Statistics II

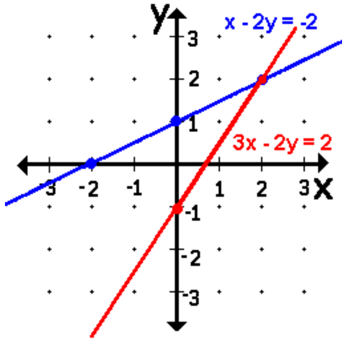
for Machine Learning

Chap. 1: ML위한 선형대수

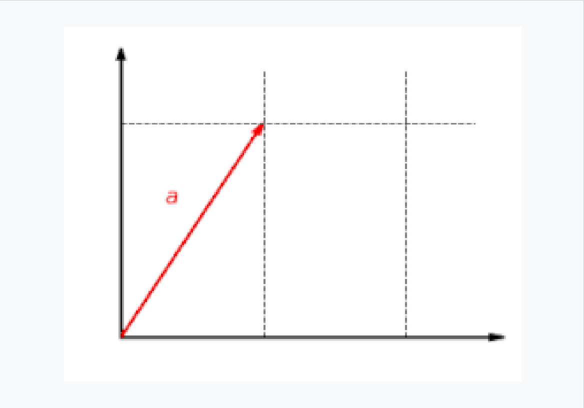
Oh, Hyung Sool

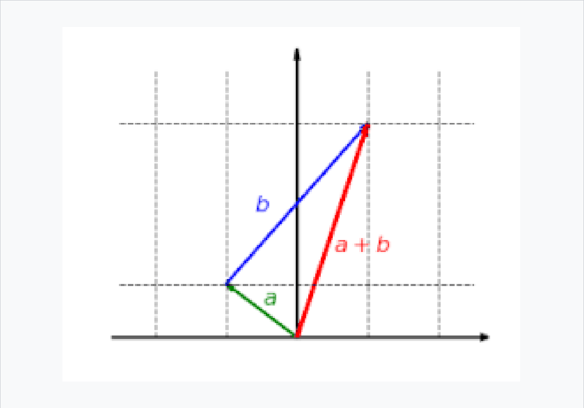
선형대수?

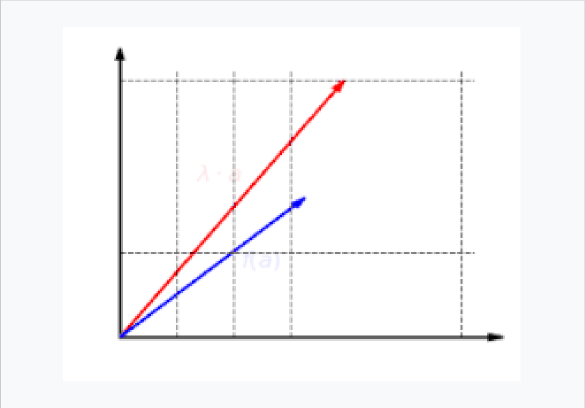
Linear algebra is the branch of mathematics concerning



- **linear equations** such as $a_1x_1 + \ldots + a_nx_n = b$
- **linear functions** such as $(x_1, \ldots, x_n) \mapsto a_1x_1 + \ldots + a_nx_n$





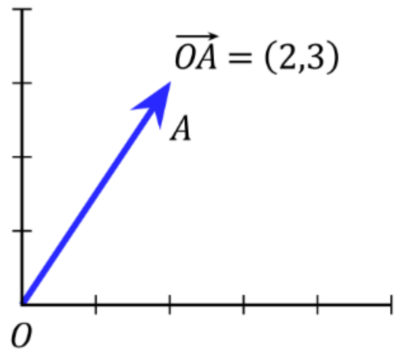


- and their representations in **vector spaces** and through **matrices**.

Vector

A **Euclidean vector** is represented by a line segment connecting an ***initial point*** *O* with a ***terminal point*** *A*.

A **Euclidean vector** (simply a **vector**) is a geometric object that has **magnitude** (or length) and **direction**



벡터는 크기(점수)와 방향(국어, 영어)의 값을 갖는 데이터를 순서대로 나열한 것

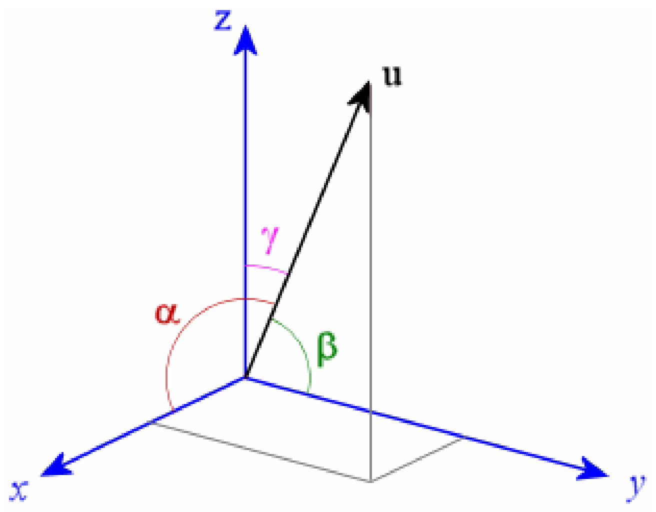
데이터를 벡터로 표현하면, 벡터를 이용하여 처리할 수 있다

$$\text{국어 점수} = \begin{bmatrix} 100 \\ 70 \\ 30 \\ 45 \\ 80 \end{bmatrix} \qquad \text{영어 점수} = \begin{bmatrix} 83 \\ 50 \\ 25 \\ 30 \\ 60 \end{bmatrix}$$

Vector Space

A **vector space** (also called a **linear space**) is a **collection of objects** called vectors, which may be **added** together and **multiplied** ("scaled") by numbers, called **scalars**.

Vector spaces are well characterized by their **dimension**, which specifies the number of **independent directions** in the space.



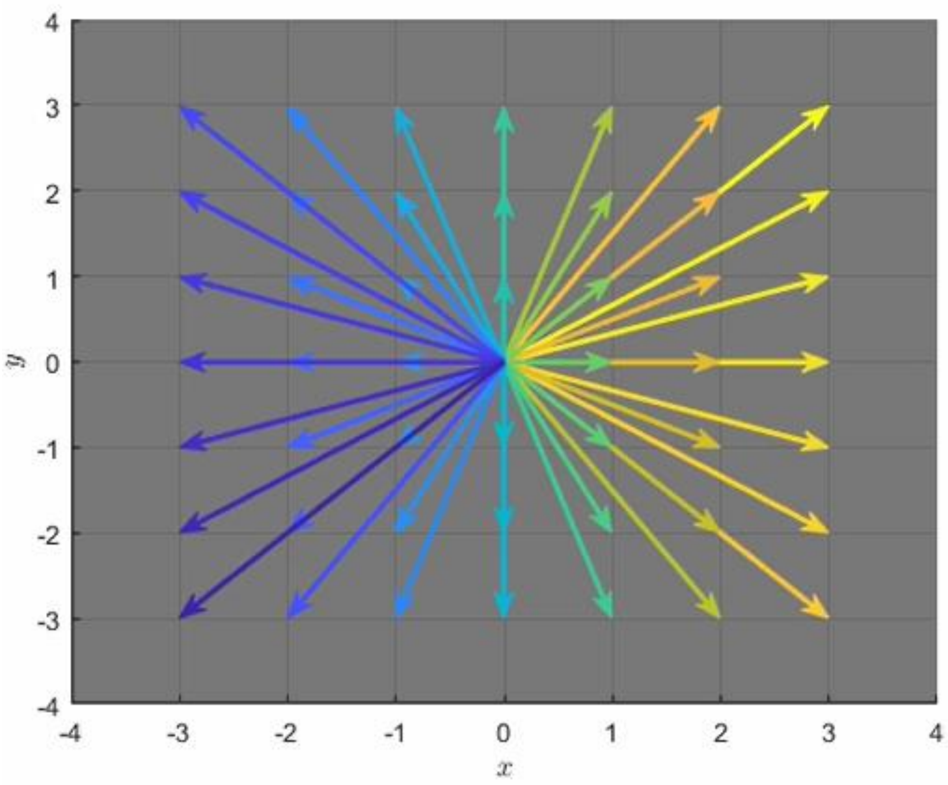
벡터공간(**vector space)** 은 벡터의 선형결합으로 표현되며, 함수를 포함하여 모든 것을 벡터의 선형결합으로 표현할 수 있다

