Similarity Metrics

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

Manhattan Distance (L1 distance)

$$\mathrm{d}_{\mathrm{L1}}(w, v) = \sum_{i=1}^{d} |w_i - v_i|, \text{ where } w, v \in \mathbb{R}^d.$$

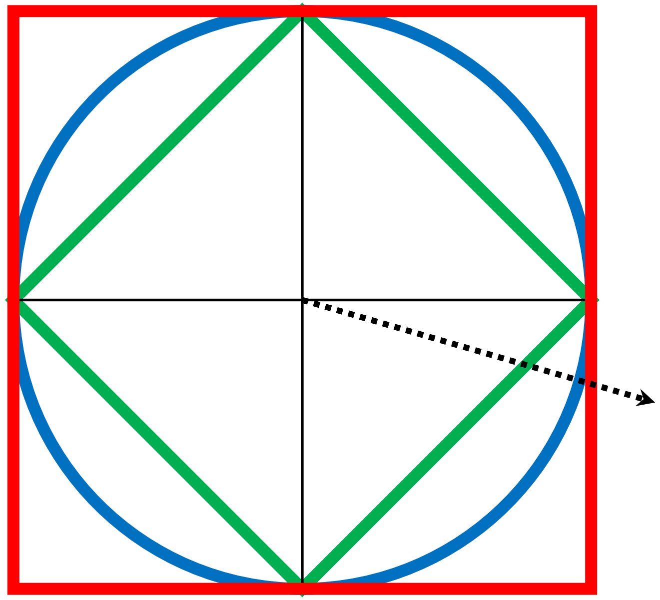
Euclidean Distance (L2 distance)

$$\mathrm{d}_{\mathrm{L2}}(w, v) = \sqrt{\sum_{i=1}^{d} (w_i - v_i)^2}, \text{ where } w, v \in \mathbb{R}^d.$$

Infinity Norm

$$d_\infty(w, v) = \max(|w_1 - v_1|, |w_2 - v_2|, \cdots, |w_d - v_d|), \text{ where } w, v \in \mathbb{R}^d$$

L1, L2 and Infinity



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

Green	L1 Norm
Blue	L2 Norm
Red	Infinity Norm

Cosine Similarity

$$\mathrm{sim}_{\cos}(w,v) = \overbrace{\frac{w \cdot v}{|w||v|}}^{\text{dot product}} = \overbrace{\frac{w}{|w|}}^{\text{unit vector}} \cdot \frac{v}{|v|}$$

$$= \frac{\sum_{i=1}^{d} w_i v_i}{\sqrt{\sum_{i=1}^{d} w_i^2}\sqrt{\sum_{i=1}^{d} v_i^2}}$$

$$\text{where } w, v \in \mathbb{R}^d$$

Summary

- L1, L2 Norm과 Infinity Norm은 강조하고자 하는 것에 따라 사용

- Cosine Similarity는 벡터의 방향을 중요시 함
    - Feature vector의 각 차원의 상대적인 크기가 중요할 때 사용