AI 비전공자를 위한 기초 수학 1: 선형 대수학

Math for AI Beginner Part 1: Linear Algebra

Week 6: Diagonalization Problem and AI Applications

Linear Algebra

What we will do during the first 4 weeks…

Review of matrices

Systems of linear algebraic equations AX=B

Finding solution(s) by row operations

Inverse of a square matrix A

Finding inverse by row operations

Determinant of a square matrix A

Calculating determinant by row operations

Bigger pictures: how are these 3 topics connected together with AI?

In between, we will look at vector space and linearly independent vectors.

Matrix eigen-problem

Diagonalisation problem

AI Applications: Deep Learning & SVM

We are now going to diagonalise a square matrix.

What does it mean to diagonalise a square matrix?

How can be it be done?

The diagonalisation problem is related to the matrix eigen-problem.

The problem of diagonalising a matrix may be stated as:

Given an $N \times N$ matrix $\mathbf{A}$, can we find an *invertible* $N \times N$ matrix $\mathbf{P}$ and an $N \times N$ *diagonal* matrix $\mathbf{D}$ such that

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} \text{ ?}$$

An example of a 5×5 diagonal matrix is:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$$

The task of diagonalising $\mathbf{A}$ is simply to find matrices $\mathbf{P}$ and $\mathbf{D}$ such that $\mathbf{A}$ can be written as $\mathbf{P}\mathbf{D}\mathbf{P}^{-1}$.

Not all square matrices can be diagonalised.

Yes, we can diagonalise the $N \times N$ matrix $\mathbf{A}$, that is, we can write $\mathbf{A} = \mathbf{PDP}^{-1}$, **if and only if $\mathbf{A}$ has N linearly independent eigenvectors.**

If $\mathbf{A}$ does not have N linearly independent eigenvectors, it is not diagonalisable.

If $\mathbf{A}$ has N linearly independent eigenvectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$, how do we construct $\mathbf{P}$ and $\mathbf{D}$?

$$\mathbf{P} = (\underbrace{\mathbf{X}_1}_{\text{1st column of P}} \quad \underbrace{\mathbf{X}_2}_{\text{2nd column of P}} \quad \ldots \quad \mathbf{X}_{N-1} \quad \underbrace{\mathbf{X}_N}_{N\text{th column of P}})$$

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & & & & \\ & \lambda_2 & & \text{O} & \\ & & \ddots & & \\ & \text{O} & & \lambda_{N-1} & \\ & & & & \lambda_N \end{pmatrix}$$

λ_j is the eigenvalue which gives the eigenvector $\mathbf{X}_j$.

Example:

Diagonalise the matrix $\mathbf{A} = \begin{pmatrix} 0 & 0 & 2 \\ -2 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix}$.

Can we find 3 linearly independent eigenvectors?

In this example, A has only 2 distinct eigenvalues: 0 and 2.

Eigenvectors are as follows:

$\lambda = 0$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = t \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

Take t = 1 to give one eigenvector, that is,

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

We cannot take a second eigenvector from here. Why?

$\lambda = 2$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = r \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + s \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Let r = 1 and s = 0 to obtain $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

Let r = 0 and s = 1 to obtain $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

We have three linearly independent eigenvectors:

$$\begin{pmatrix}1\\1\\0\end{pmatrix} \quad \begin{pmatrix}1\\0\\1\end{pmatrix} \quad \begin{pmatrix}0\\1\\0\end{pmatrix}$$

$\lambda=0$ $\quad$ $\lambda=2$ $\quad$ $\lambda=2$

A possible answer for diagonalising $\mathbf{A}$ is $\mathbf{A}=\mathbf{PDP}^{-1}$, where

$$\mathbf{P} = \begin{pmatrix}1 & 1 & 0\\ 1 & 0 & 1\\ 0 & 1 & 0\end{pmatrix}$$

$$\mathbf{D} = \begin{pmatrix}0 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 2\end{pmatrix}$$

$$\mathbf{P}^{-1} = \begin{pmatrix}1 & 0 & -1\\ 0 & 0 & 1\\ -1 & 1 & 1\end{pmatrix}$$

Check: A =PDP^{-1}.

Example :

Can we diagonalise the matrix $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 2 \end{pmatrix}$ **?**

Can we find 3 linearly independent eigenvectors?

In this example, A also has only 2 distinct eigenvalues: 1 and 2.

