

CSE 5 – Principles and Practice of AI

Worksheet 8

Learning Objectives

By the end of today's class, you will be able to:

1. Explain how computers measure similarity between words and images
2. Describe how data can be represented using features and points in a space
3. Distinguish between manually defined features and learned representations (embeddings)
4. Interpret similarity scores and compare them with human intuition

Instructions

Create a copy of this document and complete it during class.

SUBMISSION: When finished, download a **PDF** version of the completed worksheet and upload it to **Gradescope** under **In-Class - Week 4 Thursday** to receive credit.

Reminders

- Complete Quiz 3 by **Friday, April 24 11:59 PM**
- Complete and submit **both** Week 4 in-class worksheets by **Sunday, April 26 11:59 PM**
- Complete Assignment 1 by **Monday, April 27 11:59 PM**

Activity: Play Semantris

We will start by playing a game called [Semantris](#).

Your task is to play the game and observe **how the AI decides which words are related**.

Instructions

1. Open **Arcade Mode** in [Semantris](#).
2. Play for a few minutes.

As you play, try:

- Entering words that are **very closely related**
- Entering words that are **loosely related**
- Entering words that could match **multiple options**

Pay attention to how the system responds.

Optional: If you want, you may also try **Blocks Mode**, which is slower and may make it easier to observe how the system chooses between multiple possible matches.

Answer the following questions:

Q. What kinds of words seemed to work best?

Q. Did the system ever accept a word that was not a synonym? Give one example.

Q. Did the system ever behave differently than you expected? Explain.

Q. Propose a hypothesis for how you think the system decides which words are related. What information might it be using?

What might the AI be doing?

In Semantris, the system seemed to know which words were related. But how could a computer measure the **similarity** between words?

One idea is to represent words **as points in a space**, where similar words are closer together. In the next activity we will explore a simplified semantic space where each word is represented by coordinates.

Activity: Building a simple semantic space

Part A – Choosing Features

Consider the four words: *man, woman, boy, girl*

Q. What features could we use to describe these words so that a computer can **distinguish** between the words?

Feature 1: _____

Feature 2: _____

(We will decide on the final features together as a class.)

You can define more than 2 features if you want, but we will use two features for the remaining part.

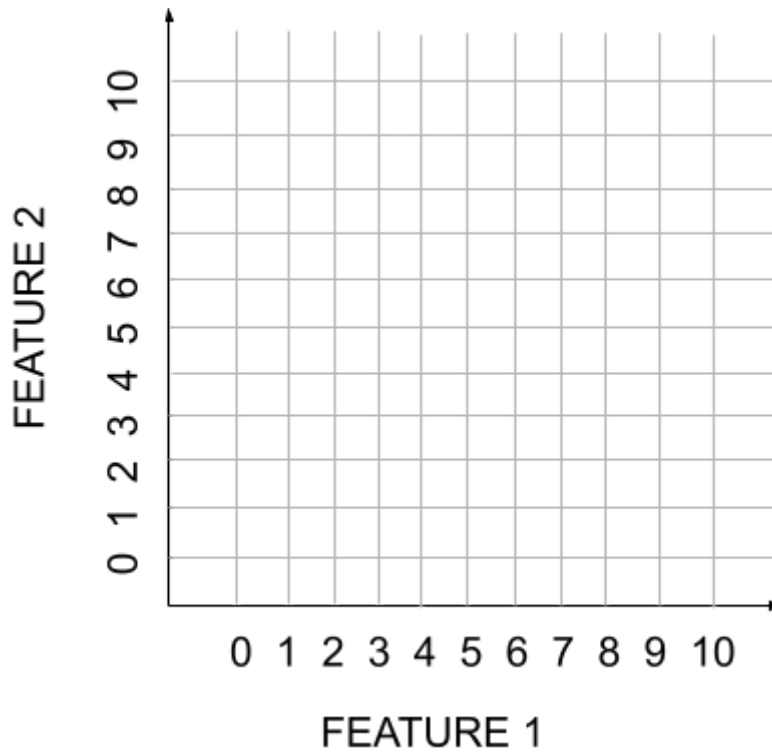
Part B – Visualizing the space and assigning Values

Now, let's use these features to assign values to each word.

Imagine a 2D graph where:

- Feature 1 is one axis
- Feature 2 is the other axis

In the graph below, rename the axes using the features we decided:



Now let's assign values to each feature for each word. Assume each feature ranges from **0 to 10**. (*We will assign these values together.*)

Word	Feature 1?	Feature 2?
man		
woman		
boy		
girl		

Now let's go back to the graph, and place markers for each word on the graph based on these assigned values.

Part C – Extending the Space

Now consider the following words:

grandmother, grandfather, grandparent, teenager, infant.

Q. Based on the feature space we created, what values would you assign these words or where would you place these words on the graph?

Go to the figure again and place these words on the graph based on these new values.

From Simple Features to Learned Representations

In our example, we described words using just two features. But real language is much more complex. It is difficult to manually define all the features needed to represent words like cucumber, smiled, or honesty.

Instead, modern systems learn these features automatically from large amounts of data. These learned representations are called **word embeddings**.

You can think of them as representing each word using many numbers, where similar words end up closer together.

Activity: Exploring Image Similarity

Just like words, images can also be represented using numbers.

Instead of defining features manually (like color, shape, or object type), systems learn these features automatically from large amounts of data.

These learned representations are called image embeddings. When images are represented this way, similar images end up closer together, allowing the system to compare them.

Part A – Setup

1. Download the dataset from Canvas - Resources >> images_no_class.zip
2. Go to: https://mediapipe-studio.webapps.google.com/studio/demo/image_embedder
NOTE - You may have to use your personal google account to access this.
3. Choose **ONE** base image from the dataset (e.g., a cat, person, object).
4. Upload this image in the left panel.

Part B – Compare Images

Select and upload images (one at a time) in the right panel and observe the similarity score. You must complete at least 3 different types of comparisons.

Comparison 1: Highly Similar

- Choose an image that you think is very similar to the base image.
- Similarity Score: _____
- Paste a screenshot of the images side by side below:

Comparison 2: Somewhat Similar

- Choose an image that is partially similar (shares some features but not all)
- Similarity Score: _____
- Paste a screenshot of the images side by side below:

Comparison 3: Not Similar

- Choose an image that is very different from the base image.
- Similarity Score: _____
- Paste a screenshot of the images side by side below:

Comparison 4: Identical Image

- Upload the same image as the base image in the right panel.
- Similarity Score: _____

Answer the following:

Q. How does the similarity score change across your comparisons? Is it what you expected?

Q. Did the similarity scores always match your intuition? Explain.

Comparison 5: When the Model Disagrees with You

Using a base image, consider two comparison images.

Can you find a case where *your intuition about similarity is different from the model's ranking*? In other words, you think Image A is more similar to the base than Image B, but the model gives a higher similarity score to Image B (or vice versa).

If you have already encountered such a case in your earlier comparisons, you may use that example. Otherwise, try different images (you may also change the base image) to find one.

Q. Report the similarity scores for both comparison images.

Q. Explain why your intuition differs from the model.

Q. What features (e.g., color, shape, background, texture) might be influencing the model's judgment?

References: <https://www.cs.cmu.edu/~dst/WordEmbeddingDemo/tutorial.html>