## DEEP LEARNING WEEK 7

1. We are training a neural network with the following patience parameter=2. When will the training stop?

| Epoch1 | Training error | Validation error |
|---|---|---|
| 1 | 4 | 5 |
| 2 | 3.5 | 3.8 |
| 3 | 3 | 3.3 |
| 4 | 2.2 | 3.6 |
| 5 | 2 | 3.6 |
| 6 | 1 | 3.4 |

a) After epoch 5
b) After epoch 4
c) After epoch 6
d) After epoch 3

**Answer:** a)
**Solution:** validation error doesn't decrease in epoch 4 and epoch 5 hence the training stops after epoch 5 since the patience parameter is 2.

2. Which of the following statements is true about the bias-variance tradeoff in deep learning? (MSQ)
a) High bias usually leads to overfitting
b) High variance usually leads to overfitting
c) High bias usually leads to underfitting
d) High variance usually leads to underfitting

**Answer:** b),c)

**Solution:** High bias occurs when the model is too simple and unable to capture the complexity of the underlying data. This leads to underfitting, where the model is unable to fit the training data well. On the other hand, high variance occurs when the model is too complex and able to fit the noise in the training data, leading to overfitting.

3. Which of the following techniques is used to reduce bias in deep learning models?
a) Dropout regularization
b) L1 regularization
c) Early stopping
d) Increasing the number of layers

**Answer:** d)

**Solution:** Increasing the number of layers increases the complexity of the model, which reduces bias by allowing it to capture more complex patterns in the data.

4. What is the purpose of adding noise to inputs in Deep Learning?
a) To make the model more complex
b) To reduce overfitting
c) To increase training speed
d) To fit the noise while training

**Answer:** b) To reduce overfitting

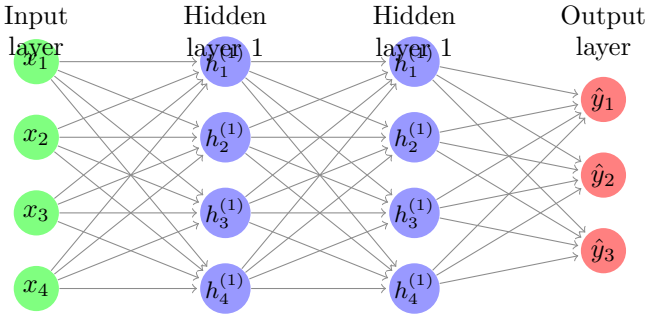
Solution: Adding noise to inputs in Deep Learning is a technique used to reduce overfitting.

5. Suppose that a model produces zero training error. What happens if we use L2 regularization, in general? (MSQ)
a) It might increase training error
b) It might decrease test error
c) Reduce the complexity of the model by decreasing the magnitude of less important weights
d) It might decrease training error

**Answer:** a),b),c)

**Solution:** Training can't go below zero .L2 regularization decreases the magnitude of weights and reduces overfitting thus often reducing test error and driving some weights towards zero.

6. Given the feed-forward network below how many thinned networks can be formed from it?




a)256
b)128
c)64
d)100

Answer: a)
Solution: No of thinned networks possible are $2^n$ and here n=8 (n is no of neurons in hidden layers).

7. We are training a neural network to distinguish dog and cat images. To increase learning effectiveness, we blur, rotate, and change some pixels. Which regularization technique are we using for the task?

a) Addition of noise during training
b) Data sharing
c) Parameter sharing
d) Data augmentation

Answer: d)
Solution: We are transforming the data in such a way that doesn't change the label hence enriching the dataset.

8. We trained different models on data and then we used the bagging technique. We observe

that our test error reduces drastically after using bagging. Choose the correct options.

a) All models had the same hyperparameters and were trained on the same features
b) All the models were correlated.
c) All the models were uncorrelated(independent).
d) All of these.

Answer: c)
Solution: If the models were correlated then the covariance of test errors would not be 0 hence test errors wouldn't reduce drastically. If all models have the same hyperparameters and train on the same set of data then they are correlated.

9. What is the usual relationship between train error and test error?
a) Train error is usually higher than test error
b) Train error is usually lower than test error
c) Train error and test error are usually the same
d) Train error and test error are unrelated

**Answer**: b)

**Solution:** In deep learning, the model is trained on a set of data and then tested on a separate set of data to measure its performance. The training error is calculated using the same data used to train the model, while the test error is calculated using new, unseen data. Since the model is optimized to fit the training data, it is expected to have a lower error on the training data than on new, unseen data. Therefore, the training error is always lower than the test error.

10. Consider two models $f_1(x) = w_0 + w_1 x$ and $f_2 = w_0 + w_1 x + w_2 x^2$. Which of these models has higher complexity?
a) $f_1$
b) $f_2$
c) Insufficient information

**Answer: b)**

**Solution:** The latter model has more parameters and can fit more functions.