

Date: _____

M T W T F S S
☐ ☐ ☐ ☐ ☐ ☐ ☐

Cosine Similarity :

Row vectors :

$$s1 = [1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0]$$

$$s2 = [1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0]$$

$$s3 = [2, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1]$$

$$\text{cos-similarity}(s1, s2) = \frac{s1 \cdot s2}{|s1| |s2|}$$

$$\text{cos-similarity}(s1, s3) = \frac{s1 \cdot s3}{|s1| \cdot |s3|}$$

$$\text{cos-similarity}(s2, s3) = \frac{s2 \cdot s3}{|s2| |s3|}$$

$$\begin{aligned} |s1| &= \sqrt{(1)^2 + (2)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2} \\ &= \sqrt{1 + 4 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1} = \sqrt{14} = 3.742 \end{aligned}$$

$$\begin{aligned} |s2| &= \sqrt{(1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (0)^2 + (1)^2 + (0)^2 + (0)^2 + (1)^2 + (1)^2 + (0)^2 + (0)^2 + (0)^2} \\ &= \sqrt{9} = 3 \end{aligned}$$

$$\begin{aligned} |s3| &= \sqrt{(2)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2 + (1)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2 + (1)^2 + (1)^2 + (0)^2 + (0)^2} \\ &= \sqrt{8} = 2.828 \end{aligned}$$

$$\begin{aligned} s1 \cdot s2 &= 1 + 2 + 1 + 1 + 1 + 1 + 0 + 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 \\ &= 8 \end{aligned}$$

$$\begin{aligned} s1 \cdot s3 &= 2 + 0 + 0 + 0 + 0 + 0 + 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 \\ &= 3 \end{aligned}$$

$$\begin{aligned} s2 \cdot s3 &= 2 + 0 + 0 + 0 + 0 + 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 \\ &= 3 \end{aligned}$$

Date: _____

M T W T F S S
☐ ☐ ☐ ☐ ☐ ☐ ☐

$$\cos_similarity(s1, s2) = \frac{8}{(3.742)(3)} = 0.713$$

$$\cos_similarity(s1, s3) = \frac{3}{(3.742)(2.828)} = 0.283$$

$$\cos_similarity(s2, s3) = \frac{3}{(3)(2.828)} = 0.3536$$

Manhattan dist :

$$M_dist(s1, s2) = \sum |s1 - s2|$$

$$= 0 + 1 + 0 + 0 + 0 + 0 + 1 + 0 + 1 + 1 + 1 + 0 + 0 + 0 + 0$$

$$= 7$$

$$M_dist(s1, s3) = \sum |s1 - s3|$$

$$= 1 + 2 + 1 + 1 + 1 + 0 + 1 + 1 + 1 + 1 + 0 + 0 + 1 + 0 + 1$$

$$= 14$$

$$M_dist(s2, s3) = 1 + 2 + 1 + 1 + 0 + 0 + 1 + 0 + 0 + 0 + 1 + 1 + 1 + 0 + 0$$

$$= 11$$

Euclidean dist :

$$E_dist(s1, s2) = \sqrt{0^2 + 1^2 + 0^2 + 0^2 + 0^2 + 0^2 + 1^2 + 0^2 + 1^2 + 1^2 + 1^2 + (-1)^2 + (-1)^2 + 0^2 + 0^2}$$

$$= \sqrt{7}$$

$$= 2.646$$

$$E_dist(s1, s3) = \sqrt{(-1)^2 + (2)^2 + (1)^2 + (1)^2 + (1)^2 + 0^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + 0^2 + 0^2 + 0^2 + 0^2}$$

$$= \sqrt{16}$$

$$= 4$$

Date: _____

M T W T F S S
☐ ☐ ☐ ☐ ☐ ☐ ☐

$$E_dist(S2, S3) = \sqrt{(-1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + 0 + 0 + (1)^2 + 0 + 0 + 0 + (1)^2 + (1)^2 + (-1)^2 + (-1)^2 + (-1)^2}$$

$$= \sqrt{11} = 3.3166$$

Cosine Similarity:

tf-idf Matrix:

$$|S1| = \sqrt{0 + (0.03)^2 + (0.015)^2 + (0.015)^2 + (0.015)^2 + 0 + (0.04)^2 + (0.015)^2 + (0.04)^2 + (0.04)^2 + (0.04)^2}$$