Eigenvectors are as follows:

$\lambda = 1$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = s \begin{pmatrix} 0 \\ 1 \\ -4 \end{pmatrix}$$

Take *s* = 1 to give one eigenvector, that is,

$$\begin{pmatrix} 0 \\ 1 \\ -4 \end{pmatrix}$$

We cannot take another eigenvector from here

$\lambda = 2$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = t \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Let *t* = 1 to obtain $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

We cannot take another eigenvector from here

Can we find 3 linearly independent eigenvectors?

No! Thus A cannot be diagonalised.

Good news!

Eigenvectors (non-trivial ones) that come from distinct eigenvalues of a square matrix are linearly independent.

If an *N*×*N* matrix $\mathbf{A}$ has *N* distinct eigenvalues, we can select one eigenvector from each eigenvalue to form *N* linearly independent eigenvectors.

Such a matrix $\mathbf{A}$ is diagonalisable.

As shown in the last two examples, if an *N*×*N* matrix $\mathbf{A}$ has less than *N* distinct eigenvalues, it may or may not be diagonalisable depending on whether *N* linearly independent vectors can be found.

Another good news!

If $\mathbf{A}$ is symmetric, it may be possible to construct $\mathbf{P}$ in such a way that $\mathbf{P}^{-1}$ can be easily found.

The transpose of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^T$.

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \qquad \mathbf{A}^{\mathrm{T}} = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$$

If $\mathbf{A} = \mathbf{A}^T$ then $\mathbf{A}$ is said to be symmetric.

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 5 \\ 3 & 0 & 6 \\ 5 & 6 & 9 \end{pmatrix} \qquad \mathbf{A}^{\mathrm{T}} = \begin{pmatrix} 1 & 3 & 5 \\ 3 & 0 & 6 \\ 5 & 6 & 9 \end{pmatrix}$$

Here is a theorem for a symmetric matrix.

If $\mathbf{A}$ is an $N \times N$ symmetric matrix whose elements are real numbers then $\mathbf{A}$ has only real eigenvalues.

Furthermore, if the symmetric matrix $\mathbf{A}$ has N distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{N-1}$ and λ_N with unit norm eigenvectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{N-1}$ and $\mathbf{X}_N$ respectively, then we can form $\mathbf{P}=(\mathbf{X}_1 \ \mathbf{X}_2 \ \ldots \ \mathbf{X}_N)$ such that $\mathbf{P}^{-1} = \mathbf{P}^{T}$. Hence, we can write $\mathbf{A} = \mathbf{PDP}^{T}$.

The norm of a vector $\mathbf{X}$ is denoted by $\|\mathbf{X}\|$.

If $\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$ **then** $\|\mathbf{X}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_N^2}$

Example:

Diagonalise $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$

Solve the matrix eigen-problem first. There are 3 distinct eigenvalues.

$\lambda = 1$ $$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -t \\ t \\ 0 \end{pmatrix}$$

$$\sqrt{(-t)^2 + t^2 + 0^2} = 1 \Rightarrow t = \frac{1}{\sqrt{2}}$$

Choose $\frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$

$\lambda = -1$ $$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -p/2 \\ -p/2 \\ p \end{pmatrix}$$

$$\sqrt{(\frac{-p}{2})^2 + (\frac{-p}{2})^2 + p^2} = 1 \Rightarrow p = \frac{2}{\sqrt{6}}$$

Choose $\frac{1}{\sqrt{6}}\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$

$\lambda = 2$ $$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} s \\ s \\ s \end{pmatrix}$$

$$\sqrt{s^2 + s^2 + s^2} = 1 \Rightarrow s = \frac{1}{\sqrt{3}}$$

Choose $\frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

A is symmetric and has 3 distinct eigenvalues.

$$\lambda = 1 \quad \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

$$\lambda = -1 \quad \frac{1}{\sqrt{6}}\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$$

$$\lambda = 2 \quad \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\mathbf{P} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix} \qquad \mathbf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$$\mathbf{P}^{-1} = \mathbf{P}^{\mathrm{T}} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}$$

Check: $\mathbf{PP}^{\mathrm{T}} = \mathbf{I}$ and $\mathbf{A} = \mathbf{PDP}^{\mathrm{T}}$.

