Exercise 1: Word Vectors

The lecture introduced distributional representations of words (word vectors) as embeddings of words in a latent space.

1. Which of the following statements about word vectors are true?
   - ❑ Word embeddings are sparse vector representations of words.
   - ❑ Embeddings are learned from surrounding context in which words appear.
   - ❑ Embeddings can only be learned for words, not subwords or sentences.
   - ❑ Embeddings can capture semantic relationships between words.
   - ❑ Embeddings can not capture syntactic relationships between words.
   - ❑ The higher the dimensionality of the embedding, the more information they can capture.

Exercise 1: Word Vectors

The lecture introduced distributional representations of words (word vectors) as embeddings of words in a latent space.

1. Which of the following statements about word vectors are true?
2. Which of the following equations should hold for "good" word embeddings?

- ❑ $\mathbf{t}_{\text{boy}} - \mathbf{t}_{\text{girl}} \approx \mathbf{t}_{\text{brother}} - \mathbf{t}_{\text{sister}}$
- ❑ $\mathbf{t}_{\text{cat}} - \mathbf{t}_{\text{doc}} \approx \mathbf{t}_{\text{puppy}} - \mathbf{t}_{\text{kitten}}$
- ❑ $\mathbf{t}_{\text{berlin}} - \mathbf{t}_{\text{tokyo}} + \mathbf{t}_{\text{japan}} \approx \mathbf{t}_{\text{germany}}$
- ❑ $\mathbf{t}_{\text{bigger}} - \mathbf{t}_{\text{big}} + \mathbf{t}_{\text{cold}} \approx \mathbf{t}_{\text{colder}}$

Exercise 1: Word Vectors

The lecture introduced distributional representations of words (word vectors) as embeddings of words in a latent space.

1. Which of the following statements about word vectors are true?
2. Which of the following equations should hold for “good” word embeddings?
3. What is the difference between static and contextualized word embeddings?

Exercise 1: Word Vectors

The lecture introduced distributional representations of words (word vectors) as embeddings of words in a latent space.

1. Which of the following statements about word vectors are true?
2. Which of the following equations should hold for “good” word embeddings?
3. What is the difference between static and contextualized word embeddings?
4. Most computational models of distributional similarity, including neural embeddings, such as word2vec often embed antonyms (e.g., “good”, “bad”) close to each other in the vector space. Briefly explain the underlying reason for this phenomenon and why it may be problematic.

Exercise 2: Word Mover Distance

- d_1: "The **cat climbed** the tree."
- d_2: "The **feline scaled** the tree."
- d_3: "The **kitten ascended** the tree."

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the WMD between d_1 and d_2.

Exercise 2: Word Mover Distance

- d_1: "The **cat climbed** the tree."
- d_2: "The **feline scaled** the tree."
- d_3: "The **kitten ascended** the tree."

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the WMD between d_1 and d_2.
   Pairwise distances between d_1 and d_2:

$$d(\text{cat}, \text{feline}) = \sqrt{(0.4 - 0.35)^2 + (0.5 - 0.55)^2 + (0.6 - 0.65)^2} = 0.0866$$

$$d(\text{cat}, \text{scaled}) = \sqrt{(0.4 - 0.75)^2 + (0.5 - 0.85)^2 + (0.6 - 0.95)^2} = 0.6062$$

$$d(\text{climbed}, \text{feline}) = \sqrt{(0.7 - 0.35)^2 + (0.8 - 0.55)^2 + (0.9 - 0.65)^2} = 0.4975$$

$$d(\text{climbed}, \text{scaled}) = \sqrt{(0.7 - 0.75)^2 + (0.8 - 0.85)^2 + (0.9 - 0.95)^2} = 0.0866$$

$$\begin{aligned} \text{WMD}(d_1, d_2) &= \min\{d(\text{cat}, \text{feline}) + d(\text{climbed}, \text{scaled}) + d(\text{tree}, \text{tree}), \\ &\qquad d(\text{cat}, \text{scaled}) + d(\text{climbed}, \text{feline}) + d(\text{tree}, \text{tree})\} \\ &= \min\{0.0866 + 0.0866 + 0, 0.6062 + 0.4975 + 0\} = 0.1732 \end{aligned}$$

Exercise 2: Word Mover Distance

- d_1: "The **`cat climbed`** the tree."
- d_2: "The **`feline scaled`** the tree."
- d_3: "The **`kitten ascended`** the tree."

