Assignment #3

CSE 447: Natural Language Processing
Eric Boris: 1976637
Collaborators: Allyson Ely, Ardi Madadi, Rahul Ram

Written Problems

---

1. A simple way to compute a vector representation of a sequence of words is to add up the vector representations of the words in the sequence. Consider a sentiment analysis model in which the predicted sentiment is given by:

$$\text{score}(w_1, ..., w_m) = \theta \cdot \sum_{i=1}^{m} \mathbf{x}_{w_i} \tag{1}$$

where w_i is the ith word and $\mathbf{x}_{w_i}$ is the embedding for the ith word; the input is of length m (in word tokens). θ are parameters. Prove that, in such a model, the following two inequalities cannot both hold:

$$\text{score}(\text{good}) > \text{score}(\text{not good}) \tag{2}$$
$$\text{score}(\text{bad}) < \text{score}(\text{not bad}) \tag{3}$$

Answer: Prove by contradiction that inequalites (2) and (3) cannot both be true when using sentiment analysis model (1).

Assume that both inequalities (2) and (3) are true using sentiment analysis model (1). Let $\mathbf{x}_{w_1}$, $\mathbf{x}_{w_2}$, $\mathbf{x}_{w_3}$ be arbitrary word embedding vectors for the words "good", "not", and "bad", respectively and let θ be an arbitrary weight vector.

Claim 1: $0 > \theta \cdot \mathbf{x}_{w_2}$

$$\begin{aligned}
\text{score}(\text{good}) &> \text{score}(\text{not good}) \\
\theta \cdot \sum_{i=1}^{m} \mathbf{x}_{w_i} &> \theta \cdot \sum_{i=1}^{m} \mathbf{x}_{w_i} \\
\theta \cdot \mathbf{x}_{w_1} &> \theta \cdot (\mathbf{x}_{w_2} + \mathbf{x}_{w_1}) \\
\theta \cdot \mathbf{x}_{w_1} &> \theta \cdot \mathbf{x}_{w_2} + \theta \cdot \mathbf{x}_{w_1} \\
0 &> \theta \cdot \mathbf{x}_{w_2}
\end{aligned}$$

Claim 2: $0 < \theta \cdot \mathbf{x}_{w_2}$

$$\begin{aligned}
\text{score}(\text{bad}) &< \text{score}(\text{not bad}) \\
\theta \cdot \sum_{i=1}^{m} \mathbf{x}_{w_i} &< \theta \cdot \sum_{i=1}^{m} \mathbf{x}_{w_i} \\
\theta \cdot \mathbf{x}_{w_3} &< \theta \cdot (\mathbf{x}_{w_2} + \mathbf{x}_{w_3}) \\
\theta \cdot \mathbf{x}_{w_3} &< \theta \cdot \mathbf{x}_{w_2} + \theta \cdot \mathbf{x}_{w_3} \\
0 &< \theta \cdot \mathbf{x}_{w_2}
\end{aligned}$$

By **Claim 1** and **Claim 2**, $0 > \theta \mathbf{x}_{w_2}$ and $0 < \theta \mathbf{x}_{w_2}$, which is a contradiction. Thus, inequalities (2) and (3) cannot both be true when using sentiment analysis model (1).

Next, consider a slightly different model:

$$\text{score}(w_1, ..., w_m) = \frac{1}{m}\left(\theta \cdot \sum_{i=1}^{m} \mathbf{x}_{w_i}\right) \tag{4}$$

Construct an example of a pair of inequalities similar to (2–3) that cannot both hold.

Answer: Prove by contradiction that the following inequalities (5) and (6) cannot both hold when using sentiment analysis model (4).

$$\text{score}(\text{good good}) > \text{score}(\text{not good}) \tag{5}$$
$$\text{score}(\text{good bad}) < \text{score}(\text{not bad}) \tag{6}$$

Assume that both inequalities (5) and (6) are true using sentiment analysis model (4). Let $\mathbf{x}_{w_1}$, $\mathbf{x}_{w_2}$, $\mathbf{x}_{w_3}$ be arbitrary word embedding vectors for the words "good", "not", and "bad", respectively and let θ be an arbitrary weight vector.

Claim 1: $\theta \cdot \mathbf{x}_{w_1} > \theta \cdot \mathbf{x}_{w_2}$

$$\begin{gathered}
\text{score}(\text{good good}) > \text{score}(\text{not good}) \\
\frac{1}{m}\left(\theta \cdot \sum_{i=1}^{m} \mathbf{x}_{w_i}\right) > \frac{1}{m}\left(\theta \cdot \sum_{i=1}^{m} \mathbf{x}_{w_i}\right) \\
\frac{1}{2}\left(\theta \cdot (\mathbf{x}_{w_1} + \mathbf{x}_{w_1})\right) > \frac{1}{2}\left(\theta \cdot (\mathbf{x}_{w_2} + \mathbf{x}_{w_1})\right) \\
\theta \cdot (\mathbf{x}_{w_1} + \mathbf{x}_{w_1}) > \theta \cdot (\mathbf{x}_{w_2} + \mathbf{x}_{w_1}) \\
\theta \cdot \mathbf{x}_{w_1} + \theta \cdot \mathbf{x}_{w_1} > \theta \cdot \mathbf{x}_{w_2} + \theta \cdot \mathbf{x}_{w_1} \\
\theta \cdot \mathbf{x}_{w_1} > \theta \cdot \mathbf{x}_{w_2}
\end{gathered}$$