If a square matrix $\mathbf{A}$ can be diagonalised, it is relatively easy to compute $\mathbf{A}^M$ for high power *M*, e.g. $\mathbf{A}^{30}$.

$$\mathbf{A} = \mathbf{PDP}^{-1}$$

E.g.

$$\mathbf{A}^5 = \mathbf{P}\,\mathbf{D}\,\mathbf{P}^{-1}\mathbf{P}\,\mathbf{D}\,\mathbf{P}^{-1}\mathbf{P}\,\mathbf{D}\,\mathbf{P}^{-1}\mathbf{P}\,\mathbf{D}\,\mathbf{P}^{-1}\mathbf{P}\,\mathbf{D}\,\mathbf{P}^{-1}$$

$$= \mathbf{P}\,\mathbf{D}\,\mathbf{I}\,\mathbf{D}\,\mathbf{I}\,\mathbf{D}\,\mathbf{I}\,\mathbf{D}\,\mathbf{I}\,\mathbf{D}\,\mathbf{P}^{-1} = \mathbf{PD}^5\mathbf{P}^{-1}$$

In general:

$$\mathbf{A}^M = \mathbf{PD}^M\mathbf{P}^{-1}$$

$\mathbf{D}^M$ **is easy to calculate if D is a diagonal matrix.**

E.g.

$$\mathbf{D} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \Rightarrow \mathbf{D}^{20} = \begin{pmatrix} a^{20} & 0 & 0 \\ 0 & b^{20} & 0 \\ 0 & 0 & c^{20} \end{pmatrix}$$

Example:

A is a 3×3 matrix which is diagonalisable such that

$$\mathbf{A} = \mathbf{P}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}\mathbf{P}^{-1}.$$

What are the eigenvalues of A^{-1}?

What can you say about the eigenvectors of A^{-1}?

What is the determinant of A^{-1}?

$$\mathbf{A} = \mathbf{P}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}\mathbf{P}^{-1}$$

Use

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

$$\Rightarrow \quad \mathbf{A}^{-1} = \left(\mathbf{P}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}\mathbf{P}^{-1} \right)^{-1} = \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}\mathbf{P}^{-1} \right)^{-1} \mathbf{P}^{-1}$$

$$= (\mathbf{P}^{-1})^{-1}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}^{-1} \mathbf{P}^{-1}$$

$$= \mathbf{P}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}\mathbf{P}^{-1}$$

$$\mathbf{A}^{-1} = \mathbf{P}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}\mathbf{P}^{-1}$$

Eigenvalues of A^{-1} are 1, 1/2 and 1/3.

Eigenvectors of A^{-1} corresponding to 1, 1/2 and 1/3 are eigenvectors of A corresponding to 1, 2 and 3 respectively.

$$\frac{1}{\det(\mathbf{P})}$$

$$\mathbf{det(A^{-1})} = \det(\mathbf{P}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}\mathbf{P}^{-1})$$

$$= \det(\mathbf{P})\det(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}\mathbf{P}^{-1}) = \det(\mathbf{P})\det(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{pmatrix})\,\boxed{\mathbf{det(P^{-1})}}$$

$$\textit{use}\ \det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$$

Linear Algebra for Basic AI

What we've done during last 3 weeks…

Review of matrices

Systems of linear algebraic equations AX=B

Finding solution(s) by row operations

Inverse of a square matrix A

Finding inverse by row operations

Determinant of a square matrix A

Calculating determinant by row operations

Bigger pictures: how are these 3 topics connected together with AI?

In between, we will look at vector space and linearly independent vectors.

