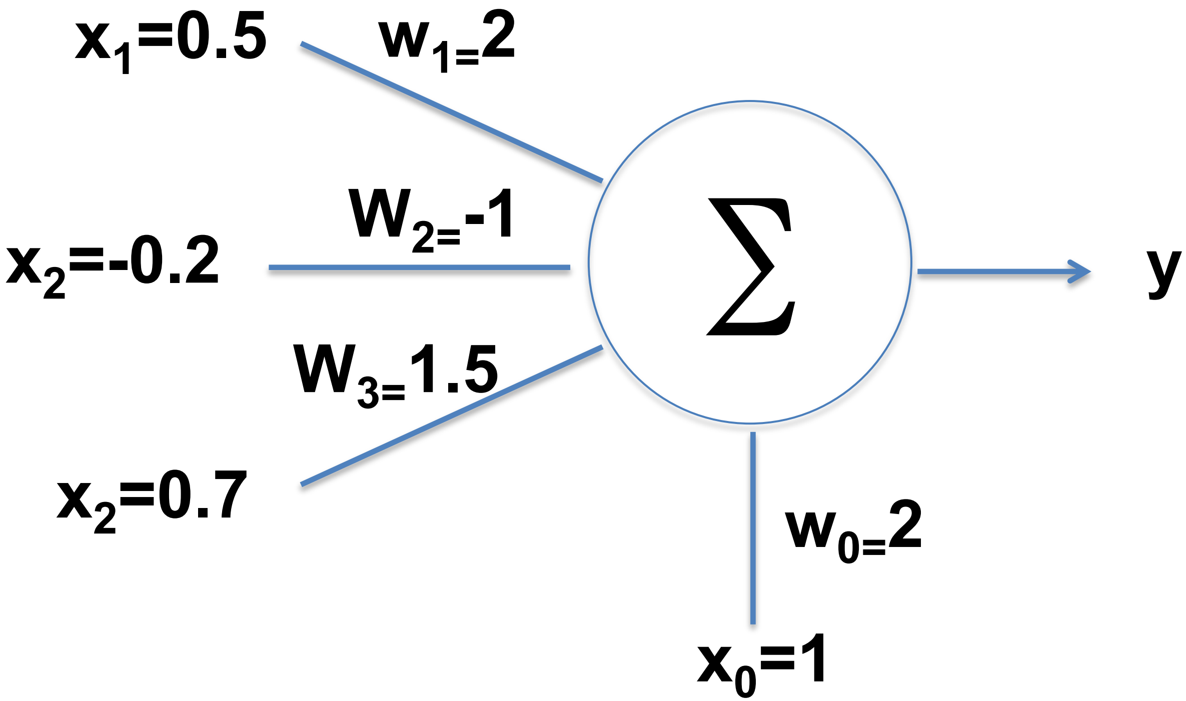
$$= w_0x_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n \quad (w_0 = b, x_0 = 1)$$

