CS344 Exercise 9

Task 1: Transfer learning and embeddings

Transfer learning is the process of reducing the dimensionality of data so that it can be visualized.

TRUE   FALSE

When using a pre-trained NN developed by someone else, it is possible to use some or all of the layers in the NN, and it is possible to re-train some, none, or all of the layers.

TRUE   FALSE

In deep learning, an embedding is a numerical representation of data that captures some of the meaning of the data.

TRUE   FALSE

Only image data can be embedded.

TRUE   FALSE

In deep learning, the features of an embedding are generally interpretable, i.e., it is straightforward to understand the meaning of each feature.

TRUE   FALSE

Using GloVe word representations is an example of using transfer learning and embeddings.

TRUE   FALSE

Using ResNet50 to calculate the similarity of images is an example of using transfer learning and embeddings.

TRUE   FALSE

t-SNE can be used to embed higher dimensional data in two-dimensional space so that the data can be visualized.

TRUE   FALSE

Task 2: Cosine similarity

The ***cosine similarity*** of two vectors is the cosine of the angle between the two vectors. It is a measure of how similar two vectors are. Two highly similar vectors will have a cosine close to 1. Two unrelated (orthogonal) vectors will have a cosine of 0. Two oppositely related vectors will have a cosine of -1.
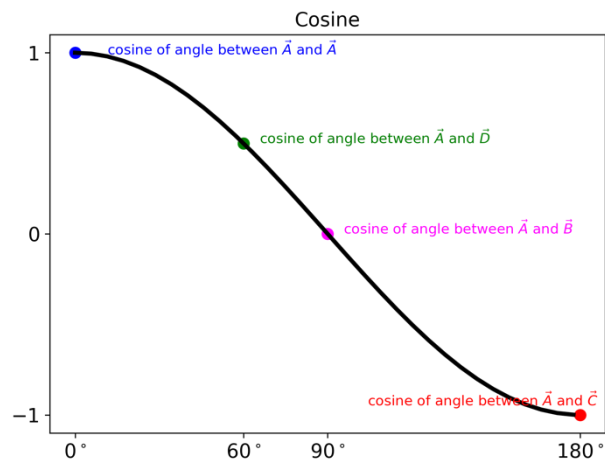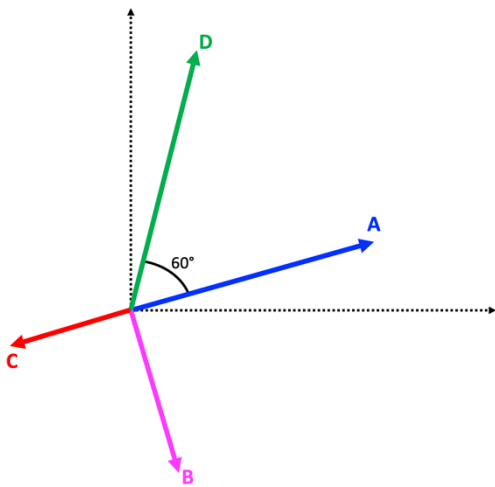
In the figures above, the angle between $\vec{A}$ and $\vec{A}$ (itself) is 0° and the cosine of this angle is 1, the angle between $\vec{A}$ and $\vec{B}$ is 90° and the cosine of this angle is 0, and the angle between $\vec{A}$ and $\vec{C}$ is 180° and the cosine of this angle is -1.

What is the cosine of the angle between $\vec{A}$ and $\vec{D}$?

For two *d*-dimensional vectors, $\vec{U}$ and $\vec{V}$, cosine similarity can be calculated as follows:

$$\text{cosine\_similarity}(\vec{U}, \vec{V}) = \frac{\vec{U} \cdot \vec{V}}{\left\| \vec{U} \right\| \left\| \vec{V} \right\|} = \frac{\sum_{i=1}^{d} U_i V_i}{\sqrt{\sum_{i=1}^{d} U_i^2} \sqrt{\sum_{i=1}^{d} V_i^2}}$$

What is the cosine similarity of the two 3-dimensional vectors [1, 2, 2] and [2, 1, 2]?

Task 3: Coding with transfer learning and embeddings

Download the Jupyter Notebook for Exercise 9 from the course website. Open the Notebook in your web browser and work through it. As you work through the Notebook, answer the following questions.

For the **Dog_Breeds** dataset, what is the model's F1-score on validation data when transfer learning *was not* used?

For the **Dog_Breeds** dataset, what is the model's F1-score on validation data when transfer learning *was* used?

How many *total* parameters does the model that uses transfer learning have?

How many *trainable* parameters does the model that uses transfer learning have?

For the **Flowers** dataset, what is your model's accuracy on *testing* data as reported by the Exercise 9 competition website?

For the **bulldog.jpg** image that we provided, what is the cosine similarity score of the most similar image?

For the **sunrise.jpg** image that we provided, what is the cosine similarity score of the most similar image?

Looking at the t-SNE plot, what other types of weather images as a group are most similar to the rain images? Cloudy or Shine or Sunrise?

What is the first value (of 50) in the embedding for the word "apple"?

Is "wellesley" in the dictionary, i.e., one of the embedded words?

What is an example of a word that is not in the dictionary (it's fine to answer this question with a proper name)?

What is the similarity of the words "excited" and "alphabet"?

Is cosine similarity a symmetric measure, i.e., is `cosine_similarity(A, B)` always the same as `cosine_similarity(B, A)`?

Using our 50-dimensional embedding, what word is predicted for completing the analogy `Argentina is to spanish as Brazil is to ?????`?

Using our 50-dimensional embedding, what word is predicted for completing the analogy `bees is to swarm as geese is to ?????`?

What are the 4 nearest neighbors to the word "wellesley"?

CS344 Exercise 9 Final Page

In the *TIME* column, please estimate the time you spent on this exercise. Please try to be as accurate as possible; this information will help us to design future exercises.

PART	TIME
Exercise	