Matrix eigen-problem

Diagonalisation problem

AI Applications: Deep Learning & SVM

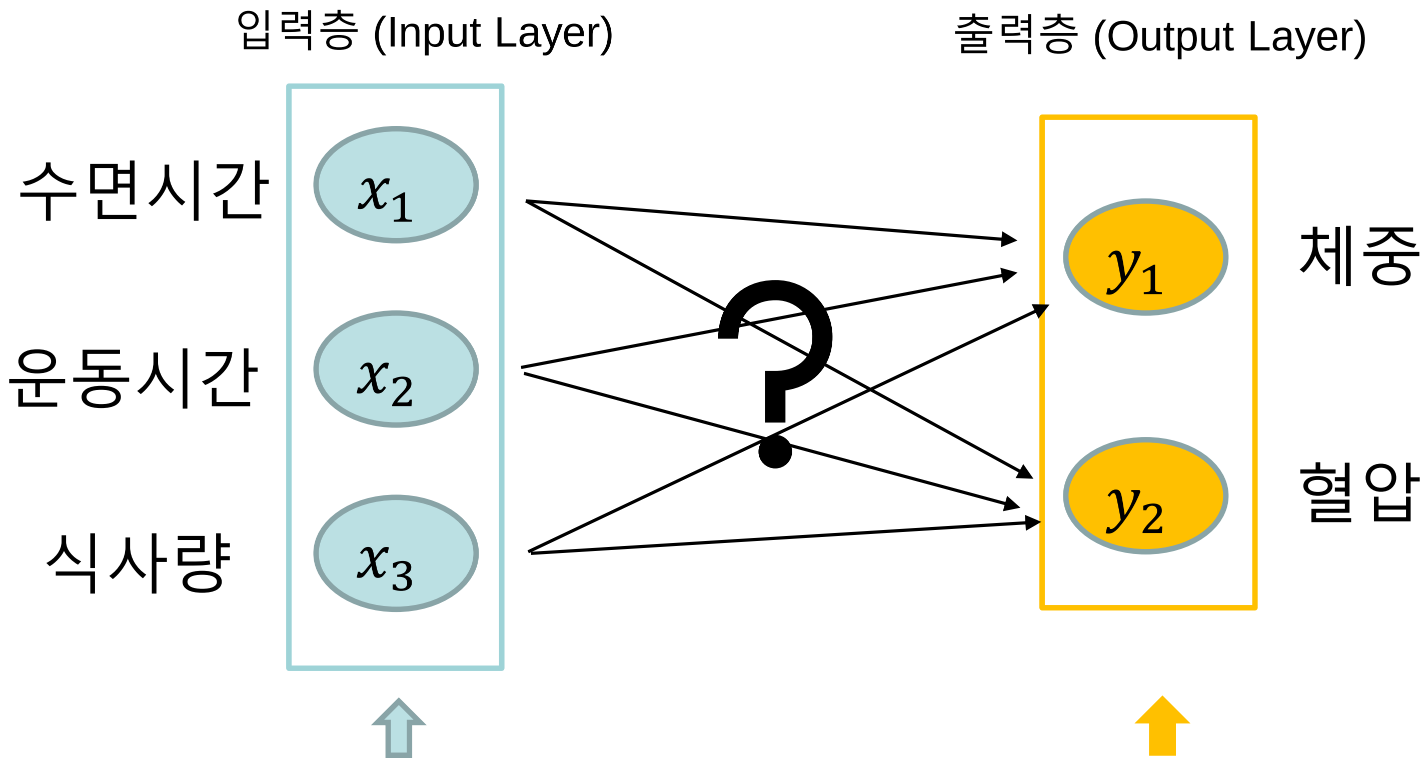
Q1 인공지능이란 무엇인가요?

Q2 인공지능은 왜 이제서야 나타났나요?

Q3 4차 산업혁명? 인공지능?

Q4 인공지능을 이해하는데 수학이 필요한가요?

인공지능과 선형대수학 (Single Layer ANN)