Linear Combination

다음의 C1, C2 는 모든 실수에 대응될 수 있으며, C1, C2 값이 바뀌면서 얻게 되는 선형결합의 결과는 2차원 실수 벡터공간의 모든 벡터에 대응된다.

이때의 두 벡터 [1,0] , [0,1] 는 기저벡터이어야 가능

$$C_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + C_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$



Matrix

A **matrix** (plural **matrices**) is a rectangular array of numbers, symbols or expressions, arranged in rows and columns.

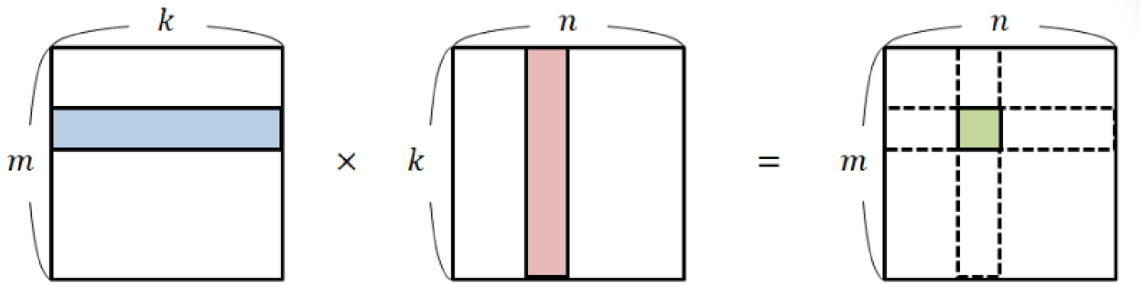
$$A = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array}\!\!\begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & n \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{array} = \in \mathbb{R}^{m \times n} \quad , \qquad \boldsymbol{A}X = \lambda X$$

행렬은 벡터를 또 다른 벡터로 변환 시키는 일종의 **연산자**로 볼 수 있다

A **major application** of matrices is to represent **linear transformations**, that is, generalizations of linear functions.

Matrix Product

행렬과 행렬의 곱 혹은 행렬과 벡터의 곱은 왼쪽 행렬의 행 요소들과 오른쪽 행렬의 열 요소들의 값들을 순서대로 곱해주고 더해주는 방법이다.



행렬 곱은 행 벡터와 열 벡터 간의 내적(inner product)을 계산함으로써 이루어진다.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 \cdot a + 2 \cdot c & 1 \cdot b + 2 \cdot d \\ 3 \cdot a + 4 \cdot c & 3 \cdot b + 4 \cdot d \end{bmatrix}$$

Matrix Product

▶ 열벡터의 선형결합

행렬의 곱을 다음의 수식표현에서처럼 **열 벡터의 선형결합으로** 이해할 수 있다.

벡터의 선형결합이 의미하는 것은 벡터 공간(vector space)의 생성을 의미

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix} = \begin{bmatrix} 1 \cdot a + 2 \cdot c \\ 3 \cdot a + 4 \cdot c \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix} = a \cdot \begin{bmatrix} 1 \\ 3 \end{bmatrix} + c \cdot \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

- 우리가 익숙한 다음의 선형 연립방정식으로 행렬의 곱에 대해 살펴보자

$$\begin{cases} x + 2y = 3 \\ 3x + 4y = 5 \end{cases} \Rightarrow \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix} \Rightarrow x \begin{bmatrix} 1 \\ 3 \end{bmatrix} + y \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

- 위의 열 벡터 선형결합 수식을 해석적 관점에서 다시 표현한다면,

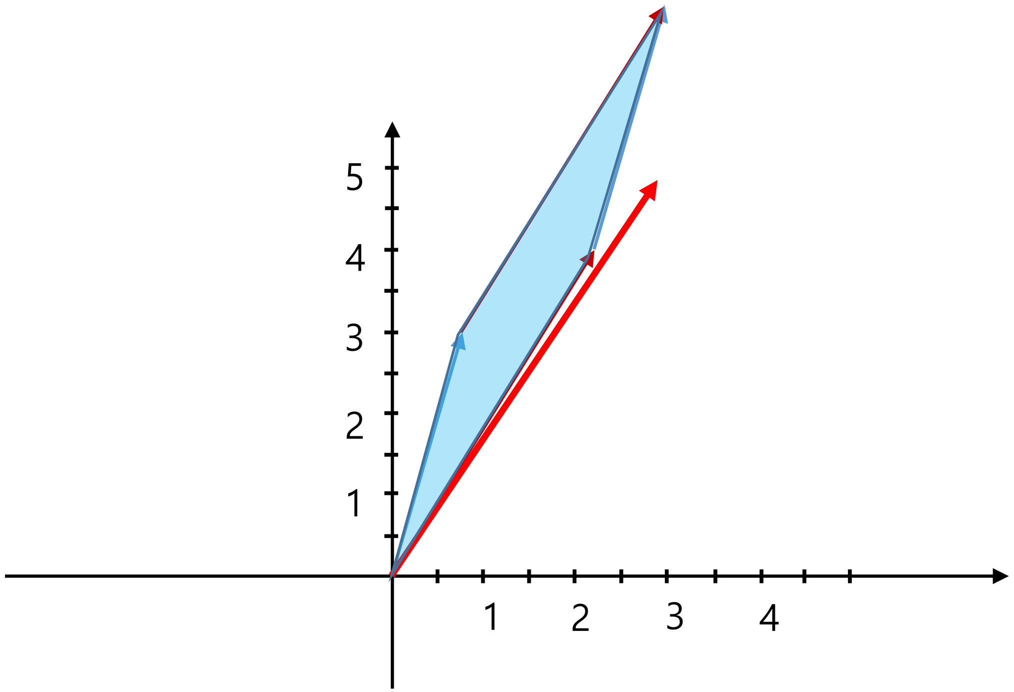
  두 벡터 $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ 과 $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$의 선형결합으로 생성된 (x, y) 벡터공간 내에 $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$가 존재하는가?

  존재한다면, 두 벡터 $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ 과 $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$을 어떻게 조합하여야 하는가? 의 문제로 이해할 수 있다.

Matrix Product

두 벡터 $\begin{bmatrix}1\\3\end{bmatrix}$ 과 $\begin{bmatrix}2\\4\end{bmatrix}$의 선형결합으로 생성된 (x, y) 벡터공간 내에 $\begin{bmatrix}3\\5\end{bmatrix}$가 존재하는가?

존재한다면, 두 벡터 $\begin{bmatrix}1\\3\end{bmatrix}$ 과 $\begin{bmatrix}2\\4\end{bmatrix}$을 어떻게 조합하여야 하는가? 의 문제로 이해할 수 있다.



