

Questions

“Training”

1. Sketch the ReLU, sigmoid, and tanh activation functions.
2. When training a neural network on a binary classification problem. Is it possible for the network’s training accuracy to stay the same even while the loss function decreases?
3. Why do we not count the input layer when we count the number of layers in a neural network?
4. When training a neural network, what does the optimizer optimize? Does the optimizer optimize the *same* quantity in each iteration, or a different one? Explain.
5. What can happen when a batch size is too small? Too large?
6. How is a training curve helpful in detecting overfitting?
7. What are hyperparameters? What are some examples of hyperparameters?
8. How do we tune hyperparameters?
9. Why is checkpointing important when you train a neural network?
10. Label the purpose of each line in the training code below:

```
for epoch in range(num_epochs):
    for data, actual in iter(train_loader):
        out = model(data)
        loss = criterion(out, actual)
        loss.backward()
        optimizer.step()
        optimizer.zero_grad()
```

Generalization

1. Bob proposes to use a function $f(x) = \min(x, 0)$. Will this activation function make the network easier to train or harder to train? (Hint: the ReLU activation can also be written as $\text{relu}(x) = \max(x, 0)$)
2. Suppose you have a dataset of 10000 labeled training data. If you choose a batch size of 50, how many iterations are in an epoch? What about if you choose a batch size of 100?
3. You find that with a batch size of 64 and a learning rate of 0.0001, your neural network trains too slowly. What can you do?

4. One “hyperparameter” we haven’t discussed is the initial (random) value of the parameters. Since neural network training does not find the *globally* optimal value of the parameters, the initial value of the parameters can be important. How can we tune this (set of) hyperparameters?
5. John and Sally are both training neural networks to classify cats vs dogs. John tuned his hyperparameters by training 20 different neural network models and choosing the best one. Sally tuned her hyperparameters by training 2000 different neural network models and choosing the best one. For the following questions, assume that the validation and test sets are fairly large.
 - Whose model do you think will have the better validation accuracy?
 - Would you expect Sally’s test accuracy be lower/higher than his validation accuracy?
 - Would you expect John’s test accuracy be lower/higher than his validation accuracy?
 - Consider the difference between the test accuracy and the validation accuracy. Would you expect this difference to be higher for Sally or for John?