WORD EMBEDDINGS

BACKGROUND MATH

EXAMPLE 1:

For the following 1-Hot word embeddings, determine whether Word A is closer/more similar to Word B or Word C by measuring the
(1) Euclidean Distance and (2) Cosine Similarity

Word A = [1, 0, 0]
Word B = [0, 1, 0]
Word C = [0, 0, 1]

- **Euclidean Distance formula**

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Euclidean Distance between Word A [1, 0, 0], Word B [0, 1, 0]=

$$d = \sqrt{(0-1)^2 + (1-0)^2 + (0-0)^2}$$

$$d = \sqrt{(-1)^2 + (1)^2 + (0)^2}$$

$$d = \sqrt{1+1+0}$$

$$d = \sqrt{2}$$

Euclidean Distance between Word A [1, 0, 0], Word C [0, 0, 1]=

$$d = \sqrt{(0-1)^2 + (0-0)^2 + (1-0)^2}$$

$$d = \sqrt{(-1)^2 + (0)^2 + (1)^2}$$

$$d = \sqrt{1+0+1}$$

$$d = \sqrt{2}$$

The Euclidean Distance between 1-hot embeddings of Word A and Word B vs Word A and Word C is the same i.e. √2. So it not possible to determine which pair is closer.

- **Cosine similarity formula**

$$\cos \alpha = \frac{\overline{a} \cdot \overline{b}}{|\overline{a}||\overline{b}|}$$

Cosine Similarity between Word A [1, 0, 0], Word B [0, 1, 0]=

Calculate dot product:

$$\overline{a} \cdot \overline{b} = a_x \cdot b_x + a_y \cdot b_y + a_z \cdot b_z = 1 \cdot 0 + 0 \cdot 1 + 0 \cdot 0 = 0 + 0 + 0 = 0$$

Calculate magnitude of a vectors:

$$|\overline{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2} = \sqrt{1^2 + 0^2 + 0^2} = \sqrt{1 + 0 + 0} = \sqrt{1} = 1$$
$$|\overline{b}| = \sqrt{b_x^2 + b_y^2 + b_z^2} = \sqrt{0^2 + 1^2 + 0^2} = \sqrt{0 + 1 + 0} = \sqrt{1} = 1$$

Calculate angle between vectors:

$$\cos \alpha = \frac{\overline{a} \cdot \overline{b}}{|\overline{a}||\overline{b}|}$$

$$\cos \alpha = \frac{0}{1 \cdot 1} = 0$$

Cosine similarity between Word A [1, 0, 0], Word C [0, 0, 1]=

Calculate dot product:

$$\overline{a} \cdot \overline{b} = a_x \cdot b_x + a_y \cdot b_y + a_z \cdot b_z = 1 \cdot 0 + 0 \cdot 0 + 0 \cdot 1 = 0 + 0 + 0 = 0$$

Calculate magnitude of a vectors:

$$|\overline{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2} = \sqrt{1^2 + 0^2 + 0^2} = \sqrt{1 + 0 + 0} = \sqrt{1} = 1$$
$$|\overline{b}| = \sqrt{b_x^2 + b_y^2 + b_z^2} = \sqrt{0^2 + 0^2 + 1^2} = \sqrt{0 + 0 + 1} = \sqrt{1} = 1$$

Calculate angle between vectors:

$$\cos \alpha = \frac{\overline{a} \cdot \overline{b}}{|\overline{a}||\overline{b}|}$$

$$\cos \alpha = \frac{0}{1 \cdot 1} = 0$$

The Cosine Similarity between 1-hot embeddings of Word A and Word B vs Word A and Word C is the same i.e. 0. So it is not possible to determine which pair is closer.