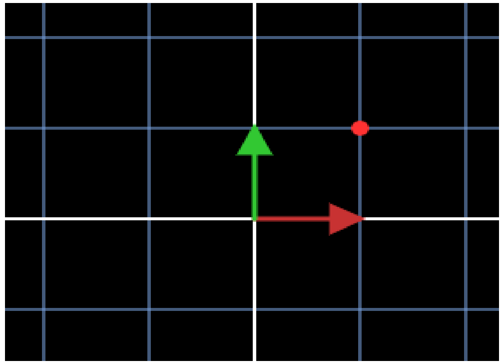
Matrix Product

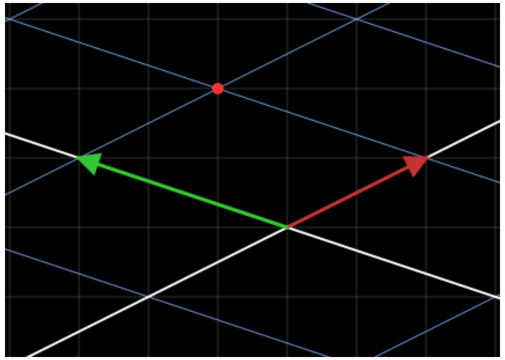
▶ 기저 벡터 변형을 통한 벡터의 선형변환

행렬의 곱 또는 행렬과 벡터의 곱을 기저 벡터의 변형을 통한 **벡터의 선형 변환**으로 해석

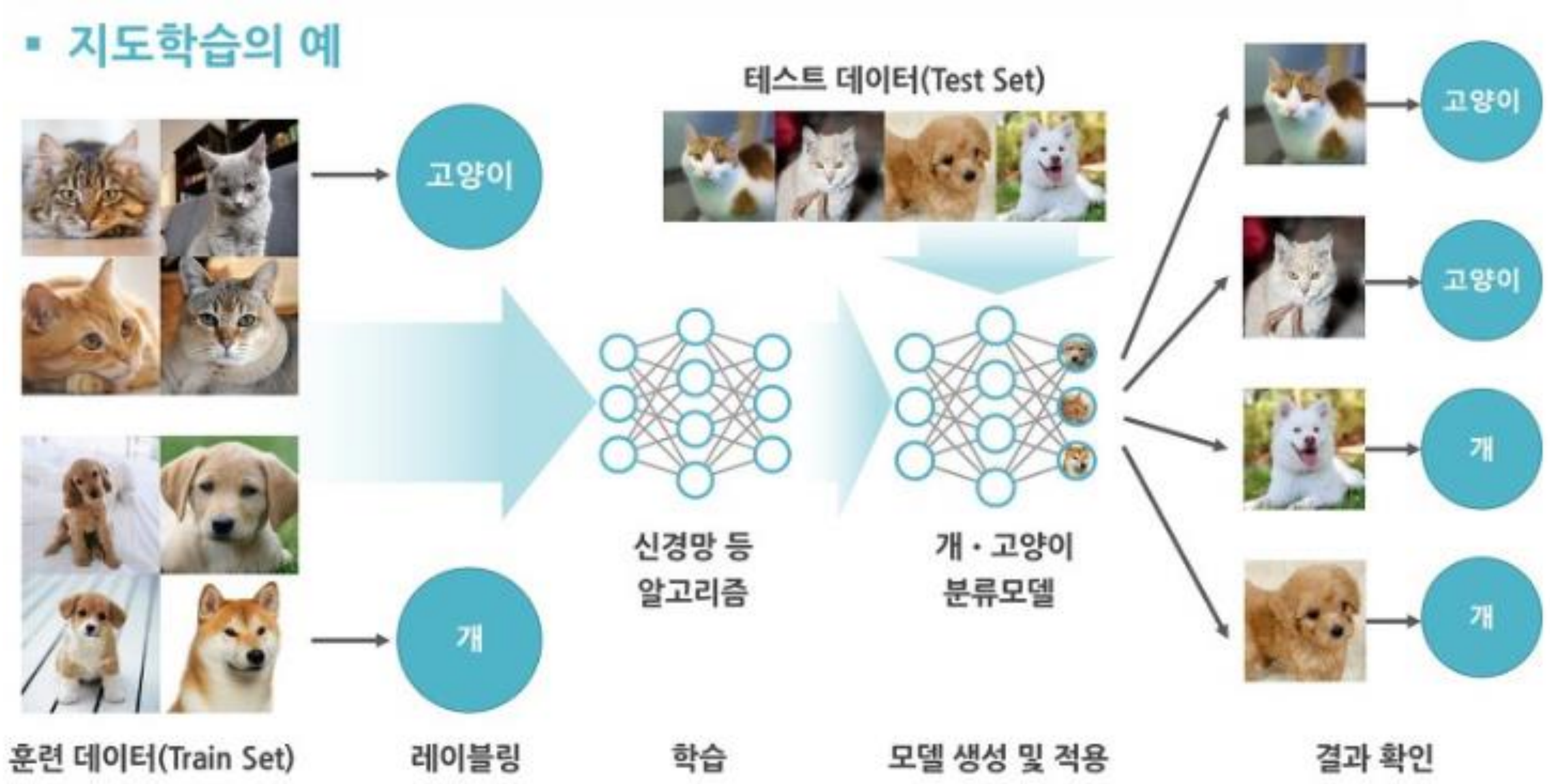
- ex) 다음의 행렬A를 이용하여 기저 벡터 $\vec{x}$ 를 변형

$$A = \begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix} \qquad \vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad A\vec{x} = \begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$





Matrix


▪ 지도학습의 예
테스트 데이터(Test Set)
고양이
개
신경망 등
알고리즘
개 · 고양이
분류모델
고양이
고양이
개
개
훈련 데이터(Train Set)
레이블링
학습
모델 생성 및 적용
결과 확인

Matrix

변수 / 관측치	X_1	…	X_i	…	X_p	Y
N_1	x_{11}	…	x_{1i}	…	x_{1p}	20.5
N_2	x_{21}	…	x_{2i}	…	x_{2p}	22.2
…	…	…	…	…	…	…
N_{n-1}	x_{n-11}	…	x_{n-1i}	…	x_{n-1p}	72.3
N_n	x_{n1}	…	x_{ni}	…	x_{np}	82.8

Linear Combination

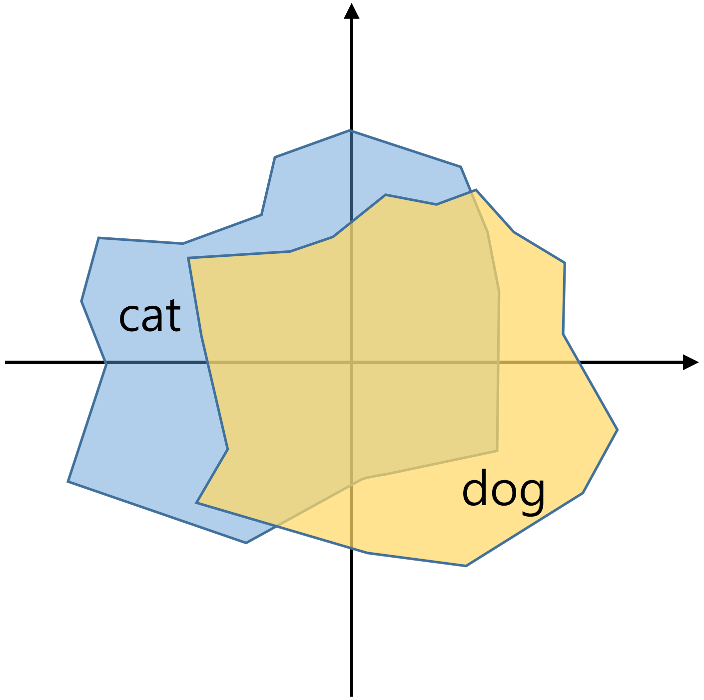
$$C_1\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}+\ C_2\begin{bmatrix}b_1\\b_2\\\vdots\\b_n\end{bmatrix}+\ \ldots+C_k\begin{bmatrix}k_1\\k_2\\\vdots\\k_n\end{bmatrix}$$

