PY-599 (Fall 2018): Applied Artificial Intelligence
Final Exam

Closed books, closed notes, closed electronic-devices, and unfortunately closed neighbor!

Name:
Department:

Special Note: My sincere gratitude goes to a select group of PY599 Applied AI students who helped me to regain a little bit of my self-confidence back! I used to be under assumption that my handwriting is the worst on the planet earth, but during grading midterm papers I realized that it is not true! Thank You! However, that would be great if this elite group of students can take a little bit of effort to make their handwriting *slightly* readable so that I can grade their paper – but not too much effort which can downgrade my handwriting to the worst ranking again resulting in shattering the little bit of self-confidence that I have regained! Thanks in advance!

1- Which item is a universal function approximator? There is just one correct choice. (5 points)

□ A feedforward neural network with at least one hidden layer with nonlinear activation functions and sufficiently large number of hidden nodes.

□ A single neuron

□ A single-layer neural network with sufficiently large number of nodes (no hidden layer in the network).

□ A feedforward neural network with at least one hidden layer with linear activation functions and sufficiently large number of hidden nodes.

2- Which item is correct about universal function approximation theory? There is just one correct choice. (5 points)

□ Universal function approximation theory tells us how to find a set of parameters to approximate any function.

□ Universal function approximation theory tells us how to use gradient descent methods to train neural networks to approximate any function.

□ Universal function approximation theory says nothing about how to find a set of parameters to approximate any function, it just talks about universality of the neural networks.

3- Which choice correctly characterizes the relationship between the depth of a neural network and its representation/learning power? There is just one correct item. (5 points)

□ Given the same number of neurons, shallow networks and deep networks are equally capable in learning and representing different patterns and functions.

□ Compared to a deeper network, a shallow network can need exponentially greater number of neurons to implement the same function.

□ Deeper networks are easier to train than the shallower networks.

□ Shallow networks are more powerful than the deep networks.

4- Which description is correct about training multilayer feedforward neural networks. There is just one correct choice. (5 points)

□ There are training algorithms with polynomial time complexity that can find the optimal set of parameters of a neural network to implement a given task or function.

□ Training a multilayer neural network is a hard problem, and therefore we should not use neural network and instead we should use models that their training is a tractable problem.

□ Training a multilayer neural network is a hard problem. However, there are methods such as gradient based techniques that can help us to find a set of near-optimal parameters to train a neural network in a reasonable time.

5- How many neurons do we need in the output layer of a neural network if we want to classify reviews into two categories, positive and negative? (9 points)

What activation function should we use in this output layer?

What type of cost function would you suggest to use for this binary classification problem?

6- How many neurons do we need in the output layer of a neural network if we want to predict the value of a house (this is a regression problem)? (9 points)

What activation function should we use in this output layer?

What type of cost function would you suggest to use for this regression problem?

7- In machine learning we use optimization techniques to fit a model to the training data. But machine learning is not just an optimization problem (or is it?). Please explain why machine learning is more than just an optimization problem. (8 points)

8- Generally speaking, can we train a model using one type/style of data and expect it to perform well on a different type/style of data? (5 points)

9- We trained a model on a training set. Please see the performance of the model on the training data and the validation data over different epochs. What is the problem? Is this result OK or not? And why? (10 points)

10- Please list 4 different methods to fight against overfitting. There is no need for explanation. (8 points)

11- Please briefly explain how Nesterov Momentum is different from the regular momentum and write a pseudocode to implement it (9 points).

12- Parameter sharing is one of the most powerful ideas used in neural networks. Which descriptions about parameter sharing are correct? There can be more than one correct choice (8 points).

□ Both CNN and RNN utilize the idea of parameter sharing. CNN uses the same parameters to process different locations of the grid input, whereas RNN applies the same parameters to process different elements of a sequence.

□ Parameter sharing can be considered as a form of regularization.

□ Parameter sharing reduces the number of parameters of a model, therefore it reduces the memory footprint of the model (we need to store fewer number of parameters).

□ Parameter sharing makes training the neural network easier and faster, but at the same time it can result in overfitting.

13- Generative adversarial networks (GANs) are one the coolest ideas of modern AI. Which descriptions about GANs are correct? There can be more than one correct description (6 points).

□ In GANs we have two models competing against each other; a generator, and a discriminator.

□ In GANs we can include an additional label input to the generator, asking it to generate a fake sample with a specific condition, and train the generator/discriminator pair accordingly.

□ In GANs we should use a pretrained discriminator that is already trained to classify inputs to different categories, and then we train the generator against this discriminator.

14- Please explain the need for reinforcement learning (why we cannot use supervised learning instead of reinforcement learning) (8 points).