Week 2 Comprehensive

TOTAL POINTS 10

1. **What does the equation for the loss function do conceptually?** 1 point

- (•) Penalize overconfidence
- ( ) Ignore historical statistical developments
- ( ) Mathematically define network outputs
- ( ) Reward indecision

2. What is overfitting? 1 point

- (•) Model complexity fits too well to training data and will not generalize in the real-world.
- ( ) Overfitting refers to the fact that more complexity is always better, which is why deep learning works.
- ( ) Model complexity is not enough to capture the nuance of the data and will under-perform in the real-world.
- ( ) Model complexity is perfectly matched to the data.

3. Why should the test set only be used once? 1 point

- ( ) More than one use can lead to overfitting.
- ( ) The model cannot learn anything new from subsequent uses.
- (•) More than one use can lead to bias.
- ( ) It is expensive to use more than once.

4. Which **two** of the following describe the purpose of a validation set? 1 point

- [x] To estimate the performance of a model.
- [x] To pick the best performing model.
- [ ] To test the performance in lieu of real-world data.
- [ ] To learn the model parameters.

5. How *do* we learn our network? 1 point

- ( ) Analytically determine global minimum
- (•) Gradient descent
- ( ) Downhill skiing
- ( ) Monte Carlo simulation

6. What technique is used to minimize loss for a large data set? 1 point

- ( ) Taylor series expansion
- ( ) Newton's method
- (•) Stochastic gradient descent
- ( ) Gradient descent

7. Which of the following are benefits of stochastic gradient descent? 1 point

- [x] With stochastic gradient descent, the update time does not scale with data size.
- [x] Stochastic gradient descent can update many more times than gradient descent.
- [ ] Stochastic gradient descent finds the solution more accurately.
- [x] Stochastic gradient descent gets near the solution quickly.
- [ ] Stochastic gradient descent finds a more exact gradient than gradient descent.

8. Why is gradient descent computationally expensive for large data sets? 1 point

- (•) Calculating the gradient requires looking at every single data point.
- ( ) Large data sets require deeper models, which have more parameters.
- ( ) There are too many local minima for an algorithm to find.
- ( ) Large data sets do not permit computing the loss function, so a more expensive measure is used.

9. What are the **two** main benefits of early stopping? 1 point

- [x] It helps save computation cost.
- [x] It performs better in the real world.
- [ ] It improves the training loss.
- [ ] There is rigorous statistical theory on it.

10. Why are optimization and validation at odds? 1 point

- ( ) Optimization seeks to generalize to the real world, while validation seeks to do as well as possible on a validation set.
- ( ) Optimization seeks to do as well as possible on a training set, while validation seeks to do as well as possible on a validation set.
- (•) Optimization seeks to do as well as possible on a training set, while validation seeks to generalize to the real world.
- ( ) They are not at odds—they have the same goal.