$$D_1\begin{bmatrix}\alpha_1\\\alpha_2\\\vdots\\\alpha_l\end{bmatrix}+\ D_2\begin{bmatrix}\beta_1\\\beta_2\\\vdots\\\beta_l\end{bmatrix}+\ \ldots+D_m\begin{bmatrix}\theta_1\\\theta k_2\\\vdots\\\theta_l\end{bmatrix}$$

Linear Combination

$$C_1\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix} + C_2\begin{bmatrix}b_1\\b_2\\\vdots\\b_n\end{bmatrix} + \ldots + C_k\begin{bmatrix}k_1\\k_2\\\vdots\\k_n\end{bmatrix}$$

$$D_1\begin{bmatrix}\alpha_1\\\alpha_2\\\vdots\\\alpha_l\end{bmatrix} + D_2\begin{bmatrix}\beta_1\\\beta_2\\\vdots\\\beta_l\end{bmatrix} + \ldots + D_m\begin{bmatrix}\theta_1\\\theta k_2\\\vdots\\\theta_l\end{bmatrix}$$

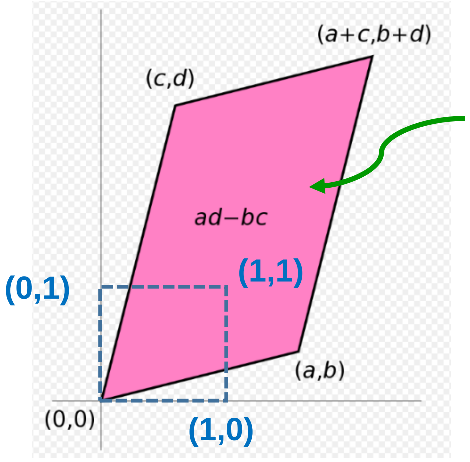


Linear Transformation

For example, the 2×2 matrix, $A = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$

Eigen Vectors, Eigen Values

can be viewed as the transform of the ***unit square*** into a parallelogram with vertices at (0, 0), (*a*, *b*), (*a* + *c*, *b* + *d*), and (*c*, *d*).



determinant of a matrix *A* : det(*A*)

행렬식(determinant)은 해당 벡터로 만들어지는 영역으로서,
해당 벡터의 행렬식 값이 0 ➔ 해당 벡터는 동일한 선상에 존재하는 벡터를 의미

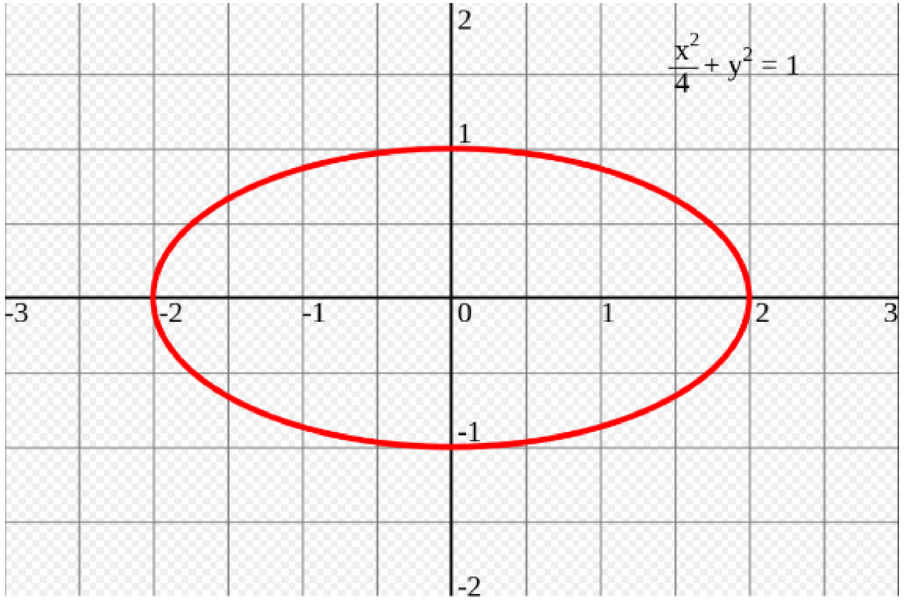
The parallelogram is obtained by multiplying **A** with each of the column vectors.

$\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *and* $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. These vectors define the vertices of the **unit square**.

Linear Transformation

For example, when we apply linear transformation A to a circle

$$A = \begin{bmatrix} 1/4 & 0 \\ 0 & 1 \end{bmatrix}$$



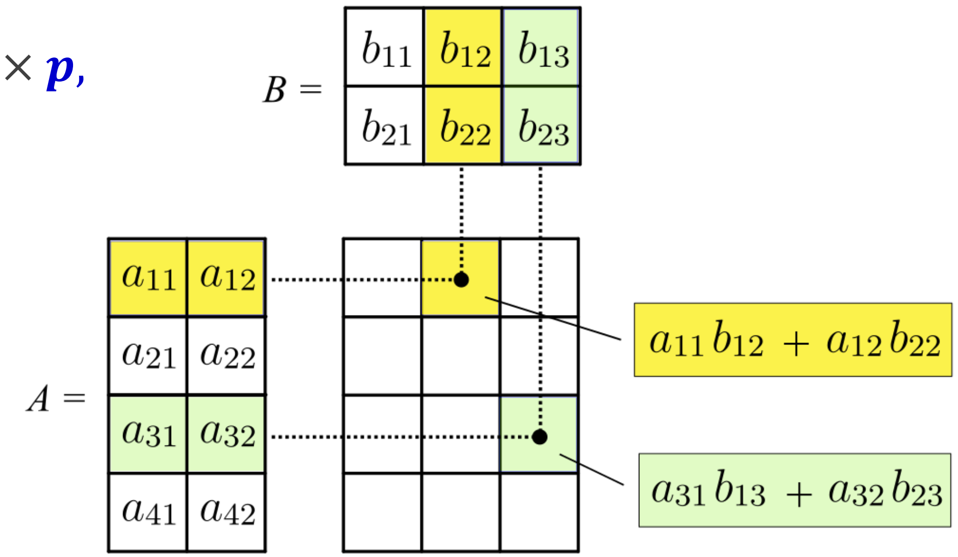
Dot Product

- The **dot product** of the Cartesian coordinates of two vectors is often called the **inner product** (or rarely **projection product**)
- 유사도(similarity) 측정에 사용

● *Algebraic definition*

Two vectors (or matrices) **A** and **B**, A *is* $\boldsymbol{m} \times n$ *and* B *is* $n \times \boldsymbol{p}$,

then the dot product $A \cdot B$ *is* $\boldsymbol{m} \times \boldsymbol{p}$ *matrix*