	`the`	`cat`	`climbed`	`tree`	`feline`	`scaled`	`kitten`	`ascended`
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the WMD between d_1 and d_2.
   WMD$(d_1, d_2) = 0.1732$
2. Calculate the WMD between d_1 and d_3.

Exercise 2: Word Mover Distance

- d_1: "The **cat climbed** the tree."
- d_2: "The **feline scaled** the tree."
- d_3: "The **kitten ascended** the tree."

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the WMD between d_1 and d_2.
   WMD$(d_1, d_2) = 0.1732$
2. Calculate the WMD between d_1 and d_3.
   Pairwise distances between d_1 and d_3:

$$d(\text{cat}, \text{kitten}) = \sqrt{(0.4 - 0.45)^2 + (0.5 - 0.57)^2 + (0.6 - 0.67)^2} = 0.1109$$
$$d(\text{climbed}, \text{ascended}) = \sqrt{(0.7 - 0.77)^2 + (0.8 - 0.87)^2 + (0.9 - 0.97)^2} = 0.1212$$

WMD calculation between d_1 and d_3:

$$\begin{aligned}\text{WMD}(d_1, d_3) &= d(\text{cat}, \text{kitten}) + d(\text{climbed}, \text{ascended}) + d(\text{tree}, \text{tree}) \\ &= 0.1109 + 0.1212 + 0 = 0.2321\end{aligned}$$

Exercise 2: Word Mover Distance

- d_1: "The **`cat climbed`** the tree."
- d_2: "The **`feline scaled`** the tree."
- d_3: "The **`kitten ascended`** the tree."

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the WMD between d_1 and d_2.
   WMD$(d_1, d_2) = 0.1732$
2. Calculate the WMD between d_1 and d_3.
   WMD$(d_1, d_3) = 0.2321$
3. Which document (d_2 or d_3) is more similar to d_1?

Exercise 3: Sentence Embeddings

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the sentence embeddings for d_1, d_2, d_3 using vector averaging.

Exercise 3: Sentence Embeddings

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the sentence embeddings for d_1, d_2, d_3 using vector averaging.
   d_1: The cat climbed the tree
   d_2: the feline scaled the tree
   d_3: the kitten ascended the tree

Exercise 3: Sentence Embeddings

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the sentence embeddings for d_1, d_2, d_3 using vector averaging.
   d_1: The cat climbed the tree

$$
\begin{aligned}
\mathbf{s}_{emb}(d_1) &= \frac{1}{5}\left(\mathbf{t}_{\text{the}} + \mathbf{t}_{\text{cat}} + \mathbf{t}_{\text{climbed}} + \mathbf{t}_{\text{the}} + \mathbf{t}_{\text{tree}}\right) \\
&= \frac{1}{5}\left(\begin{bmatrix}0.1\\0.2\\0.3\end{bmatrix} + \begin{bmatrix}0.4\\0.5\\0.6\end{bmatrix} + \begin{bmatrix}0.7\\0.8\\0.9\end{bmatrix} + \begin{bmatrix}0.1\\0.2\\0.3\end{bmatrix} + \begin{bmatrix}1.0\\1.1\\1.2\end{bmatrix}\right) \\
&= \frac{1}{5}\begin{bmatrix}2.3\\2.8\\3.3\end{bmatrix} = \begin{bmatrix}0.46\\0.56\\0.66\end{bmatrix}
\end{aligned}
$$

   d_2: the feline scaled the tree
   d_3: the kitten ascended the tree

Exercise 3: Sentence Embeddings

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the sentence embeddings for d_1, d_2, d_3 using vector averaging.