$$= \sqrt{(9 \times 10^{-4}) + (9 \times 10^{-4}) + (6.4 \times 10^{-3})} = 0.090$$

$$|S2| = \sqrt{0 + (0.02)^2 + (0.02)^2 + (0.02)^2 + (0.02)^2 + 0 + 0 + (0.02)^2 + 0 + 0 + 0 + (0.053)^2 + (0.053)^2}$$

$$= \sqrt{(2 \times 10^{-3}) + (5.618 \times 10^{-3})} = 0.087$$

$$|S3| = \sqrt{(0.08)^2 + (0.08)^2 + (0.08)^2}$$

$$= \sqrt{0.0192} = 0.138$$

$$S1.S2 = 0 + 6 \times 10^{-4} + 3 \times 10^{-4} + 3 \times 10^{-4} + 3 \times 10^{-4} + 0 + 0 + 3 \times 10^{-4} + 0$$

$$= 1.8 \times 10^{-3}$$

$$S1.S3 = 0$$

$$S2.S3 = 0$$

$$\cos_similarity(S1, S2) = \frac{1.8 \times 10^{-3}}{0.090 \times 0.087} = 0.229$$

Date: _____

M T W T F S S
☐ ☐ ☐ ☐ ☐ ☐ ☐

$$\text{cos-similarity}(s1, s3) = \frac{s1 \cdot s3}{|s1| |s3|} = \frac{0}{0.090 \times 0.138} = 0$$

$$\text{cos-similarity}(s2, s3) = \frac{s2 \cdot s3}{|s2| |s3|} = \frac{0}{0.087 \times 0.138} = 0$$

Manhattan dist :

$$\begin{aligned} M\text{-dist}(s1, s2) &= \sum |s1 - s2| \\ &= 0 + 0.01 + 5 \times 10^{-3} + 5 \times 10^{-3} + 5 \times 10^{-3} + 0.04 + 5 \times 10^{-3} + 0.04 + 0.04 + 0.04 + 0.053 + 0.053 \\ &= 0.296 \end{aligned}$$

$$\begin{aligned} M\text{-dist}(s1, s3) &= \sum |s1 - s3| \\ &= 0.03 + 0.05 + 0.05 + 0.05 + 0.04 + 0.05 + 0.04 + 0.04 + 0.04 + 0.04 + 0.08 + 0.08 + 0.08 \\ &= 0.49 \end{aligned}$$

$$\begin{aligned} M\text{-dist}(s2, s3) &= \sum |s2 - s3| \\ &= 0.02 + 0.02 + 0.02 + 0.02 + 0.02 + 0.053 + 0.053 + 0.08 + 0.08 + 0.08 \\ &= 0.446 \end{aligned}$$

Euclidean Distance :

$$\begin{aligned} E\text{-dist}(s1, s3) &= \sqrt{\sum (s1 - s3)^2} \\ &= \sqrt{0 + (0.01)^2 + (5 \times 10^{-3})^2 + (-5 \times 10^{-3})^2 + (-5 \times 10^{-3})^2 + (0.04)^2 + (-5 \times 10^{-3})^2 + (0.04)^2 + (0.04)^2 + (0.04)^2 + (0.053)^2 + (-0.053)^2} \\ &= 0.110 \end{aligned}$$

Date: _____

M T W T F S S
○ ○ ○ ○ ○ ○ ○

$$\begin{aligned} E_dist(s1, s3) &= \sqrt{\text{sum}(s1-s3)^2} \\ &= \sqrt{(0.03)^2 + (0.015)^2 + (0.015)^2 + (0.015)^2 + (0.04)^2 + (0.015)^2 + (0.04)^2 + (0.04)^2 + (0.04)^2} \\ &\quad + (-0.08)^2 + (-0.08)^2 + (-0.08)^2 \\ &= 0.165 \end{aligned}$$

$$\begin{aligned} E_dist(s2, s3) &= \sqrt{(0.02)^2 + (0.02)^2 + (0.02)^2 + (0.02)^2 + (0.02)^2 + (0.053)^2 + (0.053)^2} \\ &\quad + (-0.08)^2 + (-0.08)^2 + (-0.08)^2 \\ &= 0.163 \end{aligned}$$