Dot Product

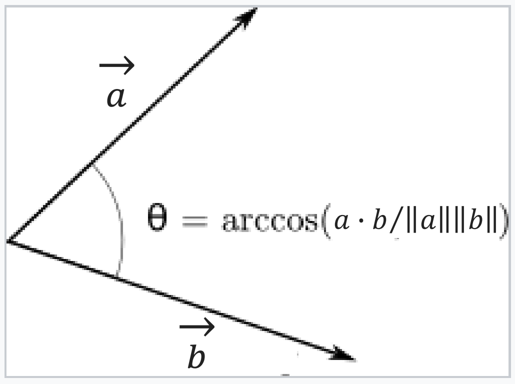
- ***Geometric definition***

Geometrically, the dot product of the two vectors is the **cosine** of the angle between them.

$$a \cdot b = \| a \| \| b \| \, cos\theta$$

correlation

The illustration pictured at right show *how to find the angle* between vectors using the dot product



✓ *If the vectors **a** and **b** are **orthogonal** (their angle is π / 2 or 90°)*

$$a \cdot b = 0, \qquad \because cos90° = 0$$

✓ *If the vectors **a** and **b** are **codirectional** (the angle between them is 0°)*

$$a \cdot b = \|a\| \|b\|, \qquad \because cos0° = 1$$

✓ *The dot product of a vector **a** with itself is*

$$a \cdot a = \|a\|^2 \ , \qquad \|a\| = \sqrt{a \cdot a} \, . \ \ That\ is\ \textbf{Euclidean length}\ of\ the\ vector\ a.$$

Dot Product

- **Data & matrix**

변수 / 관측치	X_1	…	X_i	…	X_p	Y
N_1	x_{11}	…	x_{1i}	…	x_{1p}	20.5
N_2	x_{21}	…	x_{2i}	…	x_{2p}	22.2
…	…	…	…	…	…	…
N_{n-1}	x_{n-11}	…	x_{n-1i}	…	x_{n-1p}	72.3
N_n	x_{n1}	…	x_{ni}	…	x_{np}	82.8

$$\boldsymbol{X} = \boldsymbol{Y}$$

$$\begin{bmatrix} x_{11} & & x_{1p} \\ \vdots & \cdots & \vdots \\ x_{n1} & & x_{np} \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$$

$$X = (x_1, x_2, \dots, x_n)^T$$

$$L_x = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{\boldsymbol{X}^T\boldsymbol{X}}$$

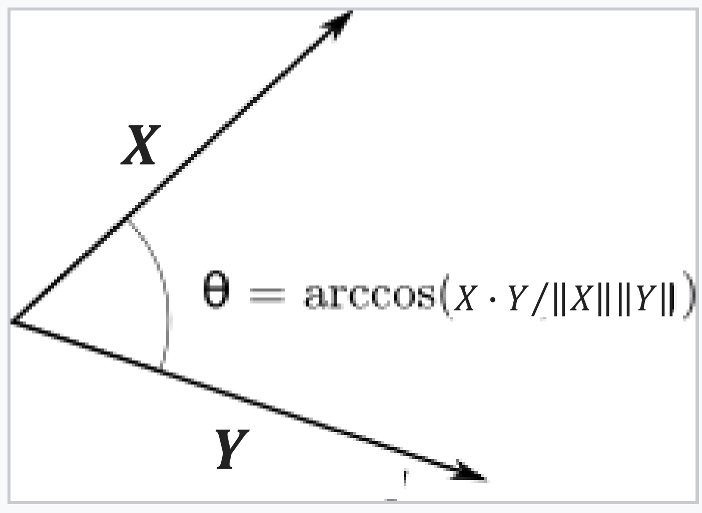
Dot Product

- **Length & Cosine angle**

$$X = (x_1, x_2, \dots, x_n)^T \quad , \quad Y = (y_1, y_2, \dots, y_n)^T$$

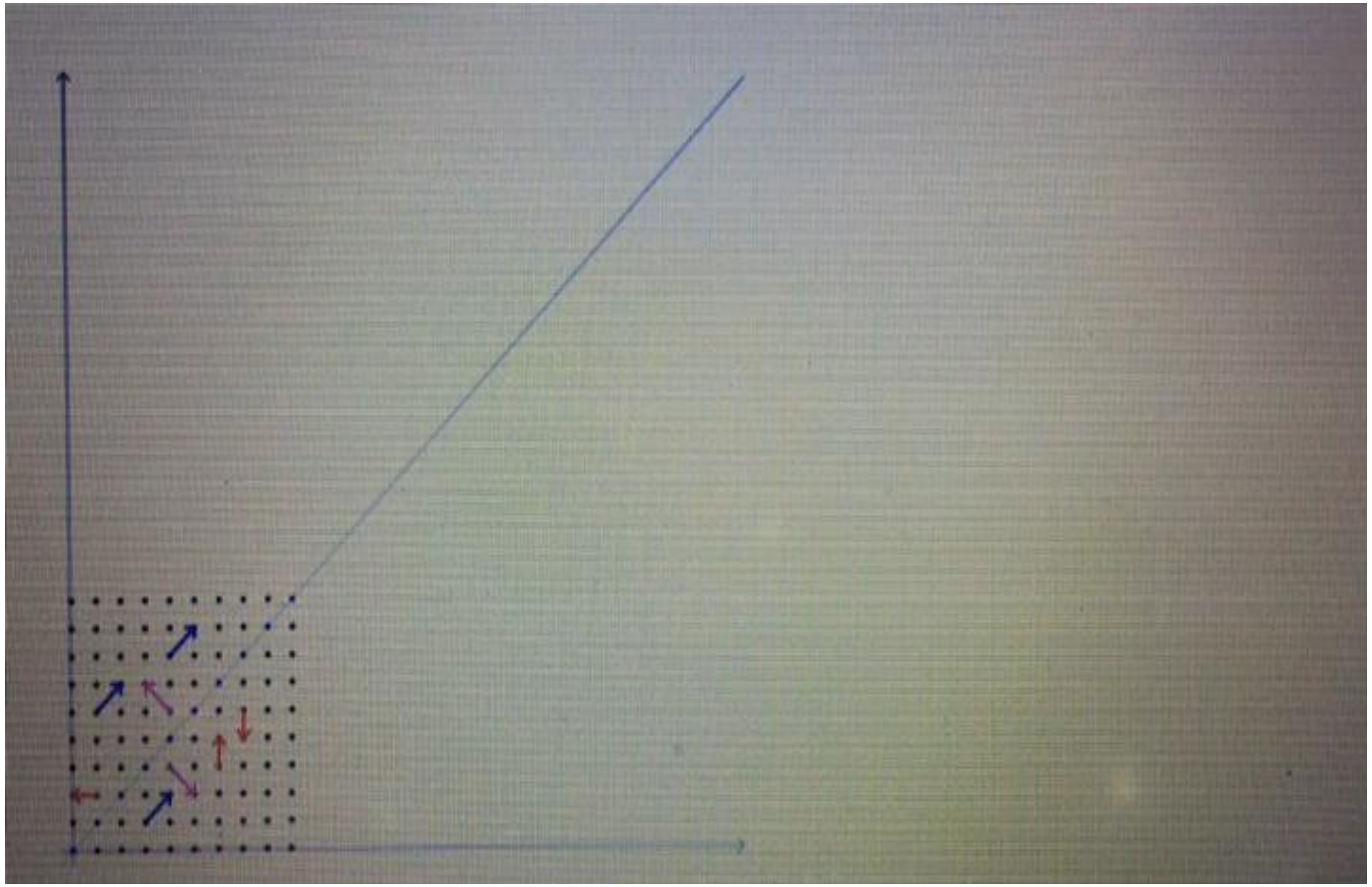
$$< X, Y > = X \cdot Y = X^T Y = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$$

$$cos\theta = \frac{X \cdot Y}{\|X\| \|Y\|} = \frac{X^T Y}{L_x L_y} = \frac{X^T Y}{\sqrt{X^T X}\sqrt{Y^T Y}}$$



선형변환 & 고유벡터

- 행렬 $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ 의 선형변환



선형변환 & 고유벡터

An ***eigenvector*** (or **characteristic vector**) of a *linear transformation* is a nonzero vector that changes at most by a *scalar factor* when that linear transformation is applied to it.

