2.5 Norms

- 어떤 벡터의 크기를 표현하고자 할 때, 'norm'이라는 '함수'를 사용.
- L^p norm

$$||\boldsymbol{x}||_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}} \tag{2.30}$$

for $p \in \mathbb{R}, p \geq 1.$

- 수학적으로 엄밀하게는, 아래 조건을 만족시키는 함수 f를 'norm'이라 함.
  - $f(\boldsymbol{x}) = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}$
  - $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$ (the *triangle inequality*)
  - $\forall \alpha \in \mathbb{R}, f(\alpha \boldsymbol{x}) = |\alpha| f(\boldsymbol{x})$
- L^2 norm (==Euclidean norm)은 많이 쓰여서 아래처럼 첨자값(==2) 없이 많이 표기함.

$$||\boldsymbol{x}||$$

  - L^2 norm의 제곱 표현이 기계학습 알고리즘에서 자주 쓰임.

$$||\boldsymbol{x}||\,||\boldsymbol{x}|| = \boldsymbol{x}^\top \boldsymbol{x}$$

  - L^2 norm의 제곱이 L^2 norm 자체보다 수학적으로 더 편리함.(ex. 편미분했을 때 전체 항과의 의존성 관계가 제곱 표현에서는 없어짐).

- L^1 norm은 0항과 non-0 항 구분이 중요할 때 사용.

$$||\boldsymbol{x}||_1 = \sum_i |x_i|.$$

  - 벡터에서 non-0 원소들의 개수 대신 쓰이기도 함.

- L^∞ norm 또는 “max norm”

$$||\boldsymbol{x}||_\infty = \max_i |x_i|.$$

- 행렬의 크기(Frobenius norm)

$$||A||_F = \sqrt{\sum_{i,j} A_{i,j}^2}$$

- 벡터 내적과 L^2 norm

$$\boldsymbol{x}^\top \boldsymbol{y} = ||\boldsymbol{x}||_2 ||\boldsymbol{y}||_2 \cos\theta$$

2.6 특수 벡터와 행렬들

- 대각행렬(diagonal matrix)
  - 행렬의 대각 성분들을 제외한 나머지 성분이 모두 0인 행렬

$$\begin{bmatrix} -3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}$$

  - v^T=[-3 2 0.2]일 때 위 행렬을 $diag(v)$와 같이 표기하기도 함.
- 대칭행렬(symmetric matrix)

$$\boldsymbol{A} = \boldsymbol{A}^\top$$
$$\boldsymbol{A}_{i,j} = \boldsymbol{A}_{j,i}$$

- 단위 벡터(unit vector)

$$||\boldsymbol{x}||_2 = 1$$

- 벡터의 직교(orthogonal)

$$\boldsymbol{x}^\top \boldsymbol{y} = 0$$

- 정규직교(orthonormal): 벡터들이 모두 단위 벡터이면서 서로서로 직교관계인 경우

- 직교행렬(orthogonal matrix): 행들이 서로 정규직교이면서 열들도 서로 정규직교인 정방(n x n)행렬

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}$$
$$\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$$