INFO 7390 –
Advances in Data Sciences and Architecture
Sample Quiz

Student Name: _____
Professor: Nik Bear Brown

Rules:
1. Actual exam will be on HackerRank
2. Use Slack to ask questions of TAs

Q1 (5 Points) Define a hyperparameter. How does one determine their values? Give an example of five hyperparameters used in CNNs.

Solution:

A model **hyperparameter** is a configuration that is external to the model and whose value cannot be estimated from data but must be specified before training.

One typically tries a range of reasonable values and look at the effect on the model.

Learning Rate
Number Of Convolutional Layers
Number Of Hidden Units In Fc Layers
Convolution Kernel Size
Stride
Batch Size
Activation Function
Weight Initialization
Dropout For Regularization
Early Stopping Threshold
Etc.

Q2 (5 Points) The number of nodes in the input layer is 10 and the first hidden layer is 5. What is the maximum number of connections from the input layer to the hidden layer?

Solution:

Fifty since MLP is a fully connected directed graph, the number of connections are a multiple of number of nodes in input layer and hidden layer.

Q3 (5 Points) The input image has been converted into a matrix of size 28 X 28 and a kernel/filter of size 7 X 7 with a stride of 1. What will be the size of the convoluted matrix?

Solution:

22 X 22 The size of the convoluted matrix is given by C=((I-F+2P)/S)+1, where C is the size of the Convoluted matrix, I is the size of the input matrix, F the size of the filter matrix and P the padding applied to the input matrix. Here P=0, I=28, F=7 and S=1. There the answer is 22.

Q4 (5 Points) For a binary classification problem how many output nodes would you choose?

Solution:

One or two output nodes.

Q5 (5 Points) Why use dropout?

Solution:

Dropout can help preventing overfitting. Dropout also gives a way to approximate by combining many different architectures.

Q6 (5 Points) What is a Policy (π) in reinforcement learning?

Solution:

Policy (π): The policy, denoted as π (or sometimes π(a|s)), is a mapping from some state s to the probabilities of selecting each possible action given that state. For example, a greedy policy outputs for every state the action with the highest expected Q-Value.

Q7 (5 Points) What defines the relationships between a given state (or state-action pair) to its successors in reinforcement learning?

Solution:

The Bellman equation defines the relationships between a given state (or state-action pair) to its successors in reinforcement learning.

$$Q^*(s,a) = \sum_{s',r} p(s',r|s,a)\left[r + \gamma \max_{a'} Q^*(s',a')\right]$$

Q8 (5 Points) What steps can we take to prevent overfitting in a Neural Network?

Solution:

Data Augmentation
Weight Sharing
Early Stopping
Dropout

Q9 (5 Points) How does q-learning differ from deep q-learning?

Solution:

In deep q-learning the algorithm uses a deep neural network as an approximator for the learning of the q-function. A q-function usually denoted as *Q(s,a)* (sometimes with a π subscript, and sometimes as *Q(s,a; ϑ)* in *Deep RL*), Q Value is a measure of the overall expected *reward* assuming the *Agent* is in *state s* and performs *action a*, and then continues playing until the end of the *episode* following some *policy* π.

Q10 (5 Points) Which of the following statement(s) is / are true for Gradient Decent (GD) and Stochastic Gradient Decent (SGD)?

1. In GD and SGD, you update a set of parameters in an iterative manner to minimize the error function.
2. In SGD, you have to run through all the samples in your training set for a single update of a parameter in each iteration.
3. In GD, you either use the entire data or a subset of training data to update a parameter in each iteration.

Solution:

Only 1

Q11 (5 Points) Can a LogLoss evaluation metric can have negative values?

Solution:

Log loss cannot have negative values.

Q12 (5 Points) Name three deep learning generative models?

Solution:

Autoencoders (AEs)
Variational Autoencoders (VAEs)
Generative Adversarial Networks (GANs)

Q13 (5 Points) What are the steps for using a gradient descent algorithm?

1. Calculate error between the actual value and the predicted value
2. Reiterate until you find the best weights of network
3. Pass an input through the network and get values from output layer
4. Initialize random weight and bias
5. Go to each neurons which contributes to the error and change its respective values to reduce the error

Solution:

4, 3, 1, 5, 2

Q14 (5 Points) Which of the following is true about model capacity (where model capacity means the ability of neural network to approximate complex functions) ?

Solution:

As number of hidden layers increase, model capacity increases

Q15 (5 Points) Why is batch normalization is helpful?

Solution:

Batch normalization is helpful because It normalizes all the input before sending it to the next layer helping numerical issues in backpropagation.

Q16 (5 Points) What if we use a learning rate that's too large?

Solution:

Network may not converge.

Q17 (5 Points) For an image recognition problem (recognizing a cat in a photo), which architecture of neural network would be better suited to solve the problem?

Solution:

Convolutional Neural Network

Q18 (5 Points) What are the factors to select the depth of neural network?

Solution:

1. Type of neural network (eg. MLP, CNN etc)
2. Input data
3. Computation power, i.e. Hardware capabilities and software capabilities
4. Learning Rate
5. The output function to map
6. Desired accuracy

Q19 (5 Points) Name five applications that GANs are useful for.

Solution:

Generate Examples for Image Datasets
Generate Photographs of Human Faces
Generate Realistic Photographs
Generate Cartoon Characters
Image-to-Image Translation
Text-to-Image Translation
Semantic-Image-to-Photo Translation
Face Frontal View Generation
Photograph Editing
Face Aging
Photo Blending
Super Resolution
Photo Inpainting
Clothing Translation
Video Prediction
3D Object Generation
Etc.

Q20 (5 Points) How does an Autoencoder (AE) differ from a Variational Autoencoder (VAE)?

Solution:

An autoencoder accepts input, compresses it, and then recreates the original input in an unsupervised manner.

A variational autoencoder assumes that the source data has some sort of underlying probability distribution (such as mixture of Gaussians) and then attempts to find the parameters of the distributions.