   d_1: The cat climbed the tree $\quad \mathbf{s}_{emb}(d_1) = (0.46, 0.46, 0.66)^\top$

   d_2: the feline scaled the tree

$$\mathbf{s}_{emb}(d_2) = \frac{1}{5}\left(\mathbf{t}_{\text{the}} + \mathbf{t}_{\text{feline}} + \mathbf{t}_{\text{scaled}} + \mathbf{t}_{\text{the}} + \mathbf{t}_{\text{tree}}\right)$$

$$= \frac{1}{5}\left(\begin{bmatrix}0.1\\0.2\\0.3\end{bmatrix} + \begin{bmatrix}0.35\\0.55\\0.65\end{bmatrix} + \begin{bmatrix}0.75\\0.85\\0.95\end{bmatrix} + \begin{bmatrix}0.1\\0.2\\0.3\end{bmatrix} + \begin{bmatrix}1.0\\1.1\\1.2\end{bmatrix}\right)$$

$$= \frac{1}{5}\begin{bmatrix}2.3\\2.9\\3.4\end{bmatrix} = \begin{bmatrix}0.46\\0.58\\0.68\end{bmatrix}$$

   d_3: the kitten ascended the tree

Exercise 3: Sentence Embeddings

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the sentence embeddings for d_1, d_2, d_3 using vector averaging.
   d_1: The cat climbed the tree $\mathbf{s}_{emb}(d_1) = (0.46, 0.46, 0.66)^\top$
   d_2: the feline scaled the tree $\mathbf{s}_{emb}(d_2) = (0.46, 0.58, 0.68)^\top$
   d_3: the kitten ascended the tree

$$
\begin{aligned}
\mathbf{s}_{emb}(d_3) &= \frac{1}{5}\left(\mathbf{t}_{\text{the}} + \mathbf{t}_{\text{kitten}} + \mathbf{t}_{\text{ascended}} + \mathbf{t}_{\text{the}} + \mathbf{t}_{\text{tree}}\right) \\
&= \frac{1}{5}\left(\begin{bmatrix}0.1\\0.2\\0.3\end{bmatrix} + \begin{bmatrix}0.45\\0.57\\0.67\end{bmatrix} + \begin{bmatrix}0.77\\0.87\\0.97\end{bmatrix} + \begin{bmatrix}0.1\\0.2\\0.3\end{bmatrix} + \begin{bmatrix}1.0\\1.1\\1.2\end{bmatrix}\right) \\
&= \frac{1}{5}\begin{bmatrix}2.42\\2.94\\3.44\end{bmatrix} = \begin{bmatrix}0.48\\0.59\\0.69\end{bmatrix}
\end{aligned}
$$

Exercise 3: Sentence Embeddings

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the sentence embeddings for d_1, d_2, d_3 using vector averaging.
   $\mathbf{s}_{emb}(d_1) = (0.46, 0.46, 0.66)^\top$ $\mathbf{s}_{emb}(d_2) = (0.46, 0.58, 0.68)^\top$
   $\mathbf{s}_{emb}(d_3) = (0.48, 0.59, 0.69)^\top$
2. Which sentence embedding pair is more similar (d_1, d_2) or (d_1, d_3)? Use the Manhattan distance.

Exercise 3: Sentence Embeddings

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the sentence embeddings for d_1, d_2, d_3 using vector averaging.
   $\mathbf{s}_{emb}(d_1) = (0.46, 0.46, 0.66)^\top$ $\mathbf{s}_{emb}(d_2) = (0.46, 0.58, 0.68)^\top$
   $\mathbf{s}_{emb}(d_3) = (0.48, 0.59, 0.69)^\top$
2. Which sentence embedding pair is more similar (d_1, d_2) or (d_1, d_3)? Use the Manhattan distance.

$$
\begin{aligned}
d(\mathbf{s}_{emb}(A), \mathbf{s}_{emb}(B)) &= |0.46 - 0.46| + |0.56 - 0.58| + |0.66 - 0.68| \\
&= 0 + 0.02 + 0.02 = 0.04 \\
d(\mathbf{s}_{emb}(A), \mathbf{s}_{emb}(C)) &= |0.46 - 0.48| + |0.56 - 0.59| + |0.66 - 0.69| \\
&= 0.02 + 0.03 + 0.03 = 0.08
\end{aligned}
$$

Exercise 3: Sentence Embeddings

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the sentence embeddings for d_1, d_2, d_3 using vector averaging.
   $\mathbf{s}_{emb}(d_1) = (0.46, 0.46, 0.66)^\top$ $\mathbf{s}_{emb}(d_2) = (0.46, 0.58, 0.68)^\top$
   $\mathbf{s}_{emb}(d_3) = (0.48, 0.59, 0.69)^\top$
2. Which sentence embedding pair is more similar (d_1, d_2) or (d_1, d_3)? Use the Manhattan distance.
3. Name at least two limitations of the vector averaging approach for generating sentence embeddings.