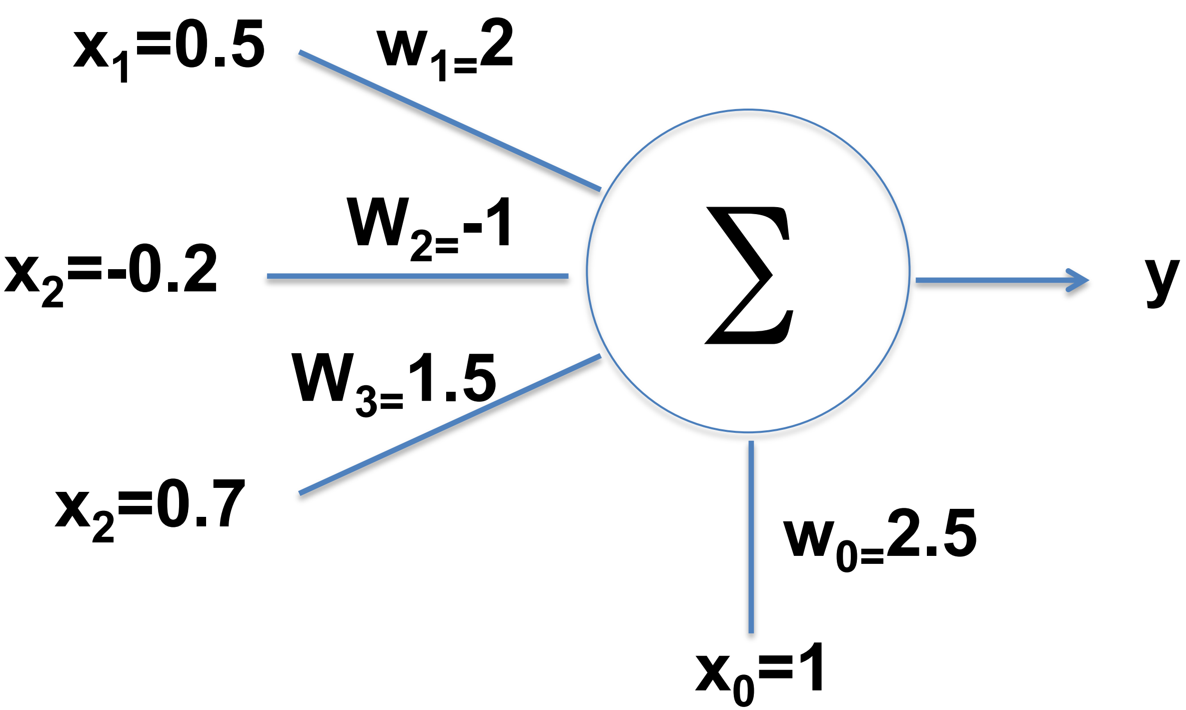
$$= \mathrm{W}^T.\mathrm{X}_p$$





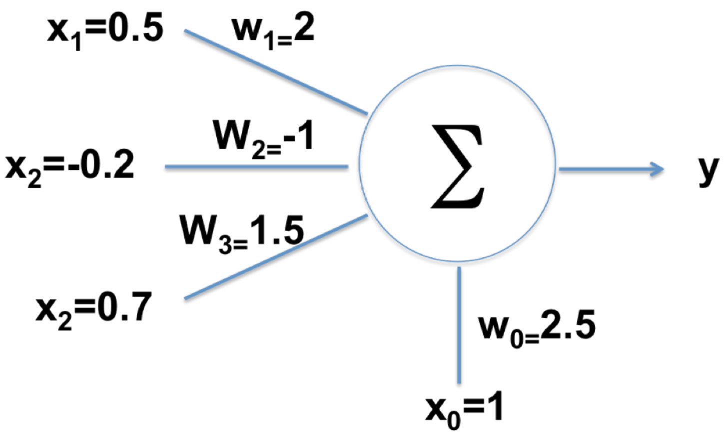

$$output = f(\mathrm{X}_{\mathrm{p}}) = \begin{cases} 1\ if\ \mathrm{W}^T.\mathrm{X}_p > 0 \\ 0\ if\ otherwise \end{cases}$$


x1=0.5
w1=2
x2=-0.2
W2=-1
W3=1.5
x2=0.7
y
w0=2
x0=1




$$W = \begin{bmatrix} 2.5 \\ 2 \\ -1 \\ 1.5 \end{bmatrix}$$

$$X_p = \begin{bmatrix} 1 \\ 0.5 \\ -0.2 \\ 0.7 \end{bmatrix}$$




$$X_p = \begin{bmatrix} 1 \\ 0.5 \\ -0.2 \\ 0.7 \end{bmatrix} \qquad W = \begin{bmatrix} 2.5 \\ 2 \\ -1 \\ 1.5 \end{bmatrix}$$

$$y = W^T X_p = \begin{bmatrix} 2.5 & 2 & -1 & 1.5 \end{bmatrix} . \begin{bmatrix} 1 \\ 0.5 \\ -0.2 \\ 0.7 \end{bmatrix} = 2.5 + 1 + 0.2 + 1.05 = 4.75$$

