

Ethan Gillespie

INFX 499

Ch. 4 Exercises

10/11/21

1. The linear regression training algorithms you can use if you have a training set with millions of features are mini-batch gradient descent, stochastic gradient descent, and batch gradient descent.
2. If your training set has different scales, the function that will have an elongated bowl is the cost function. This causes the gradient descent algorithms to take a longer time to converge. What we can do about this is scaling your data before training your model.
3. The Gradient Descent cannot get stuck because the cost function is convex.
4. All gradient descent algorithms will eventually end up being similar if the problem is convex and the learning rate is not high.
5. If you notice that at every epoch there is a validation error then you can assume that the learning rate can be too high, and diverging can be happening to the algorithm. You can either reduce the learning rate or stop the learning process.
6. No you should save at regular intervals and if the model never beats the record that was set then you can go back to the best saved model.
7. The algorithm with the fastest training is stochastic gradient descent because it only looks at one training occurrence at a time. With that being said it is the fastest to get to the vicinity of the global optimum. The batch gradient descent is the only algorithm that will converge.
8. The problem that is occurring is the model is overfitting the training set. To fix this you can lower the polynomial degree, or you can regularize the model. This can lower the degrees of freedom of the model.
9. The problem that is happening is the model is underfitting and to fix this you can lower the regularization hyperparameter.
10. You would want to use ridge regression over plain linear regression because ridge regression has some regularization and performs better. Ridge regression is preferably used instead of lasso because lasso tends to push weights down to zero and this can lead to sparse models. Lastly, elastic net is used more than lasso because lasso can sometimes be inconsistent and uses an extra hyperparameter.
11. To classify pictures for daytime/nighttime and outdoor/indoor, you should use the logistic regression classifiers.
12. On GitHub.