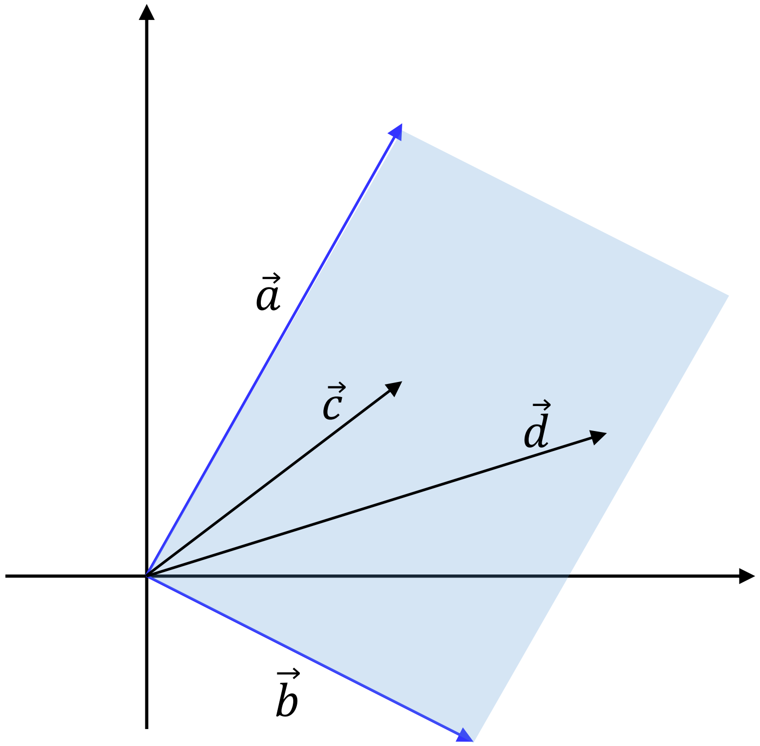
$$\boldsymbol{AX} = \lambda \boldsymbol{X}$$

- $\boldsymbol{A}$: $matrix\ of\ the\ data, ie\ linear\ transformation$
- $\boldsymbol{X}$: $eigen\ vector\ for\ the\ linear\ transformation$
- $\boldsymbol{\lambda}$: $eigen\ value\ for\ the\ linear\ transformation$

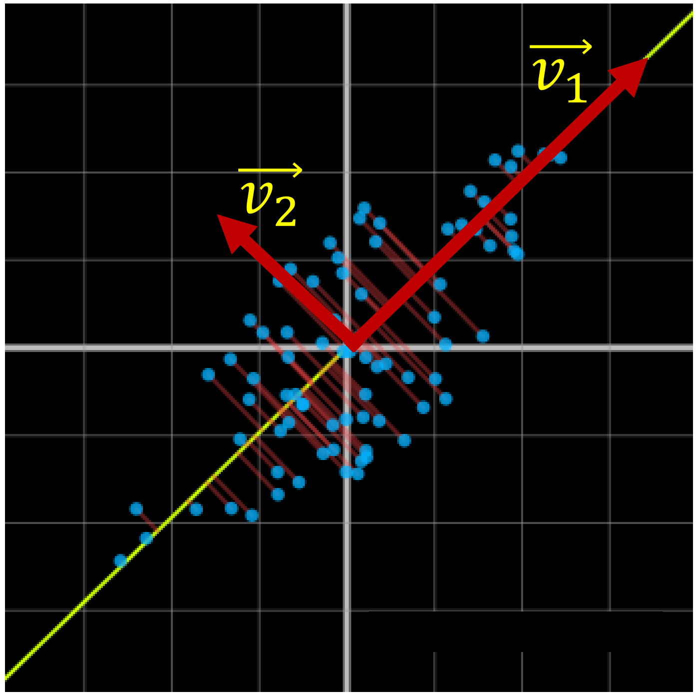
즉, 행렬 $\boldsymbol{A}$ 의 선형변환을 ***eigenvector*** 와 ***eigenvalue*** 성분으로 **분해**한다는 것이다.

특성	기저 벡터	고유벡터
정의	선형공간의 모든 벡터를 선형결합으로 나타낼 수 있는 벡터들의 집합	선형결합에 의해 방향이 변하지 않는 벡터
방향	변하지 않음	변함
크기	변하지 않음	변함(고유값이 변화의 크기임)
선형 독립성	선형독립일 필요 없음	선형독립이어야 함
사용목적	**선형공간의 구조**를 이해하는데 사용	**선형변환(행렬)**을 이해하는데 사용

- 기저 벡터: $\vec{a}$ $\vec{b}$



- 고유 벡터



선형변환 & 고유벡터

- **Eigen Vector & Eigen Value**

The meaning of eigenvectors and eigenvalues for *linear transformation* matrix ***A***

✓ Eigen vector: ***direction*** of the *linear transformation*

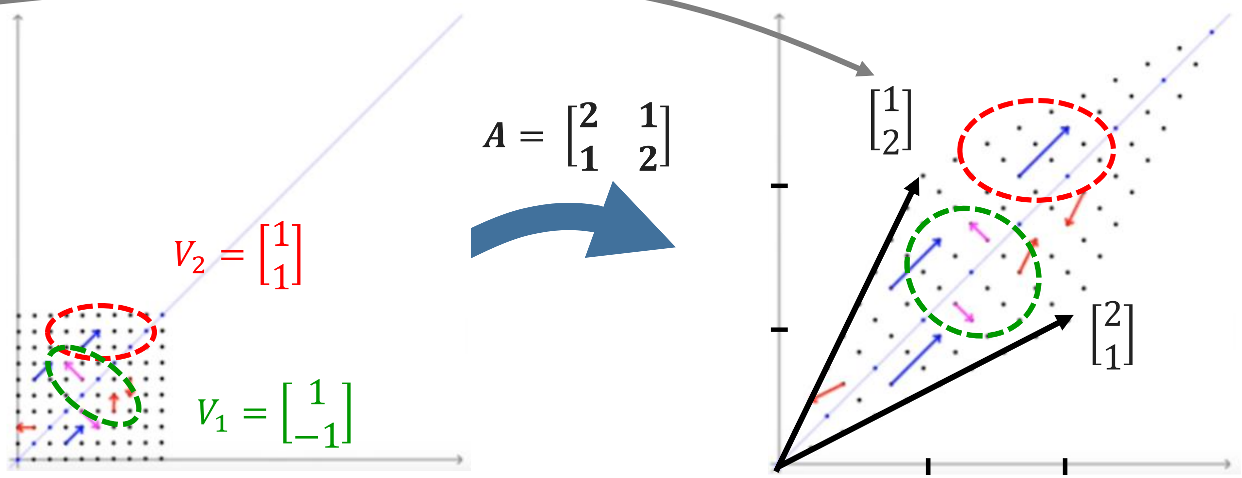
✓ Eigen value: ***magnitude*** of the *linear transformation*