# Data Matrix X

**k is the number of features.**
**n is number of training data.**

$$X=\begin{bmatrix} 0.3 & \cdots & -0.1 \\ \vdots & \ddots & \vdots \\ 0.9 & \cdots & 0.43 \end{bmatrix}$$

(k,n)

$$X_p = \begin{bmatrix} 1 & \cdots & 1 \\ 0.3 & & -0.1 \\ \vdots & \ddots & \vdots \\ 0.9 & \cdots & 0.43 \end{bmatrix}$$

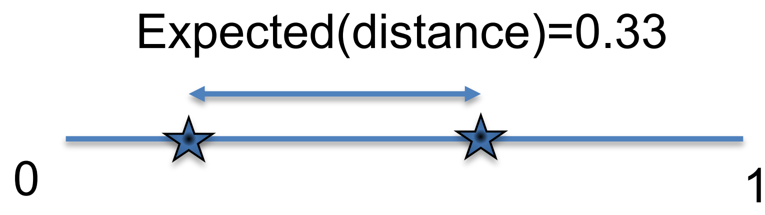
$$Y = W^T X_p$$

(1,n) (1,k+1) (k+1,n)

# The Curse of Dimensionality

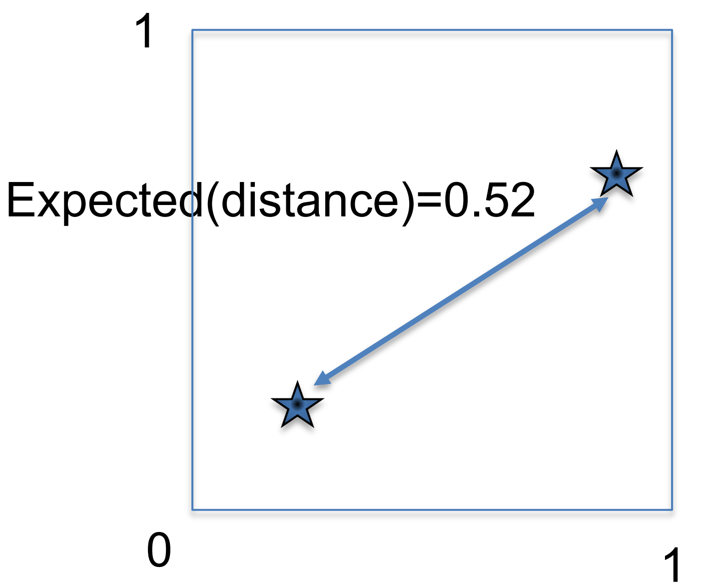
# 1-D

- If you select two points randomly in a unit line, the average distance between them=0.33.



