



# Representations and metrics

6 questions

1  
point

1.

Suppose you'd like to perform nearest neighbor search from the following set of houses:

	Price (USD)	Number of rooms	Lot size (sq. ft.)
House 1	500000	3	1840
House 2	350000	2	1600
House 3	600000	4	2000
House 4	400000	2	1900

Since the features come in wildly different scales, you decide to use scaled Euclidean distances. Choose the set of weights  $a_i$  (as presented in the video lecture) that place each feature on the same scale in the distance calculation. (Hint: Look at the form of the scaled Euclidean distance equation carefully before selecting an answer below.)

Note:

1  $a_{\text{price}}$  = weight assigned to price (USD)  
2  $a_{\text{room}}$  = weight assigned to number of rooms  
3  $a_{\text{lot}}$  = weight assigned to lot size (sq.ft.)

- ☐  $a_{\text{price}} = 1, a_{\text{room}} = 1, a_{\text{lot}} = 1$
- ☐  $a_{\text{price}} = 1, a_{\text{room}} = 1, a_{\text{lot}} = 1e-6$
- ☒  $a_{\text{price}} = 1e-10, a_{\text{room}} = 1, a_{\text{lot}} = 1e-6$
- ☐  $a_{\text{price}} = 1e-5, a_{\text{room}} = 1, a_{\text{lot}} = 1e-3$



a\_price = 1e5, a\_room = 1, a\_lot = 1e3

---

1  
point

2.

Consider the following two sentences.

- Sentence 1: The quick brown fox jumps over the lazy dog.
- Sentence 2: A quick brown dog outpaces a quick fox.

Compute the Euclidean distance using word counts. Round your answer to 3 decimal places.

Note. To compute word counts, turn all words into lower case and strip all punctuation, so that "The" and "the" are counted as the same token.

3.606

---

1  
point

3.

Refer back to the two sentences given in Question 2 to answer the following:

Recall that we can use cosine similarity to define a distance. We call that distance cosine distance.

Compute the **cosine distance** using word counts. Round your answer to 3 decimal places.

Note: To compute word counts, turn all words into lower case and strip all punctuation, so that "The" and "the" are counted as the same token.

Hint. Recall that we can use cosine similarity to define a distance. We call that distance cosine distance.

0.565

---

1  
point

4.

(True/False) For positive features, cosine similarity is always between 0 and 1.

- ☒ True
- ☐ False
- 

1  
point

5.

Using the formula for TF-IDF presented in the lecture, complete the following sentence:

A word is assigned a zero TF-IDF weight when it appears in \_\_\_\_ documents. (N: number of documents in the corpus)

- ☒  $N - 1$
- ☐  $N/2$
- ☐  $N$
- ☐  $0.1 * N$
- ☐ 100
- 

1  
point

6.

Which of the following does **not** describe the word count document representation?

- ☐ Ignores the order of the words
- ☐ Assigns a high score to a frequently occurring word
- ☒ Penalizes words that appear in every document
- 

Submit Quiz