Claim 2: $\theta \cdot \mathbf{x}_{w_1} < \theta \cdot \mathbf{x}_{w_2}$

$$\begin{gathered}
\text{score}(\text{good bad}) < \text{score}(\text{not bad}) \\
\frac{1}{m}\left(\theta \cdot \sum_{i=1}^{m} \mathbf{x}_{w_i}\right) < \frac{1}{m}\left(\theta \cdot \sum_{i=1}^{m} \mathbf{x}_{w_i}\right) \\
\frac{1}{2}\left(\theta \cdot (\mathbf{x}_{w_1} + \mathbf{x}_{w_3})\right) < \frac{1}{2}\left(\theta \cdot (\mathbf{x}_{w_2} + \mathbf{x}_{w_3})\right) \\
\theta \cdot (\mathbf{x}_{w_1} + \mathbf{x}_{w_3}) < \theta \cdot (\mathbf{x}_{w_2} + \mathbf{x}_{w_3}) \\
\theta \cdot \mathbf{x}_{w_1} + \theta \cdot \mathbf{x}_{w_3} < \theta \cdot \mathbf{x}_{w_2} + \theta \cdot \mathbf{x}_{w_3} \\
\theta \cdot \mathbf{x}_{w_1} < \theta \cdot \mathbf{x}_{w_2}
\end{gathered}$$

By **Claim 1** and **Claim 2**, $\theta \cdot \mathbf{x}_{w_1} > \theta \cdot \mathbf{x}_{w_2}$ and $\theta \cdot \mathbf{x}_{w_1} < \theta \cdot \mathbf{x}_{w_2}$, which is a contradiction. Thus, inequalities (5) and (6) cannot both be true when using sentiment analysis model (4).

---

2. Continuing in the style of problem 1, consider this model:

$$\text{score}(w_1, ..., w_m) = \theta \cdot \text{ReLU}\left(\sum_{i=1}^{m} \mathbf{x}_{w_i}\right) \tag{7}$$

Show that, in this case, it is possible to achieve the inequalities (2–3). The recommended way to do this is to provide weights θ and embeddings for the words good, bad, and not. The problem can be solved with four dimensions.
Hint: problems 1 and 2 are related to the "XOR" problem (2) from assignment 2.

Answer: Prove using sentiment analysis model (5), the word embedding vectors $\mathbf{x}_{\text{good}} = [-1, 1]^T$, $\mathbf{x}_{\text{not}} = [-1, -1]^T$, and $\mathbf{x}_{\text{bad}} = [1, -1]^T$ for the words "good", "not", and "bad", respectively, and weight vector $\theta = [-1, 1]$ that inequalities (2) and (3) hold.

Claim 1: score(good) > score(not good)

$$
\begin{aligned}
\text{score(good)} &> \text{score(not good)} \\
\theta \cdot \text{ReLU}\left(\sum_{i=1}^{m} \mathbf{x}_{w_i}\right) &> \theta \cdot \text{ReLU}\left(\sum_{i=1}^{m} \mathbf{x}_{w_i}\right) \\
\theta \cdot \text{ReLU}\left(\mathbf{x}_{\text{good}}\right) &> \theta \cdot \text{ReLU}\left(\mathbf{x}_{\text{not}} + \mathbf{x}_{\text{good}}\right) \\
\theta \cdot \text{ReLU}\left([-1,\, 1]^T\right) &> \theta \cdot \text{ReLU}\left([-1,\, -1]^T + [-1,\, 1]^T\right) \\
\theta \cdot \text{ReLU}\left([-1,\, 1]^T\right) &> \theta \cdot \text{ReLU}\left([-2,\, 0]^T\right) \\
\theta \cdot [0,\, 1]^T &> \theta \cdot [0,\, 0]^T \\
[-1,\, 1] \cdot [0,\, 1]^T &> [-1,\, 1] \cdot [0,\, 0]^T \\
1 &> 0
\end{aligned}
$$

Claim 2: score(bad) < score(not bad)

$$
\begin{aligned}
\text{score(bad)} &< \text{score(not bad)} \\
\theta \cdot \text{ReLU}\left(\sum_{i=1}^{m} \mathbf{x}_{w_i}\right) &< \theta \cdot \text{ReLU}\left(\sum_{i=1}^{m} \mathbf{x}_{w_i}\right) \\
\theta \cdot \text{ReLU}\left(\mathbf{x}_{\text{bad}}\right) &< \theta \cdot \text{ReLU}\left(\mathbf{x}_{\text{not}} + \mathbf{x}_{\text{bad}}\right) \\
\theta \cdot \text{ReLU}\left([1,\, -1]^T\right) &< \theta \cdot \text{ReLU}\left([-1,\, -1]^T + [1,\, -1]^T\right) \\
\theta \cdot \text{ReLU}\left([1,\, -1]^T\right) &< \theta \cdot \text{ReLU}\left([0,\, -2]^T\right) \\
\theta \cdot [1,\, 0]^T &< \theta \cdot [0,\, 0]^T \\
[-1,\, 1] \cdot [1,\, 0]^T &< [-1,\, 1] \cdot [0,\, 0]^T \\
-1 &< 0
\end{aligned}
$$

Thus, because **Claim 1** and **Claim 2** are true, the inequalities (2) and (3) hold when using sentiment analysis model (5) and the described word embedding vectors and weight vector.