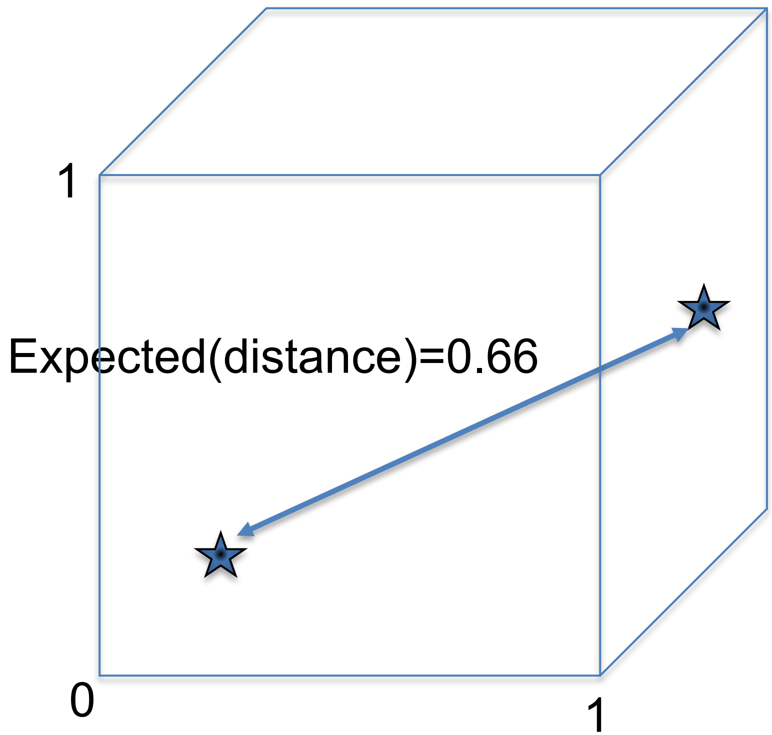
# 2-D

- If you select two points randomly in a unit square, the average distance between them≈0.52.



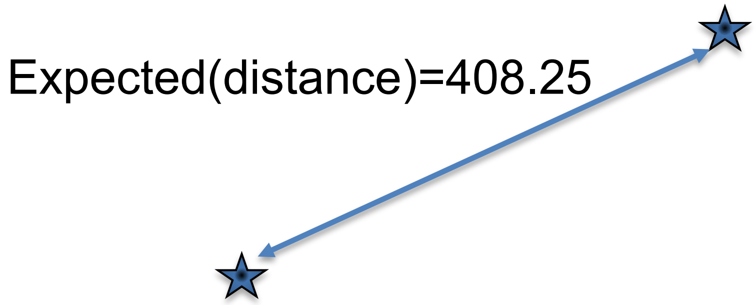
# 3-D

- If you select two data randomly in a unit 3D cube, the average distance between them≈0.66.



# 1,000,000-D

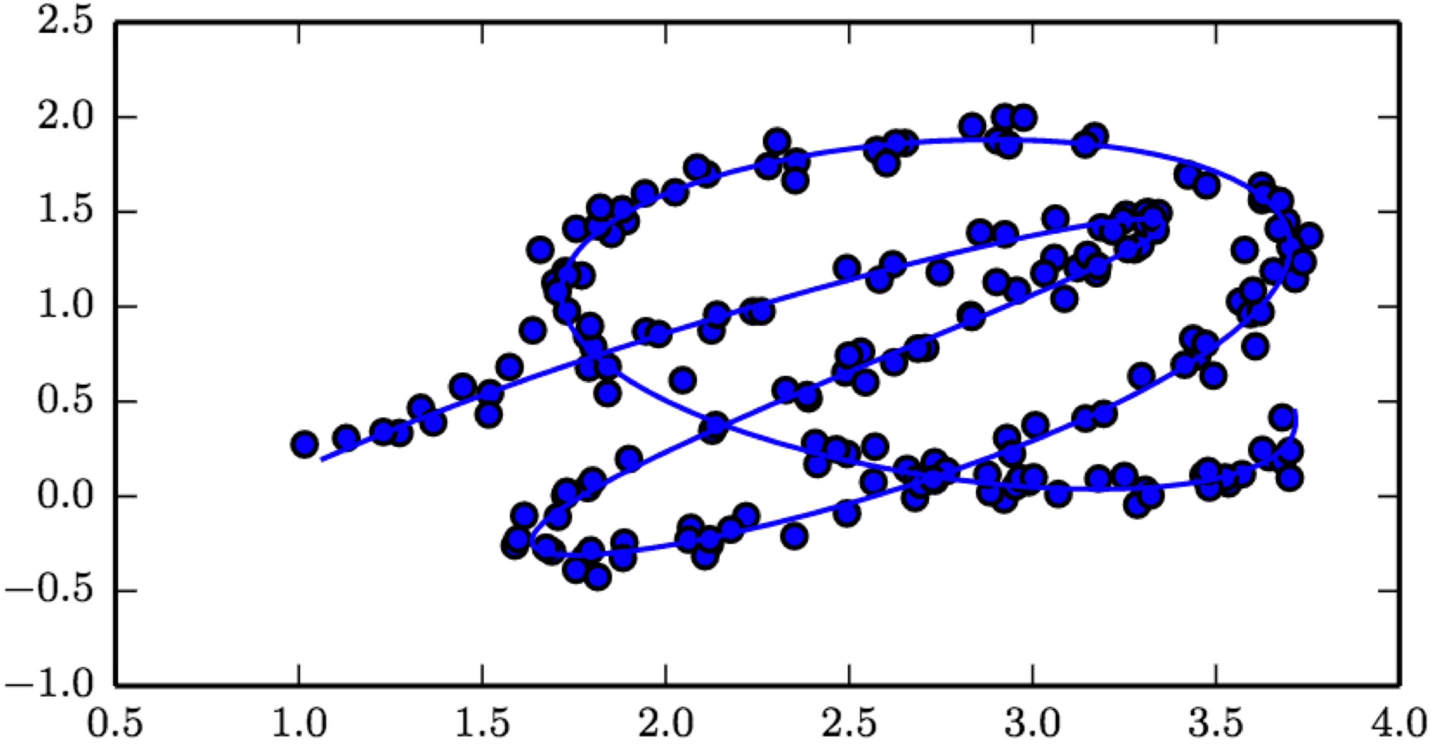
- If you select two points randomly in a unit one million-dimensional hypercube, the average distance between them≈408.25.