똘이) 수면시간: 6 운동시간: 3 식사량 2500 → 체중 70 kg/ 혈압 110

순이) 수면시간: 4 운동시간: 1 식사량 1800 → 체중 50 kg/ 혈압 100

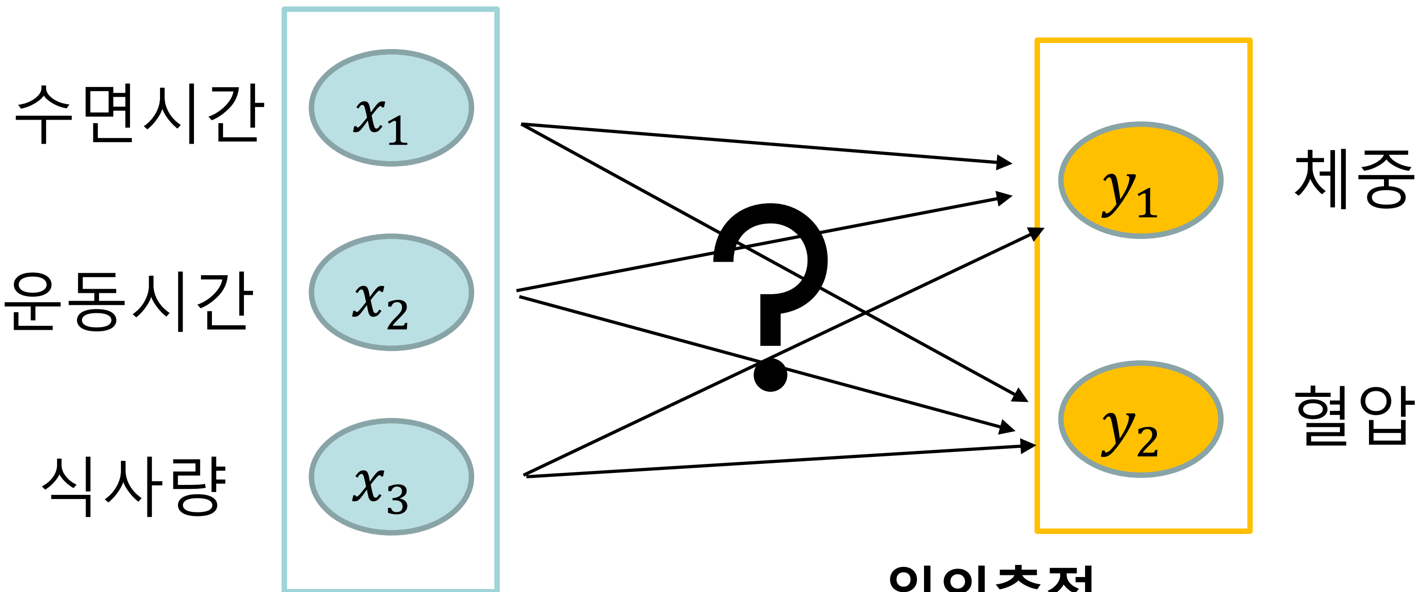
토니) 수면시간: 8 운동시간: 0 식사량 4000 → 체중 100 kg/ 혈압 130

인공지능과 선형대수학 (Single Layer ANN)

똘이) 수면시간: 6 운동시간: 3 식사량 2500 → 체중 70 kg/ 혈압 110

입력층 (Input Layer)

출력층 (Output Layer)



임의추정

[100 180]

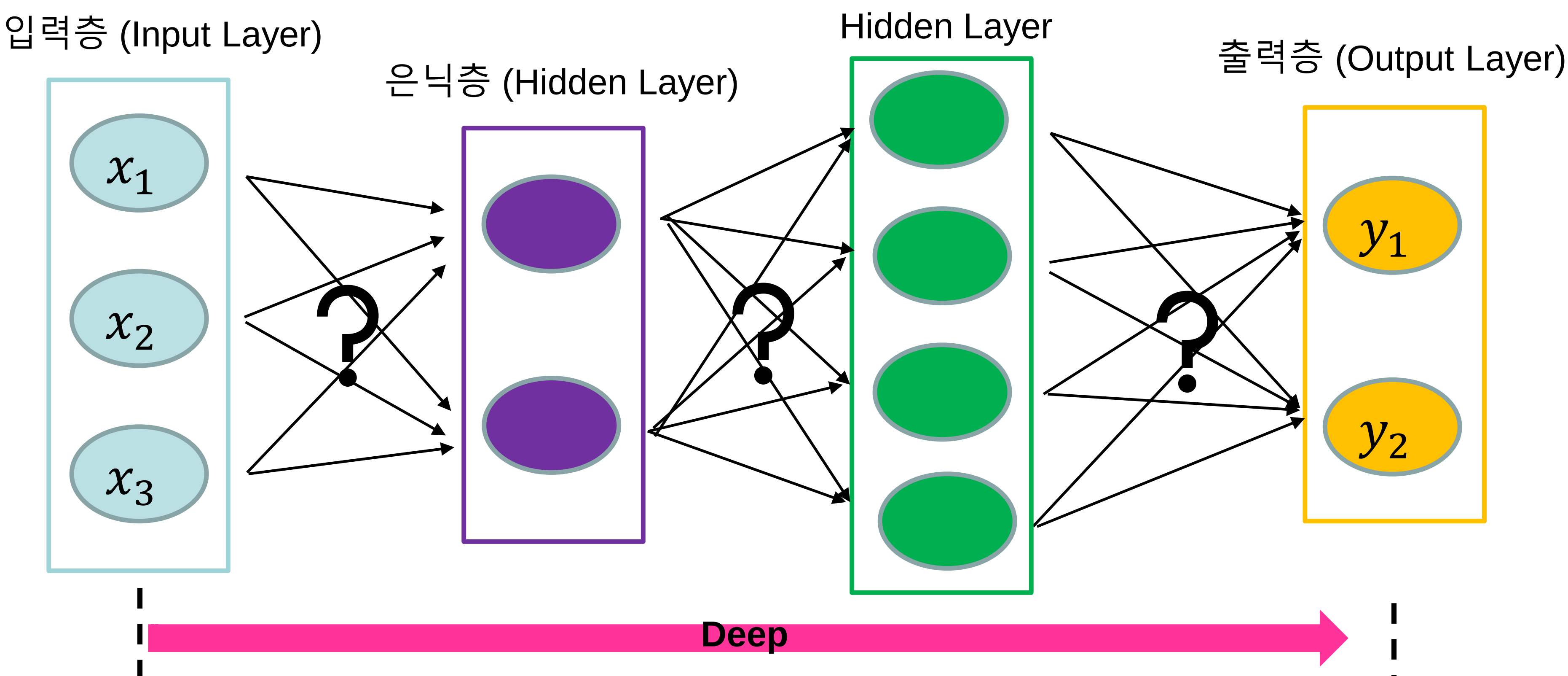
$$[\,x_1\ x_2\ x_3\,] \begin{bmatrix} 3 \times 2 \end{bmatrix} = [y_1 y_2] \leftarrow$$ **추정 출력 값**