Exercise 3: Sentence Embeddings

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the sentence embeddings for d_1, d_2, d_3 using vector averaging.
   $\mathbf{s}_{emb}(d_1) = (0.46, 0.46, 0.66)^\top$ $\quad\mathbf{s}_{emb}(d_2) = (0.46, 0.58, 0.68)^\top$
   $\mathbf{s}_{emb}(d_3) = (0.48, 0.59, 0.69)^\top$
2. Which sentence embedding pair is more similar (d_1, d_2) or (d_1, d_3)? Use the Manhattan distance.
3. Name at least two limitations of the vector averaging approach for generating sentence embeddings.
4. Interpret the following cosine similarity values between two sentence embeddings:
   $\textit{sim}_{cosine}(\mathbf{s}_1, \mathbf{s}_2) = -1$ $\quad\textit{sim}_{cosine}(\mathbf{s}_1, \mathbf{s}_2) = 0$ $\quad\textit{sim}_{cosine}(\mathbf{s}_1, \mathbf{s}_2) = 1$

Exercise 3: Sentence Embeddings

	the	cat	climbed	tree	feline	scaled	kitten	ascended
d_1	0.1	0.4	0.7	1.0	0.35	0.75	0.45	0.77
d_2	0.2	0.5	0.8	1.1	0.55	0.85	0.57	0.87
d_3	0.3	0.6	0.9	1.2	0.65	0.95	0.67	0.97

1. Calculate the sentence embeddings for d_1, d_2, d_3 using vector averaging.
   $\mathbf{s}_{emb}(d_1) = (0.46, 0.46, 0.66)^\top$ $\mathbf{s}_{emb}(d_2) = (0.46, 0.58, 0.68)^\top$
   $\mathbf{s}_{emb}(d_3) = (0.48, 0.59, 0.69)^\top$
2. Which sentence embedding pair is more similar (d_1, d_2) or (d_1, d_3)? Use the Manhattan distance.
3. Name at least two limitations of the vector averaging approach for generating sentence embeddings.
4. Interpret the following cosine similarity values between two sentence embeddings:
   $sim_{cosine}(\mathbf{s}_1, \mathbf{s}_2) = -1$ $sim_{cosine}(\mathbf{s}_1, \mathbf{s}_2) = 0$ $sim_{cosine}(\mathbf{s}_1, \mathbf{s}_2) = 1$

5. Which sentence is more similar to d_1 according to cosine similarity?

$$\mathit{sim}_{cosine}(\mathbf{s}_{emb}(d_1), \mathbf{s}_{emb}(d_2)) = \frac{\begin{bmatrix}0.46\\0.56\\0.66\end{bmatrix} \cdot \begin{bmatrix}0.46\\0.58\\0.68\end{bmatrix}}{\left\|\begin{bmatrix}0.46\\0.56\\0.66\end{bmatrix}\right\| \cdot \left\|\begin{bmatrix}0.46\\0.58\\0.68\end{bmatrix}\right\|} = 0.999911$$

$$\mathit{sim}_{cosine}(\mathbf{s}_{emb}(d_1), \mathbf{s}_{emb}(d_3)) = \frac{\begin{bmatrix}0.46\\0.56\\0.66\end{bmatrix} \cdot \begin{bmatrix}0.48\\0.59\\0.69\end{bmatrix}}{\left\|\begin{bmatrix}0.46\\0.56\\0.66\end{bmatrix}\right\| \cdot \left\|\begin{bmatrix}0.48\\0.59\\0.69\end{bmatrix}\right\|} = 0.999992$$