# The Curse of Dimensionality

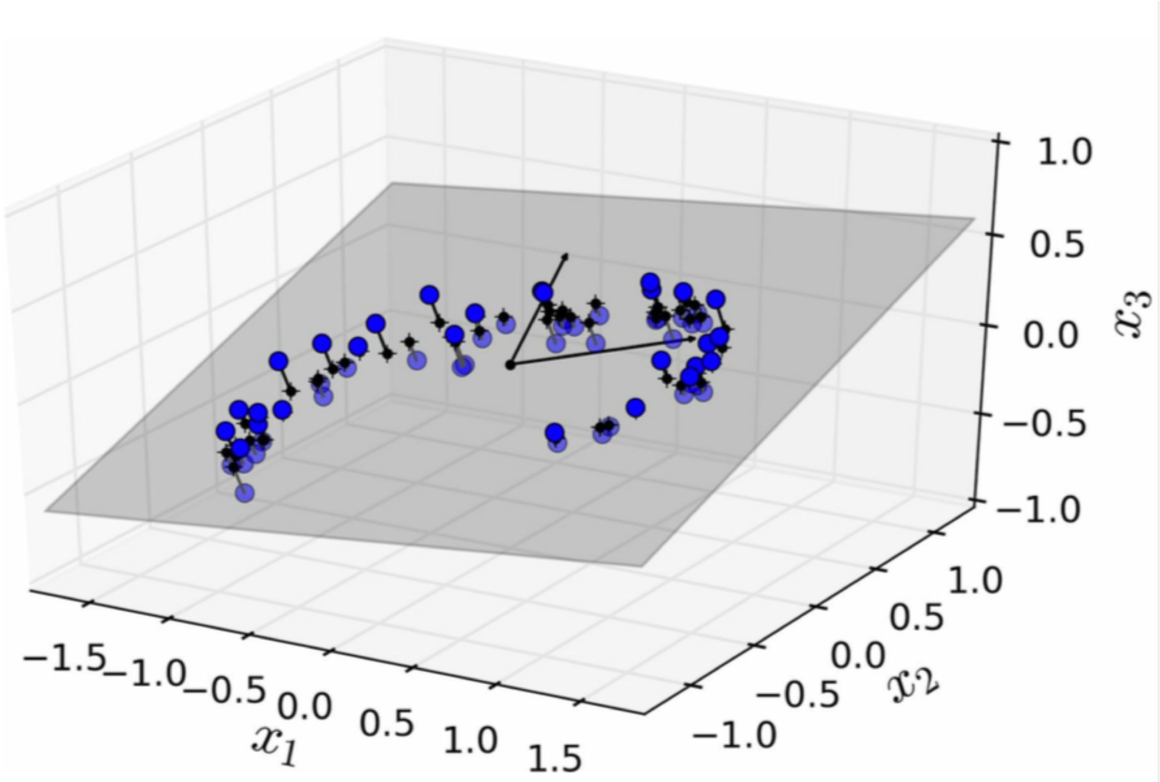
# How Scatter Plot of Data Set X looks like in the k dimensional Feature Space?

