CS 6320.002: Natural Language Processing
Fall 2020

Homework 4 — 45 points
Issued 01 Nov. 2021
Due 11:59pm 15 Nov. 2020

Deliverables: Answers can be typed directly into Gradescope. See the assignment guide for more details.

What does it mean to "show your work?" Write out the math step-by-step; we should be able to clearly follow your reasoning from one step to another. (You can combine "obvious" steps like simplifying fractions or doing basic arithmetic.) The point of showing your work is twofold: to get partial credit if your answer is incorrect, and to show us that you worked the problem yourself and understand it. We will deduct points if steps are missing.

1 Semantic Role Labeling — 10 points

The questions in this section are based on material covered in Week 10.

1.1

Identify the predicate, the arguments, and the semantic role of each argument in the following sentences:

- Alice hid the key in a flowerpot.
- The cold weather killed Bob's rosebushes.
- Eve picked the lock with a paper clip.

You can assume that there is only one predicate per sentence. Use the set of semantic role labels from lecture.

1.2

Convert the following predicate-argument structures into English sentences:

- sank(submarine - agent, battleship - theme, torpedo - instrument)
- booked(Alice - agent, plane ticket - theme, Bob - beneficiary)
- walked(Eve - agent, McDermott Library - source, ECSS - goal)

Write two different sentences for each predicate-argument structure.

2 Word Embeddings — 15 points

The questions in this section are based on material covered in Week 11.
Suppose we have the following table of word co-occurrence counts, where the rows correspond to words w and the columns correspond to contexts c:

(w, c)	leash	matrix	collar	network	pet	graph
vector	0	9	0	7	1	2
walk	6	1	3	2	1	2
embedding	0	7	1	8	0	3
dog	9	0	5	0	3	0
neural	0	9	0	10	1	6
ball	2	0	1	0	1	0

2.1

What are the pointwise mutual information (PMI) embeddings for this co-occurrence table? Show your work. You can round your answers to the hundreth place and leave blank any entires for which the PMI is undefined.

2.2

Now use α-smoothing with $\alpha = 0.75$ to calculate positive pointwise mutual information (PPMI) embeddings for this co-occurrence table. Show your work. You can round your answers to the hundreth place.

3 Neural Networks — 25 points

The questions in this section are based on material covered in Week 12.

3.1 The XOR Problem

Suppose we have a neural network with a hidden layer $\mathbf{h}$ defined as follows:

$$h_1 = x_1 \quad h_2 = x_1 \quad h_3 = x_1 x_2 \quad h_4 = x_1^2 \quad h_5 = x_2^2$$

We also have an output layer $\hat{y} = \text{RELU}(\mathbf{w} \cdot \mathbf{h} + b)$. We want to use this neural network to solve the XOR problem. Give a set of values for the parameters $\mathbf{w}$ and b such that the network correctly models the XOR function. (There is more than one correct answer.) Show that using your parameter values give the correct output values.

3.2   Batched Forward Pass

Suppose we have the following feed-forward neural network:

$$\mathbf{H} = \text{RELU}(\mathbf{W}_1 \cdot \mathbf{X} + \mathbf{b}_1)$$
$$\mathbf{Y} = \text{RELU}(\mathbf{W}_2 \cdot \mathbf{H} + \mathbf{b}_2)$$

The network has already been trained, with parameter values

$$\mathbf{W}_1 = \begin{bmatrix} 1 & -1 & 2 & 3 & 0 \\ 4 & 0 & -1 & 1 & 3 \\ 2 & 1 & 3 & -5 & -4 \\ 4 & -3 & 2 & 1 & -3 \end{bmatrix} \quad \mathbf{b}_1 = \begin{bmatrix} -1 \\ 2 \\ -4 \\ 3 \end{bmatrix}$$

$$\mathbf{W}_2 = \begin{bmatrix} 2 & -1 & -1 & 3 \\ -2 & 1 & -5 & 4 \end{bmatrix} \quad \mathbf{b}_b = \begin{bmatrix} 12 \\ 3 \end{bmatrix}$$

and input is a batch of two (column) vectors, combined into a single matrix $\mathbf{X} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 5 & 3 \\ 1 & 1 \\ 4 & 2 \end{bmatrix}$.

3.2.1

What is the value of $\mathbf{H}$ after running this input batch through the trained network? Show your work.

3.2.2

What is the value of $\mathbf{Y}$ after running this input batch through the trained network? Show your work.

3.3 Backpropagation

Suppose we have a neural network with a hidden layer $\mathbf{h}$ defined as follows:

$$h_1 = \sigma x_1 \quad h_2 = ax_2 + bx_3 + cx_2x_3 \quad h_3 = x_4^2 + d$$

where a, b, c, d are learned parameters. We also have an output layer $\hat{y} = h_1 + h_2 + h_3$. We want to train this network using some loss function $L(\hat{y})$ and learning rate $\eta = 0.1$. Suppose the current values of the parameters are

$$a = 3 \quad b = 4 \quad c = 2 \quad d = 2$$

and we have a training input $\mathbf{x} = (0, 2, -1, 2)$.

3.3.1

What are the values of $h_1, h_2, h_3, \hat{y}$ after running this training input through the network (ie. the forward pass of the computation graph)? Show your work. (You don't have to draw the computation graph for submission, but you might find it helpful to do so for yourself.)

3.3.2

Suppose we have $\dfrac{\partial L}{\partial \hat{y}} = 3$. What are the values of the parameters a, b, c, d after performing backpropagation (ie. the backward pass of the computation graph) and update? Show your work. (You don't have to draw the computation graph for submission, but you might find it helpful to do so for yourself.)