Hyperparemter Tuning Quiz

Daniel Abraham, Ganesh Vurimi,
Matthew Bronars, Warren Deng, Rasika Iyer

October 2020

1 Questions

1. What is the difference between a hyperparameter and a parameter?
2. What are some examples of hyperparameters that we have learned so far?
3. What happens to training error as lambda is increased in Ridge Regression?
4. What are some examples of hyperparameter tuning methods?
5. What is Grid Search?
6. What is Random Search?
7. When is Random Search better than Grid Search?
8. Why are hyperparameters hard to tune?
9. Are some hyperparameters in the same model more important than others?
10. We have created a classifier that requires us to tune its learning rate and ridge weight. This model is completely new to us, so we don't have a good understanding of what these values should be set to. Describe a method that will get us well tuned hyperparameters while keeping a modest run time.

2 Answers

1. Hyperparameters control how the model learns and are set before learning. Parameters are set by learning.
2. Learning Rate in gradient descent, Number of layers in neural network, Lambda in ridge regression
3. Training error increases since we care less about fitting the points and more about the size of our solution.
4. Grid Search, Random Search, Bayesian Optimization, Gradient-based optimization
5. Grid Search is a hyperparameter tuning technique that involves searching through every combination of hyperparameters in the model. It is an exhaustive search.
6. Random Search is a hyperparameter tuning technique that involves searching through a predetermined number of random combinations of the hyperparameters.
7. Random Search is typically better than Grid Search when there are a large number of hyperparameters. This is because Grid Search has to look through every combination of hyperparameters, which could be very inefficient.

8. There are a large variety of hyperparameters, which differ not only in dimensionality but in type (for example type of activation function and degree are both hyperparameters).

9. Yes, hyperparameters do not equally affect accuracy. For example in a neural network, learning rate may be more important than number of neurons in a single hidden layer.

10. You can start with a random search to find the general region where these parameters have low loss. From here you can perform a more localized grid search and select your minimum. Another acceptable answer would be performing grid search over different degrees of these parameters to find an appropriate magnitude for each value. Again, once we have this we can perform another more localized grid search and select the values that lead to minimum loss.