1 x 3   1 x 2

[6 3 2500]   **최적화 (미분)**   [70 110]

정답과의 차이

인공지능과 선형대수학 (Deep Learning)



$$[\,x_1\ x_2\ x_3\,]\begin{bmatrix}\cancel{3}\times\cancel{2}\end{bmatrix}\begin{bmatrix}\cancel{2}\times\cancel{4}\end{bmatrix}\begin{bmatrix}\cancel{4}\times 2\end{bmatrix} = [y_1 y_2]$$

$1 \times \cancel{3}$ $\qquad$ 1×2

Linear Algebra & Support Vector Machine (SVM)

The **Support Vector Machine (SVM)** has been shown to be able to achieve good **generalization performance** for **classification of high-dimensional data sets** and its training can be framed as solving a **quadratic programming** Problem.

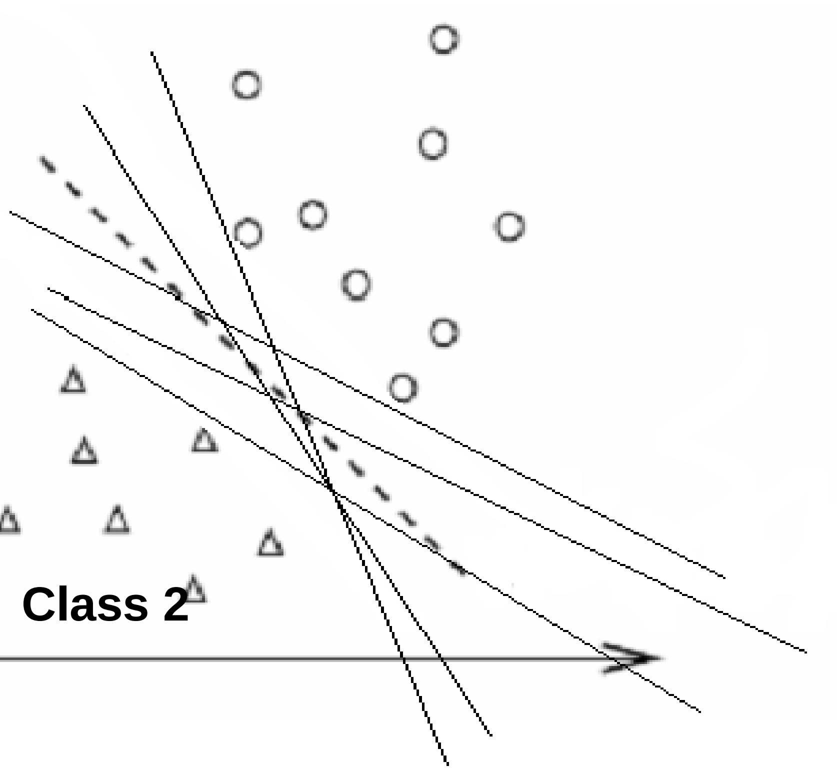
Usually we try to maximize **classification performance** for the **training data**

However, if the classifier is **too fit for the training data**, the [classification ability for **unknown data = generalization ability**] **is degraded**

SVM is trained so that the **direct decision function maximizes** the **generalization ability**

SVM 은 Machine Learning 중 **분류** 문제에 주로 쓰임!