For example,

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \lambda_1 = 1\,,\ \lambda_2 = 3$$

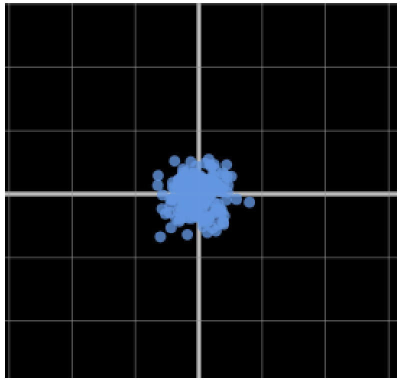
$$\lambda_1 = 1\,, \quad V_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

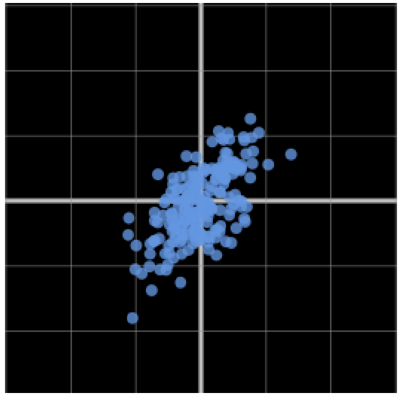
$$\lambda_2 = 3\,, \quad V_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$



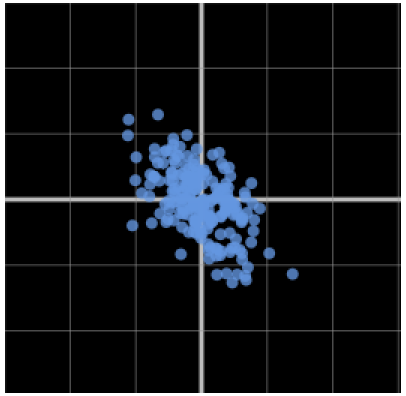
❖ **고유벡터들은,** $\boldsymbol{V} = [V_1\ V_2] = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ **, 서로 직교한다. 방향은 그대로, 크기만 변화.**

선형변환 & 고유벡터

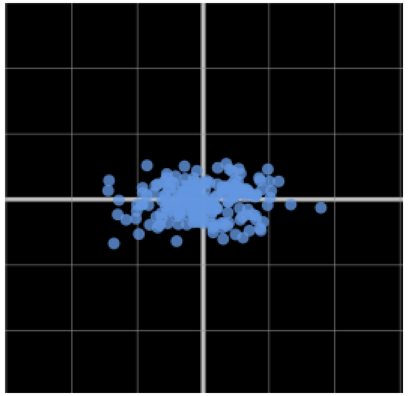




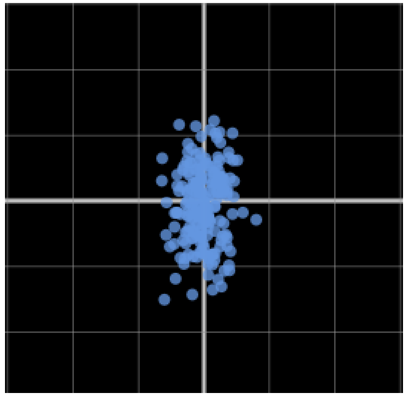
$$\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}$$



$$\begin{bmatrix} 3 & -2 \\ -2 & 4 \end{bmatrix}$$



$$\begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

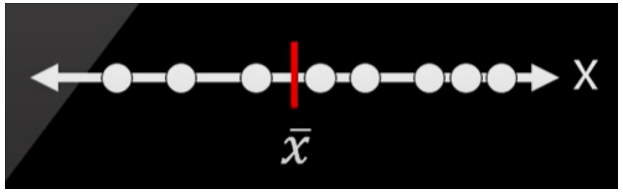
공분산(Covariance)

ID	국어	영어
1	95	95
2	90	95
3	80	75
4	60	70
5	40	35
6	80	80
7	95	90
8	30	25
9	15	10
10	60	70
평균	64.5	64.5
분산	808	925
공분산	762	

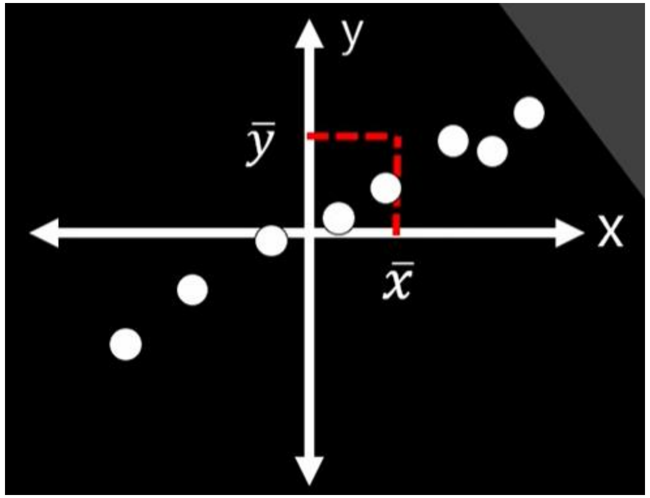
ID	국어	영어
1	9.5	9.5
2	9.0	9.5
3	8.0	7.5
4	6.0	7.0
5	4.0	3.5
6	8.0	8.0
7	9.5	9.0
8	3.0	2.5
9	1.5	1.0
10	6.0	7.0
평균	6.45	6.45
분산	8.08	9.25
공분산	7.62	

- ✓ $Cov(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
- ✓ 공분산: X의 편차와 Y의 편차를 곱한 값들의 평균으로, **데이터**에 따라 공분산 **크기가 달라진다**.
  - Cov(X,Y) > 0 : X↑ , Y↑
  - Cov(X,Y) < 0 : X↑ , Y↓
  - Cov(X,Y) = 0 : No linear relationship
- ✓ 공분산은 두 변수 간에 **양의 상관**관계가 있는지? **음의 상관**관계가 있는지? 정도만 알려줌
- ✓ 상관관계가 얼마나 큰지는 제대로 반영하지 못함

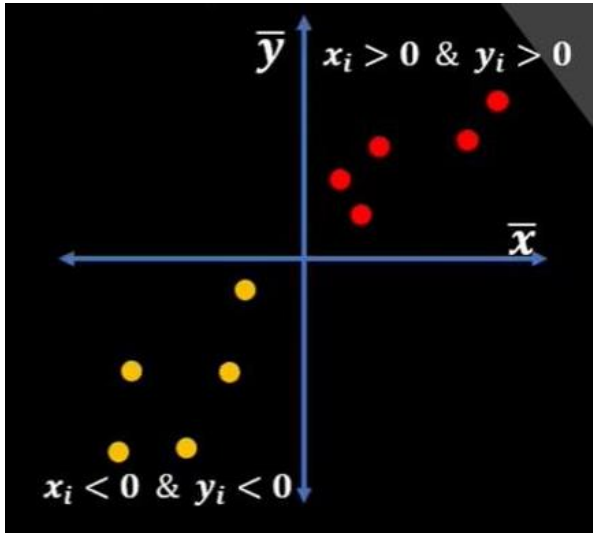
분산 & 공분산



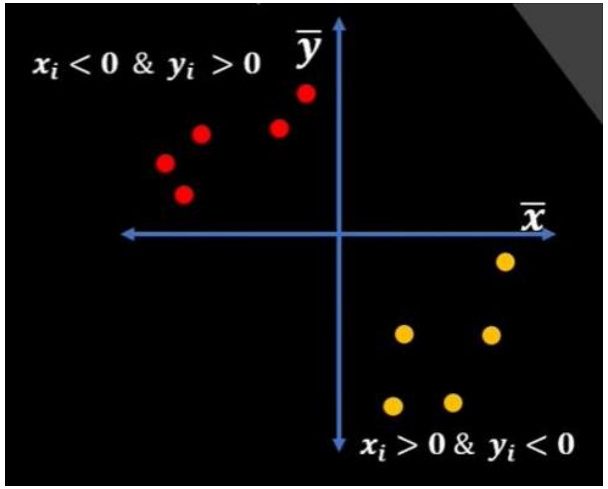
$$Var(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$



$$Cov(x, y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$



$$Cov(x, y) = +$$



$$Cov(x, y) = -$$

상관계수(Correlation Coefficient)

