* 주의: axis랑 p랑 무관하다.

원점으로부터의 거리...?

① norm의 정의 (수학적으로)

norm은 기본적으로 벡터공간의 함수이다.

$\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}$

벡터공간의 예제) $\mathbb{R}^2$, $\mathbb{R}^3$, 2×2 matrix 모여둔 공간 등..

$\|\cdot\|$ (norm) 이 함수를 f라고 하면

$f: V \to [0, \infty)$

i) $f(ax) = af(x)$
ii) $f(x+y) \le f(x) + f(y)$
iii) $f(x)=0 \iff x=0$
iv) $f(x) \ge 0$

이 조건을 다 만족시키면 'norm 함수'이다.

② Lp-norm 이란?

수많은 norm들 중에서 하나...

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$$

P=1

$\boxed{p: 0 \sim \infty}$

$x = (x_1, x_2, \cdots x_n)$

$p=2$

$$= \sqrt{(x_1)^2 + |x_2|^2 + \cdots |x_n|^2}$$

$(0,0,\cdots 0) \longleftrightarrow (x_1, x_2, \cdots x_n)$

P=1 $d = |x_1| + |x_2| + |x_3| + \cdots |x_n|$

Lp-norm의 정의 (암기하자!)

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$$

그리고 L0, L1, L2, L∞ norm은 꼭 알아야 함!

L1-norm

$\uparrow$

p

따라서,

그리고 $x = (x_1, x_2, \cdots x_n)$라고 하면

Lo-norm : $\|x\|_0 = \left(|x_1|^0 + |x_2|^0 + \cdots |x_n|^0\right)^{\frac{1}{0}}$ $\frac{1}{0} = \infty$?

(복잡) . . . .

≒ 0이 아닌것의 갯수!

(0과 ∞는 정의가 잘 안되어서

약속으로 처리)

ex) $d = (1, 0, 1)$의 Lo-distance = 2

$(0, 0, 2)$ $\xrightarrow{d_2}$ 1

L1-norm : $\|x\|_1 = \left(|x_1| + |x_2| + \cdots |x_n|\right)$

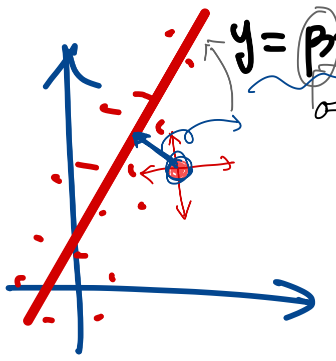
= 절댓값의 합!

L2-norm : $\|x\|_2 = \sqrt{(x_1)^2 + \cdots (x_n)^2}$

= 우리가 아는 일반적인 길이.

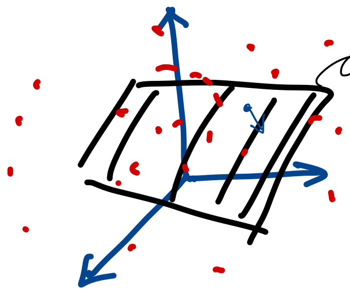
L∞-norm : $\|x\|_\infty = \left(|x_1|^\infty + \cdots |x_n|^\infty\right)^{\frac{1}{\infty}}$

$\doteqdot \max(x_k)$



$y = \beta x + k$

여기서 β를 찾는건 쉽지만...



$ax + by + cz + d = 0$

즉, β는 변수가 3개.

⇒ 상수가 아닌 행렬로 문제를 풀 필요를 느낀다.

$$\beta = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

선형회귀의 목적 → L2 - distance의 최소화.

*가정: 현재 d차원이다. 즉, $\beta = (\beta_1, \beta_2, \beta_3, \cdots \beta_k, \cdots \beta_d)$

원래 data들은 $\mathbf{y}$로 정의된다. $y = (y_1, y_2 \cdots y_d)$

예상한 평면위의 점은 $X\beta$로 표현된다.

$$L2\text{-distance} = \| y - X\beta \|_2$$

$$(L2\text{-distance}: \|x\|_2 = (|x_1|^2 + \cdots |x_d|^2)^{\frac{1}{2}} = \left(\sum_{k=1}^{d} |x_k|^2\right)^{\frac{1}{2}}$$

$$= \left(\sum_{k=1}^{d} (y_k - X_k\beta_k)^2\right)^{\frac{1}{2}}$$

β_k에 대해 편미분하면, $\dfrac{\partial\left(\sqrt{(y_1 - x_1\beta_1)^2 + \cdots + (y_k - x_k\beta_k)^2 + \cdots + (y_d - x_d\beta_d)^2}\right)}{\partial(\beta_k)}$

β=N개

$(x^a)' = (a x^{a-1})$ 이므로,

$$\frac{\partial L}{\partial \beta_k} = \cancel{\tfrac{1}{2}}\left(\|y - X\beta\|^2\right)^{-\frac{1}{2}} \times \cancel{2}(y_k - x_k\beta_k) \times (-x_k)$$

우리가 구하는 값은 임의의 β_k에 대한 편미분이 아니라 행렬 β로 편미분이므로

확장하면,

$$\frac{\partial L}{\partial \beta} = \left(\frac{\partial L}{\partial \beta_1}, \frac{\partial L}{\partial \beta_2}, \cdots \frac{\partial L}{\partial \beta_d}\right)$$

$$= \left(\frac{-x_1(y_1 - x_1\beta_1)}{\|y - X\beta\|_2}, \frac{-x_2(y_2 - x_2\beta_2)}{\|y - X\beta\|_2}, \cdots \frac{-x_d(y_d - x_d\beta_d)}{\|y - X\beta\|_2}\right)$$

$$= \left(\frac{-X^T(y - X\beta)}{\|y - X\beta\|_2}\right)$$

* N은 차원이 커지며 항이 커지는걸 방지하는 효과가 있다! (normalize).