Linear Algebra & Support Vector Machine (SVM)



- **<u>Which line will classify the unseen data well?</u>**

(2차원: 직선, 3차원: 평면, 4차원 이상: Hyperplane)

Hyperplane 의 일반식

$$W^T X + b = 0$$

***where, W*: normal vector of hyperplane**
***b*: bias**

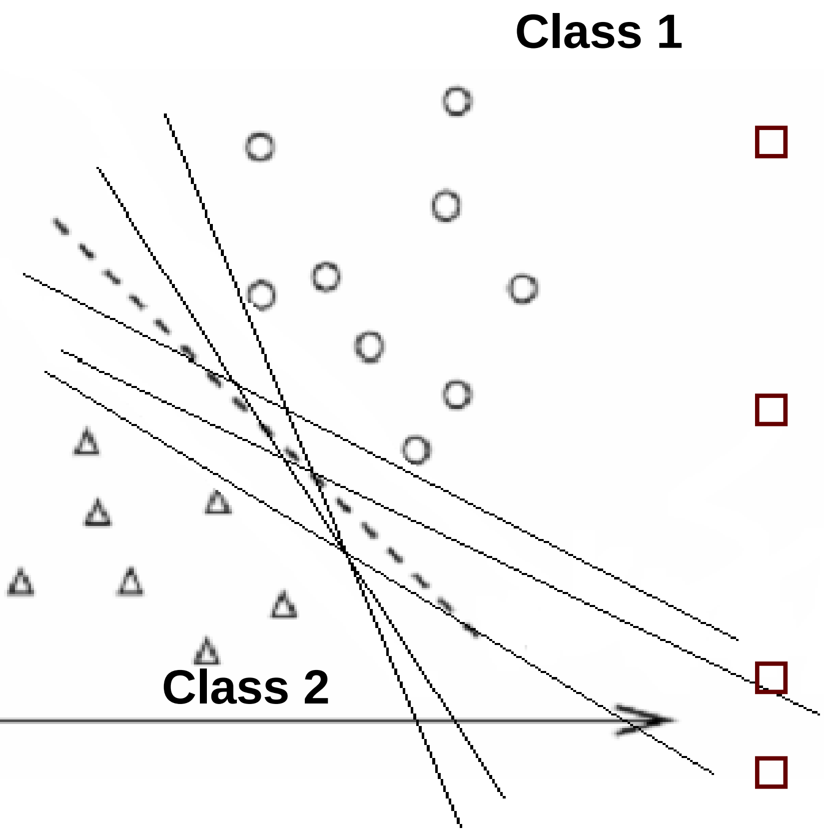
2차원 (x, y) 좌표에서의 직선: ax+by+c=0

$$\boldsymbol{W^T = [a\ b], X = \begin{bmatrix} x \\ y \end{bmatrix}, c = b}$$

3차원 (x, y,z) 좌표에서의 평면: ax+by+cz+d=0

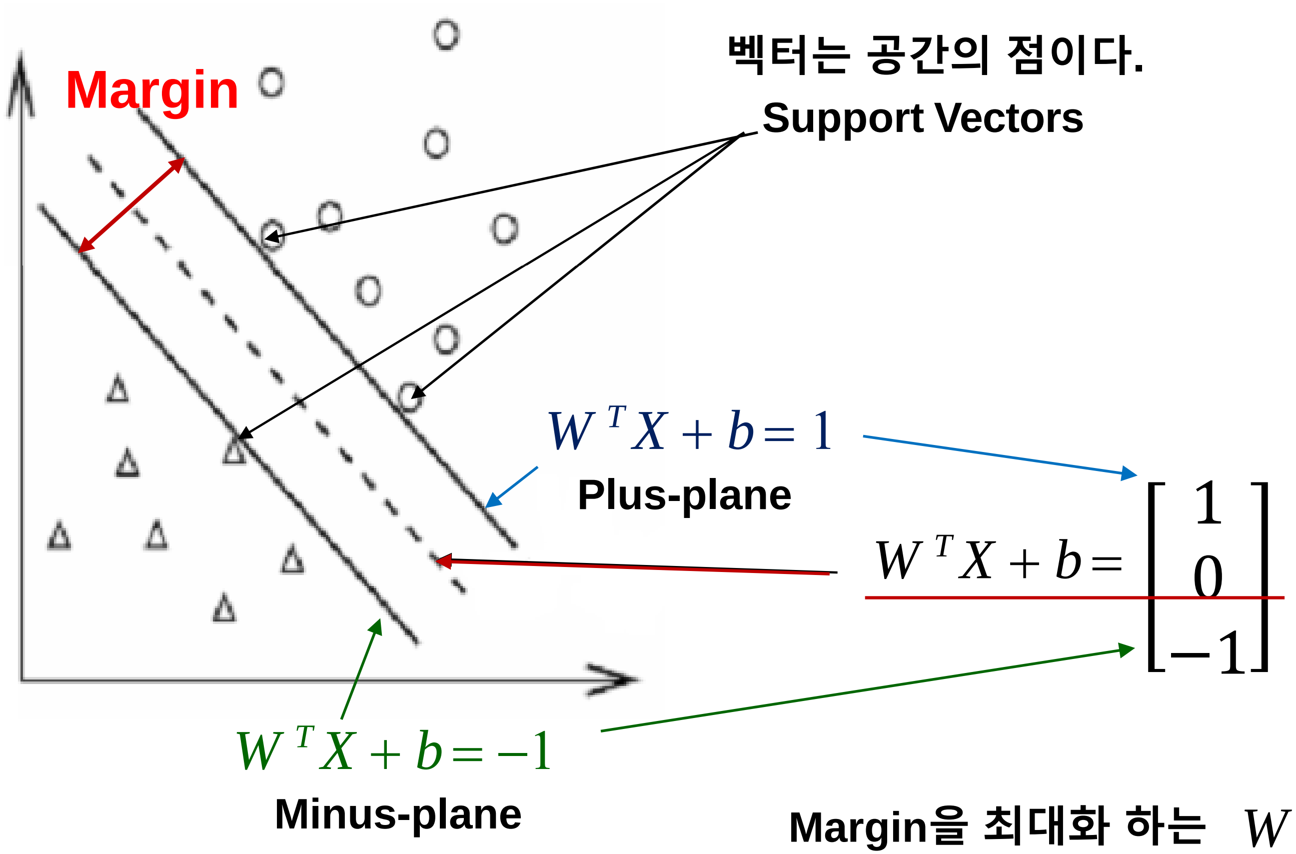
$$\boldsymbol{W^T = [a\ b\ c], X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, d = b}$$

Linear Algebra & Support Vector Machine (SVM)



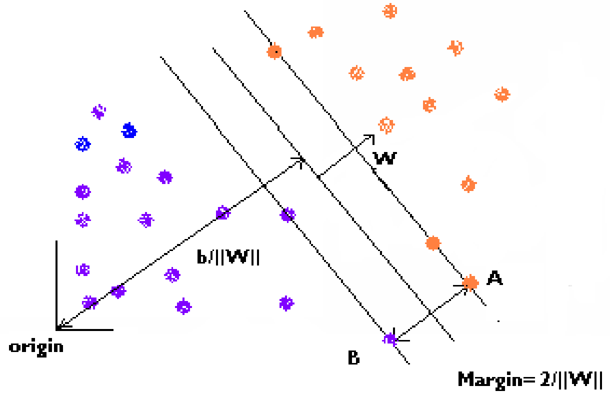
- Two class classification 문제
- 두 class 를 나누는 hyperplane 은 무한히 많음
- 어떤 hyperplane 이 가장 좋은 hyperplane 인가?
- "좋다"는 것의 기준은?
- → Maximizing margin over the training set = good prediction performance
- So What is Margin?

Linear Algebra & Support Vector Machine (SVM)



Margin을 최대화 하는 $W^T X + b = 0$ **을 찾자!**

Linear Algebra & Support Vector Machine (SVM)



$$W^T X + b = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

- Distance of a point (u, v) from Ax+By+C=0, is given by |Ax+By+C|/||n|| Where ||n|| is norm of vector n(A,B)
- Distance of hyperpalne from origin = $\dfrac{b}{\| W \|}$
- Distance of point A from origin = $\dfrac{b+1}{\| W \|}$
- Distance of point B from Origin = $\dfrac{b-1}{\| W \|}$
- Distance between points A and B (Margin) = $\dfrac{2}{\| W \|}$

Linear Algebra & Support Vector Machine (SVM)

$$\max_{W,b} \frac{2}{||W||} = \min_{W,b} \frac{1}{2} W^T W$$

목적식

Such that

$$Y^{(i)}(W^T X^{(i)} + b) \geq 1 \quad for \quad \forall i$$

제약식

Notice: $W^T W = || W ||^2$

It is a convex quadratic optimization problem !

→ We need 1) Linear Algebra and 2) Vector Calculus!