hw\_07

1. **What is/are true about the advantages of batch normalization? (a, c, d, e)**
2. Selecting the initial weight value is no longer important.
3. Even at a very high learning rate, learning is always done properly.
4. BN can reduce the impact of changes in the distribution of learning data.
5. Increase the generalization performance of neural networks.
6. We can reduce the size of the feature space of the input data.
7. **What is/are true about Batch normalization? (b,c,d,e)**
8. normalize the output of each layer to mean 0 and variance 1
9. applies between hidden layers of neural networks
10. can speed up learning
11. can ease off overfitting
12. can convert the output of each layer into a normal distribution
13. **What function does the code below have? (d)**

| batch\_norm = nn.BatchNorm2d(num\_features) |
| --- |

1. Techniques for normalizing batch data
2. Normalizing the output of each layer to mean 0 and variance 1
3. Techniques for speeding up learning
4. Normalization technique applied between hidden layers of neural networks
5. Use only the number of classes entered in num\_features
6. **What is the running average of batch normalization? (a)**
7. To record the calculated mean and variance for each batch and update the recorded mean and variance for the previous batch with the following batch
8. To have the same distribution by aligning the input data calculated at each layer of the network with an average of 0 and a variance of 1
9. How to record updated data for each batch and proceed with learning
10. To connect each layer of a neural network
11. To record all batch data and turn it over to the next batch
12. **Which of the following statements about dropout is/are correct? (a,d)**
13. Dropout prevents complex co-adaptations in which a feature detector is only helpful in the context of several other specific feature detectors.
14. Dropout is more similar to L2 regularization than L1 regularization during training.
15. Dropout is active during training and testing.
16. Dropout can be viewed as a form of ensemble learning
17. The amount of dropout, p, can be optimized through standard stochastic gradient descent (SGD) methods