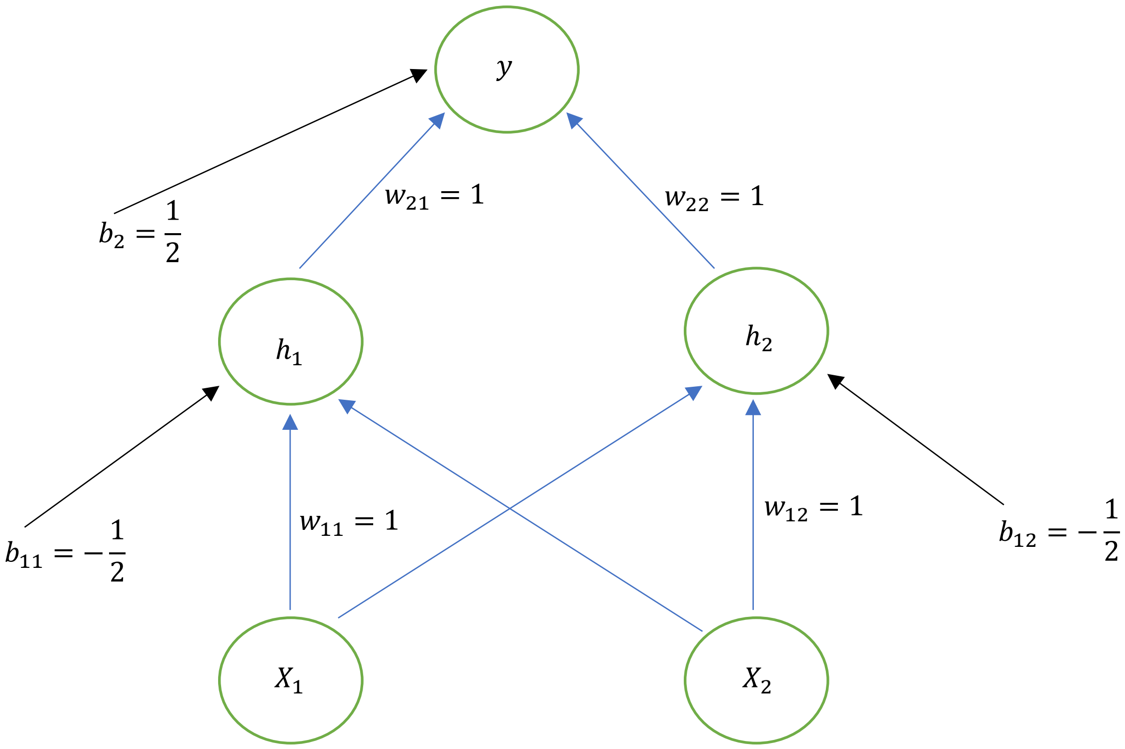
Name: Janmejay Mohanty
CWID: 20009315
Email: jmohanty@stevens.edu

Course: CS 583 A
Course Name: Deep Learning

Quiz1A
Date: 11th October 2022

Solutions

Ans1:



$w_1 = [1 \quad 1]$

$b_1 = \left[-\frac{1}{2} \quad -\frac{1}{2}\right]$

$w_2 = [1 \quad 1]$

$b_2 = \frac{1}{2}$

Activation Function: $f(h) = \begin{cases} 1, & if\ w_x + b > 0 \\ 0, & else \end{cases}$

X_1	X_2	h_1	h_2	y
1	1	1	1	1
1	0	1	1	1
0	1	1	1	1
0	0	0	0	0

Ans2:

(1)

$\boldsymbol{Pr(C) \rightarrow Prior\ Probability}$: The prior probability of an event is the probability of the event computed before the collection of new data. One begins with a prior probability of an event and revises it in the light of new data.

$\boldsymbol{Pr(X|C) \rightarrow Class\ Conditional\ Probability}$: Conditional probability is a measure of probability to an event that is occurring, given other event has already occurred.

$\boldsymbol{Pr(C|X) \rightarrow Discrimitive\ Model}$: Also referred as conditional models. This is a class of logistical models used for classification or regression. It should distinguish decision boundary through observed data.

$\boldsymbol{Pr(C).Pr(X|C) \rightarrow Generative\ Model}$: A generative model describes how a dataset is generated, in terms of a probabilistic model. By sampling this model, we generate data.

(2)

$\text{Beam Size k} = 1\text{: } \langle\text{BOS}\rangle\text{ Montreal a great playground.}$

$\text{Beam Size k} = 2\text{: } \langle\text{BOS}\rangle\text{ Montreal, giant playground.}$

$\boldsymbol{Pros}$: More words would be traversed and chosen. Therefore, a higher probability to get better results.

$\boldsymbol{Cons}$: More computational resources and more memory are required.

Ans3:

(1) Trigram model of language modelling: A trigram model restricts the conditional information to the previous two words. Using this method, the conditional distribution can calculate a certain word combination frequency based on the previous two words.

(2) Procedure of 5-fold cross-validation: Dataset is split into 5 sets. Each one set is taken as test set by turn while other four are training sets. Eventually, we'll have 5 accuracies and the average is the accuracy of 5-fold cross-validation.
Pros: Avoiding the randomness and bias by training and testing all the data.

(3) Bagging: Bootstraps the training set, estimates many copies of a model on the resulting samples and then averages their predictions.
Boosting: Sequentially reweights the training samples forcing the model to attend the training examples with higher loss.
Stacking: Used a separate validation set to train a meta-model that combines predictions of multiple models.