# How Scatter Plot of Data Set X looks like in the K dimensional Feature Space?



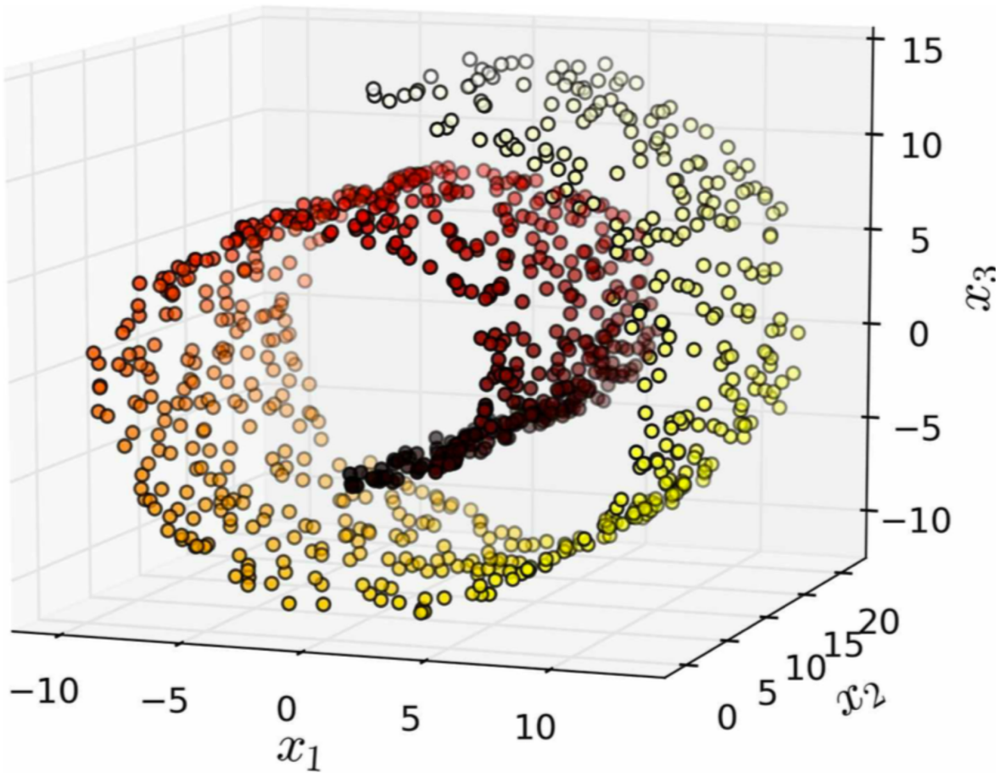

Picture is from: Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. Cambridge: MIT press; 2016 Nov 18

# How Scatter Plot of Data Set X looks like in the K dimensional Feature Space?




Picture is from: Géron, Aurélien. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. " O'Reilly Media, Inc.", 2017.

# How Scatter Plot of Data Set X looks like in the K dimensional Feature Space?

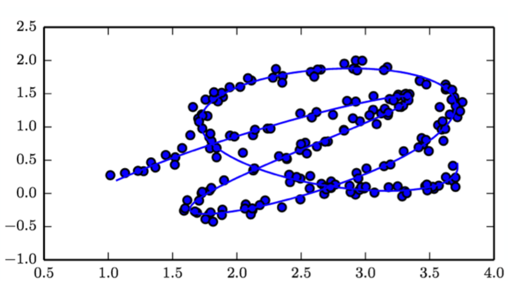


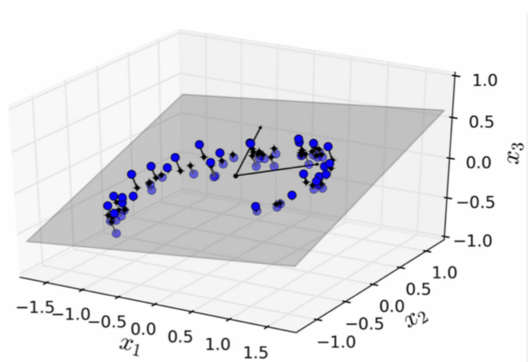

Picture is from: Géron, Aurélien. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. " O'Reilly Media, Inc.", 2017.

