

Large Scale Machine Learning and Deep Learning

Review Questions 5

1. Is it OK to initialize all the weights to the same value as long as that value is selected randomly using He initialization? Is it okay to initialize the bias terms to 0?

2. In which cases would you want to use each of the following activation functions: ELU, leaky ReLU (and its variants), ReLU, tanh, logistic, and softmax?

3. What is batch normalization and why does it work?

4. Does dropout slow down training? Does it slow down inference (i.e., making predictions on new instances)?

5. What may happen if you set the `momentum` hyperparameter too close to 1 (e.g., 0.99999) when using a `MomentumOptimizer`?