Deep Learning CS583 Fall 2022
Quiz 1 - Section A

October 10th, 2022

Instructor: Jia Xu

Student name: _____

Student ID: _____

Student email address: _____

- **Read these instructions carefully**
- Fill-in your personal info, as indicated above.
- You have 24 hours.
- There are three questions. Each question worths the same.
- Both computer-typed and hand-writing in the very clear form are accepted.
- This is an open-book test.
- You should work on the exam only by yourself.
- Submit your PDF/Doc/Pages **by 18:00 Oct 12th** on Canvas.

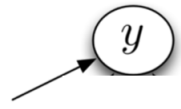
good luck!

1 Question

You are given one or several hidden nodes "h", two inputs x_1, x_2, and the output y. Draw a neural network and assign the weights and bias that performs OR operation:

- if $x_1 = 0$, $x_2 = 0$, then $y = 0$
- if $x_1 = 1$, $x_2 = 0$, then $y = 1$
- if $x_1 = 0$, $x_2 = 1$, then $y = 1$
- if $x_1 = 1$, $x_2 = 1$, then $y = 1$

The activation function outputs 1 if the input is greater than zero and outputs 0 otherwise.

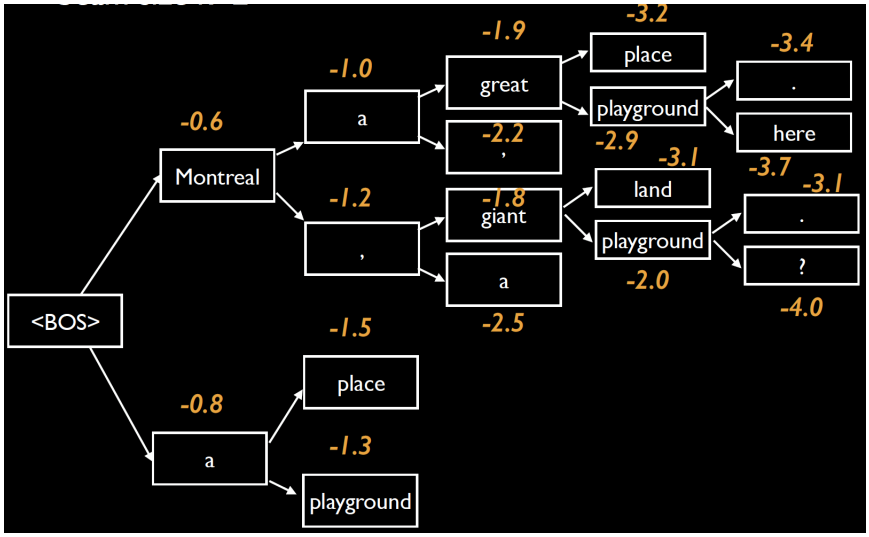


2 Question

- Align each term with its explaination, given c as class, x as input observation.

$Pr(c)$	discriminative model
$Pr(x\|c)$	class conditional probability
$Pr(c\|x)$	generative model
$Pr(c) \cdot Pr(x\|c)$	prior probability

- In a lecture slide below, the orange color indicates the log-likelihood of the partial path from the beginning of the sentence (BOS) until the current output. What is the prediction output if we change the beam size.



  - beam size k=1, output word sequence:

  - beam size k=2, output word sequence:

  - What are the advantage and the disadvantage of a larger beam size?

3 Question

- Briefly explain the trigram method of language modeling.

- What is the procedure of 5-fold cross-validation, and what is its advantage over the traditional approach of simply splitting one's available data into a training set and a validation set?

- We have seen that averaging the outputs from multiple models typically gives better results than using just one model. Let's say that we're going to average the outputs from 10 models. Of course, we want 10 good models, i.e. models that also perform well individually. What additional property of a collection of 10 models makes that collection a good candidate for output averaging?