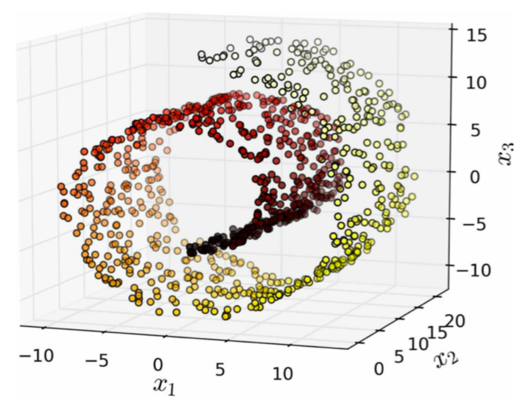
# Manifold Assumption (Manifold Hypothesis)

- Manifold Assumption: Real-world high-dimensional data sets lie close to a much lower-dimensional manifold.

- Manifold is a connected subset of a higher dimensional space. (rough definition)





# Dimension Reduction

# Dimension Reduction

- Linear Algebra provides many techniques, including Principal Component Analysis, for dimension reduction.

- Many of these techniques are implemented in Python Modules (e.g. in scikit-learn).

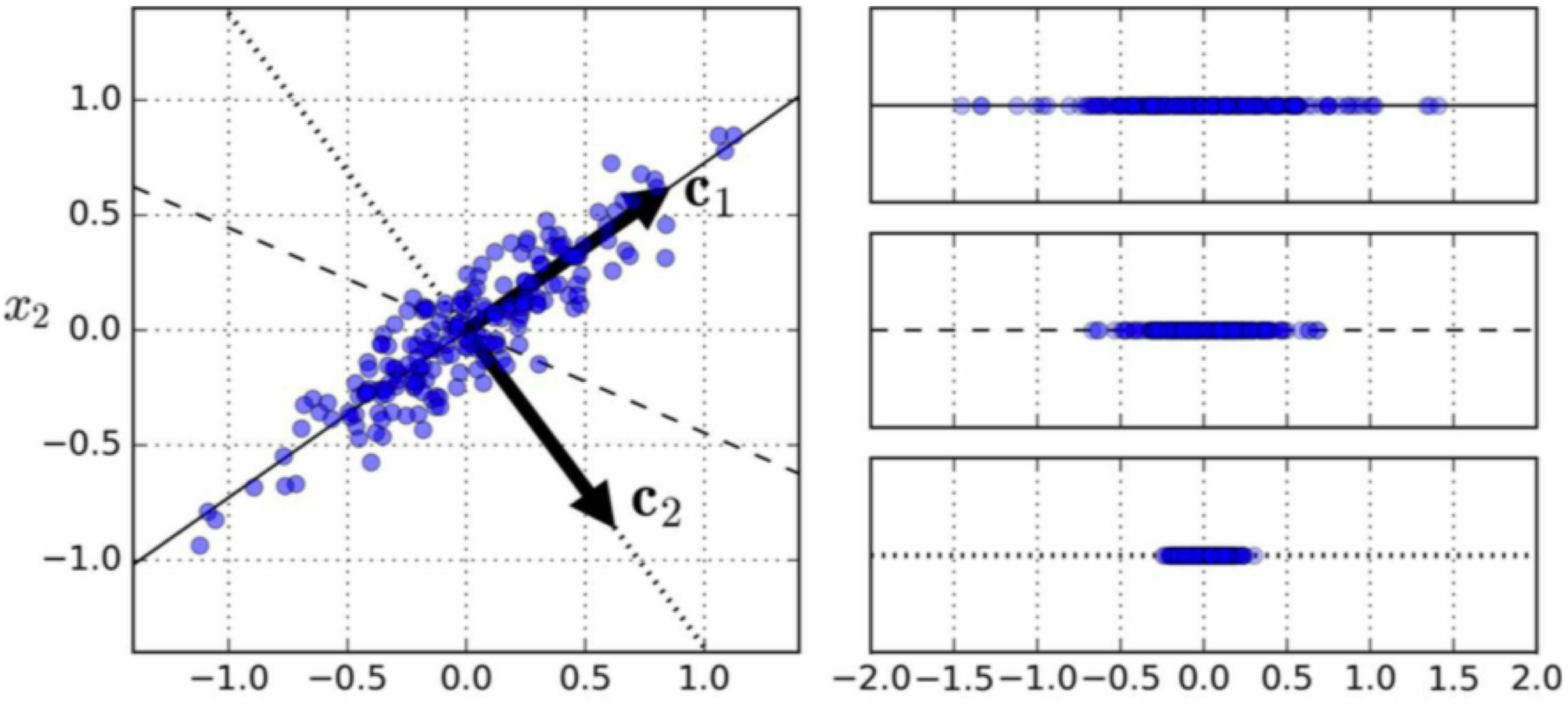
# Principal Component Analysis (PCA)