EXAMPLE 2:

For the following word embeddings (dense representation), determine whether Word A is closer/more similar to Word B or Word C by measuring the
(1) Euclidean Distance and (2) Cosine Similarity

Word A = [-3, 2.5, 1]
Word B = [0, 1, -4.5]
Word C = [0.5, -0.25, 0]

- **Euclidean Distance formula**

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Euclidean Distance between Word A [-3, 2.5, 1], Word B [0, 1, -4.5] =

$$d = \sqrt{(0 - (-3))^2 + (1 - (2.5))^2 + (-4.5 - (1))^2}$$

$$d = \sqrt{(3)^2 + (-1.5)^2 + (-5.5)^2}$$

$$d = \sqrt{9 + 2.25 + 30.25}$$

$$d = \sqrt{41.5}$$

$$d = 6.442049$$

Euclidean Distance between Word A [-3, 2.5, 1], Word C [0.5, -0.25, 0] =

$$d = \sqrt{(0.5 - (-3))^2 + (-0.25 - (2.5))^2 + (0 - (1))^2}$$

$$d = \sqrt{(3.5)^2 + (-2.75)^2 + (-1)^2}$$

$$d = \sqrt{12.25 + 7.5625 + 1}$$

$$d = \sqrt{20.8125}$$

$$d = 4.562072$$

The Euclidean Distance between Word A and Word B is 6.44 and between Word A and Word C is 4.56. So Word A and Word C are closer to each other.

- **Cosine similarity formula**

$$\cos \alpha = \frac{\bar{a} \cdot \bar{b}}{|\bar{a}||\bar{b}|}$$

Cosine Similarity between Word A [-3, 2.5, 1], Word B [0, 1, -4.5] =

Calculate dot product:

$$\bar{a} \cdot \bar{b} = a_x \cdot b_x + a_y \cdot b_y + a_z \cdot b_z = (-3) \cdot 0 + 2.5 \cdot 1 + 1 \cdot (-4.5) = 0 + 2.5 - 4.5 = -2$$

Calculate magnitude of a vectors:

$$|\bar{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2} = \sqrt{(-3)^2 + (2.5)^2 + 1^2} = \sqrt{9 + 6.25 + 1} = \sqrt{16.25} = \frac{\sqrt{65}}{2}$$

$$|\bar{b}| = \sqrt{b_x^2 + b_y^2 + b_z^2} = \sqrt{0^2 + 1^2 + (-4.5)^2} = \sqrt{0 + 1 + 20.25} = \sqrt{21.25} = \frac{\sqrt{85}}{2}$$

Calculate angle between vectors:

$$\cos \alpha = \frac{\bar{a} \cdot \bar{b}}{|\bar{a}||\bar{b}|}$$

$$\cos \alpha = \frac{-2}{\sqrt{65}/2 \cdot \sqrt{85}/2} = -\frac{8\sqrt{221}}{1105} \approx -0.10762764703941$$

Cosine similarity between Word A [-3, 2.5, 1], Word C [0.5, -0.25, 0] =

Calculate dot product:

$$\bar{a} \cdot \bar{b} = a_x \cdot b_x + a_y \cdot b_y + a_z \cdot b_z = (-3) \cdot 0.5 + 2.5 \cdot (-0.25) + 1 \cdot 0 = -1.5 - 0.625 + 0 = -\frac{17}{8}$$

Calculate magnitude of a vectors:

$$|\bar{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2} = \sqrt{(-3)^2 + (2.5)^2 + 1^2} = \sqrt{9 + 6.25 + 1} = \sqrt{16.25} = \frac{\sqrt{65}}{2}$$

$$|\bar{b}| = \sqrt{b_x^2 + b_y^2 + b_z^2} = \sqrt{(0.5)^2 + (-0.25)^2 + 0^2} = \sqrt{0.25 + 0.0625 + 0} = \sqrt{0.3125} = \frac{\sqrt{5}}{4}$$

Calculate angle between vectors:

$$\cos \alpha = \frac{\bar{a} \cdot \bar{b}}{|\bar{a}||\bar{b}|}$$

$$\cos \alpha = \frac{-17/8}{\sqrt{65}/2 \cdot \sqrt{5}/4} = -\frac{17\sqrt{13}}{65} \approx -0.9429903335828895$$

The Cosine Similarity between between Word A and Word B is 0.107 and between Word A and Word C is 0.94. So Word A and Word C are more similar to each other.