ID	국어	영어
1	95	95
2	90	95
3	80	75
4	60	70
5	40	35
6	80	80
7	95	90
8	30	25
9	15	10
10	60	70
평균	**64.5**	**64.5**
분산	**808**	**925**
공분산	**762**	

$$\gamma = \frac{Cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}} = \frac{762}{\sqrt{808}\sqrt{925}} = 0.88$$

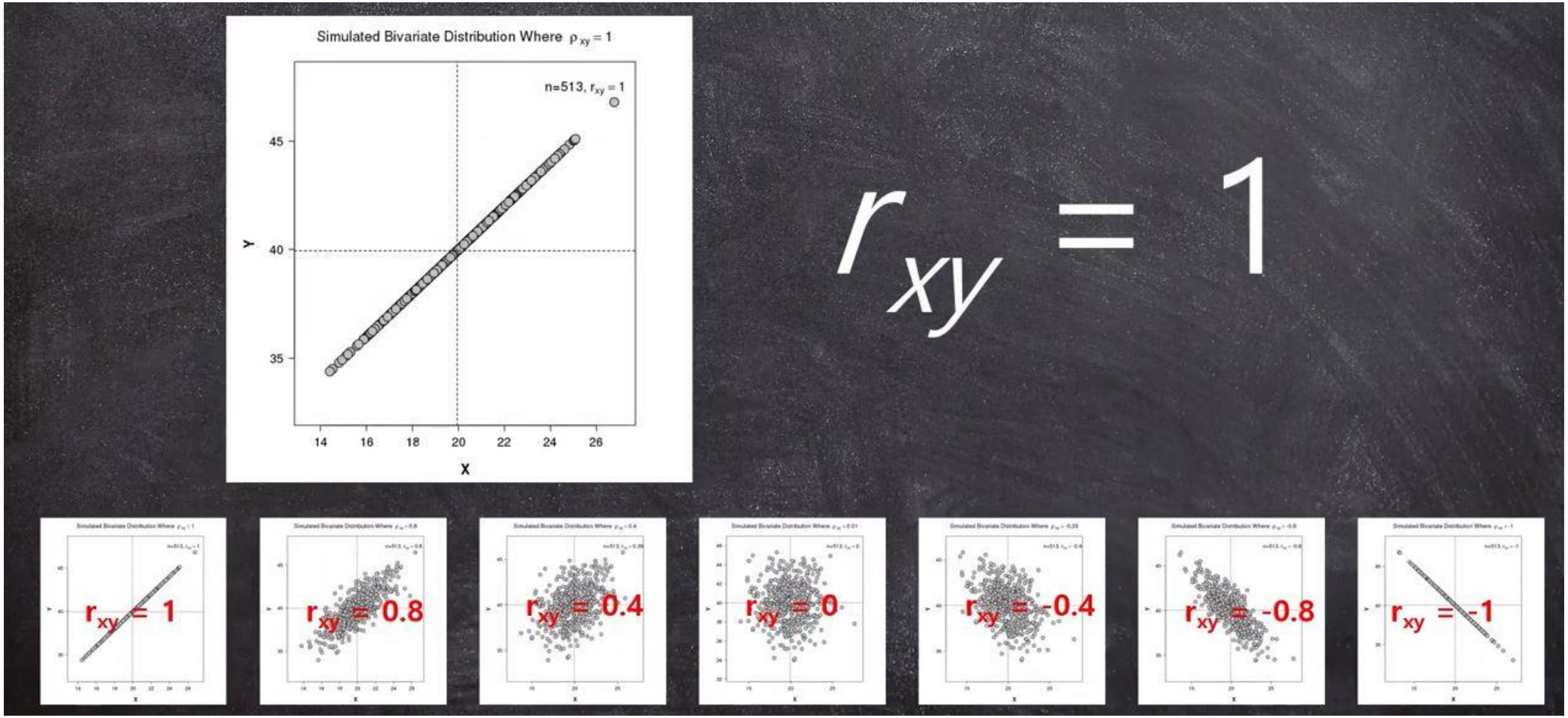
ID	국어	영어
1	9.5	9.5
2	9.0	9.5
3	8.0	7.5
4	6.0	7.0
5	4.0	3.5
6	8.0	8.0
7	9.5	9.0
8	3.0	2.5
9	1.5	1.0
10	6.0	7.0
평균	**6.45**	**6.45**
분산	**8.08**	**9.25**
공분산	**7.62**	

$$\gamma = \frac{Cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}} = \frac{7.62}{\sqrt{8.08}\sqrt{9.25}} = 0.88$$

상관계수(Correlation Coefficient)

- $$\gamma = \frac{Cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}}$$
- 상관계수: **공분산을 표준화한 값**
- 절대값은 1보다 작거나 같음
- X, Y가 완벽한 선형관계이면, γ =1 or -1
- 상관계수가 1 또는 -1에 근접할수록 힘(power)가 크다는 것
- 여기서의 힘이란 점들이 모여있는 정도

상관계수(Correlation Coefficient)



내적 & 상관계수

$$Cov(X,Y) = E(X-\mu_X)(Y-\mu_Y) = \boxed{E[XY]} - \mu_X\mu_Y$$
$$= \frac{1}{N}\boxed{\sum_{i=1}^{N} X_iY_i} - \mu_X\mu_Y$$
$$= \frac{1}{N}\boxed{< X,Y >} - \mu_X\mu_Y$$

$$\boldsymbol{\rho} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i - \mu_X}{\sigma_X}\right)\left(\frac{Y_i - \mu_Y}{\sigma_Y}\right)$$

$if\ \mu_X \,\&\, \mu_Y = 0$

$$= \frac{1}{N}\sum_{i=1}^{N}\frac{X_iY_i}{\sigma_X\sigma_Y} = \frac{< X,Y >}{\sqrt{\sum X^2}\sqrt{\sum Y^2}} = \frac{< \boldsymbol{X},\boldsymbol{Y} >}{\|\boldsymbol{X}\|\|\boldsymbol{Y}\|}$$
$$= cos\theta$$

- ✓ 내적 $a \cdot b$ ($a^T b$, $<a, b>$) 결과 행렬로부터 공분산을 구하기 위해서는 원래의 data matrix에서 평균이 zero가 되도록 **표준화** 해주어야 한다.
- ✓ 상관계수 $\boldsymbol{\rho}$ 는 두 벡터 간의 $cos\theta$ 값이 되기 때문에 $-1 \leq \boldsymbol{\rho} \leq \mathbf{1}$ 갖게 된다.
- ✓ **표준화** 한 경우에는, 상관계수 $\boldsymbol{\rho}$ = 내적의 $\boldsymbol{cos\theta}$ 성립하며,

$$\boldsymbol{Cov(X,Y) = < X,Y >}$$