- PCA is the most common dimension reduction algorithm.
- There are many other dimension reduction algorithms as well.

# Principal Component Analysis (PCA)

- PCA identifies the closest axes to the data set.
- And Projects the data onto them.

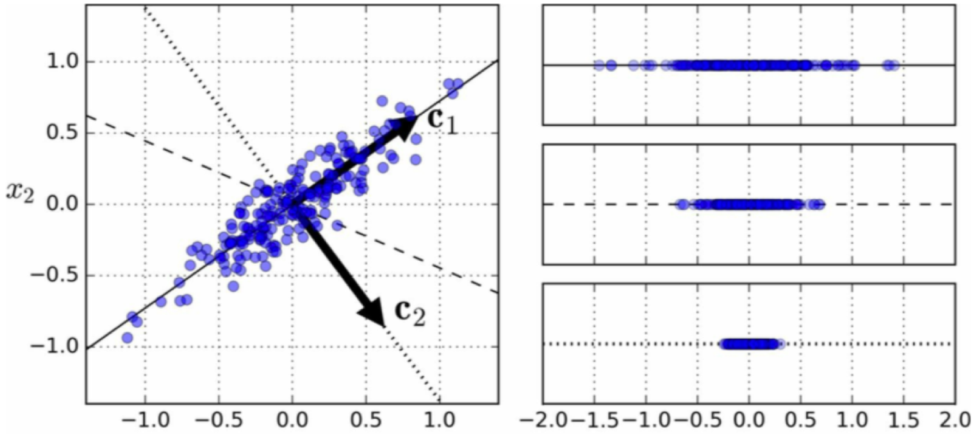
# Selecting the Subspace




Picture is from: Géron, Aurélien. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. " O'Reilly Media, Inc.", 2017.

# Selecting the Subspace

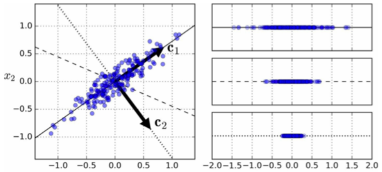
- We are better to choose the axis (the subspace) that preserves the maximum amount of variance.
- Variance=information!
- This is the main idea behind PCA.




Picture is from: Géron, Aurélien. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. " O'Reilly Media, Inc.", 2017.

# Selecting the Subspace

- We are better to choose the axis (the subspace) that preserves the maximum amount of variance.
- Variance=information!
- This is the main idea behind PCA.
- If dataset is k-dimensional, PCA returns back k principal components (axes), ranked in terms of the projected variance.
- And you can pick and choose how many components (dimensions) you like to keep, or how much variance preserve.



Picture is from: Géron, Aurélien. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. " O'Reilly Media, Inc.", 2017.

# Suggested Reading

Chapter 8: Dimension Reduction of

Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. " Géron, Aurélien. O'Reilly Media, Inc.", 2017.