Answer the following questions and submit a pdf file with your answers in one week.

1. Why do we implement the empirical risk minimization in the machine learning domain?
2. Why do we use minibatch algorithms other than batch algorithms for some optimization problems?
3. Why are neural networks generally not identifiable?
4. Why are local minima not a common problem plaguing neural network optimization?
5. Why do recurrent neural networks have the vanishing and exploding gradient problem?
6. How to integrate momentum into the stochastic gradient descent (SGD) optimization?
7. Why do some criteria for initial weights in deep networks often not lead to optimal performance?