ECE523: Engineering Applications of Machine Learning and Data Analytics

I acknowledge that this exam is solely my effort. I have done this work by myself. I have not consulted with others about this exam in any way. I have not received outside aid (outside of my own brain) on this exam. I understand that violation of these rules contradicts the class policy on academic integrity.

Name: ____________________________

Signature: ____________________________

Date: ____________________________

Instructions: There are five problems. Partial credit is given for answers that are partially correct. No credit is given for answers that are wrong or illegible. Write neatly.

Problem 1: __________

Problem 2: __________

Problem 3: __________

Problem 4: __________

Problem 5: __________

Total: __________

Quick note about answers submitted for the exam

This exam is a take home. You may use your notes and textbook to help you answer the questions on the exam; however, you are not allowed to use help from other students or the ***internet*** (this includes Greg's video lectures). Any violation of that rule is a severe breach in trust and academic integrity. Finally, you must answer each of the questions in your own words. This means that you should not be copying and pasting from my lecture or scribe notes to answer the questions.

Good luck!

Problem #1 – Neural Networks (10 Points)

(a) One method for preventing the neural networks' weights from overfitting is to add regularization terms. You will now derive the update rules for the regularized neural network. Recall that the non-regularized gradient descent update rule for w_1^{t+1} is:

$$w_{ji}^{t+1} = w_{ji}^{t} + \eta \sum_{n=1}^{N} e_j(n)\phi'\left(v_j(n)\right) y_i(n) \tag{1}$$

Derive the update rule for w_{ji}^{t+1} in the regularized neural net loss function which penalizes based on the square of each weight. Use $\lambda \geq 0$ to denote the regularization parameter. Use the following regularizer:

$$R(w) = \lambda \sum_i w_i^2$$

Re-express the regularized update rule so that the only difference between the regularized setting and the unregularized setting above is that the old weight w_{ji}^t is scaled by some constant. Explain how this scaling prevents overfitting.

(b) The definition of a sigmoid function is given by

$$\phi(x) = \frac{1}{1 + \mathrm{e}^{-x}}$$

Show that $\phi'(x) = \phi(x)(1 - \phi(x))$

Problem #2 – GANs (10 Points)

Describe what a generative adversarial network is and how they can be used.

Problem #3 – Random Short Answer (10 Points)

(SA:1) In the context of a Multi-Arm Bandit, describe the difference between regret and pseudo-regret. Why is it generally easier to work with pseudo-regret than regret? Explain the differences between an adversarial bandit and stochastic bandit. Feel free to support your response with equations.

(SA:2) Describe the two step optimization approach used in the k-means clustering algorithm.

(SA:3) In class, we discussed bounding the risk of a model using Hoeffding's inequality and the union bound. Describe why we are used the supremum in the bound below.

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}}\left|\widehat{R}_n(f) - R(f)\right| \geq \epsilon\right) \leq 2N\exp\left(-2n\epsilon^2\right)$$

(SA:4) Explain the difference between backpropagation and backpropagation through time.

(SA:5) We discussed semi-supervised learning, and in particular, self-training and co-training. Describe the differences between them and provide a real-world example for each one of the algorithms. *Note that you will not get credit if you answer with an example we provided in class. Think outside the box!*

Problem #4 – True/False: A Gamblers Ruin (10 Points)

(1) [True/False] (1 point): Semi-supervised learning is guaranteed to improve the performance of a classifier that is trained using the unlabeled data.

(2) [True/False] (1 point): k-means clustering places data into groups; however, the groups do not necessarily mean the cluster is a class.

(3) [True/False] (1 point): SMOTE generates data that lie outside the convex hull of the minority class data, which is why SMOTE is so effective.

(4) [True/False] (1 point): Passive-Aggressive online learning will never perform and update to $\mathbf{w}_t$ if a data $\mathbf{x}$ is correctly classified.

(5) [True/False] (1 point): The VC-dimension allows us to bound on risk even when $\mathcal{F}$ is infinite.

(6) [True/False] (1 point): The gradients cannot explode in an RNN, which is a desirable property of the backpropagation through time.

(7) [True/False] (1 point): Classification error is typically the best way to measure the performance of an RNN language model.

(8) [True/False] (1 point): Using a sigmoid activation function in a deep neural network trained with backpropagation is one way to avoid the vanishing gradient problem.

(9) [True/False] (1 point): A discriminator network, D, after enough training in a GAN will always be able to identify if a sample came from the data set or the generator network, G.

(10) [True/False] (1 point): In backpropagation, the only difference between updating a hidden node versus an output node is how the local gradient is calculated.

Problem #5 – Recurrent Neural Networks (10 Points)

Answer the following questions about recurrent neural networks:

1. Describe the backpropagation through time algorithm and why it is different than the standard backpropagation algorithm.
2. In the context of a language model, why is using a class of words needed to train the model?
3. Outside of the language model, provide an example where they can be used.

Scratch Paper (not graded)