Machine Learning & Discovery *ECS 271, Fall 2023*

Final Exam

Name:

1. What is vanishing gradient, when it happens, and how can we avoid it?
2. What is the role of momentum in training a deep network?
3. How do we choose when to stop training a deep network?
4. Given a deep network $y = f(x)$, how can we construct an adversarial example?
5. Assuming that we want to to PCA on data to keep only k dimensions, does scaling one of the features by a constant scalar change the error of the PCA? How about rotating all the feature vectors with a constant rotation matrix?
6. What is the loss function in k-means and how can we optimize it? Please write the pseudo code.
7. What is the effect of large or small batch in stochastic gradient descent? Which one is better in what cases?