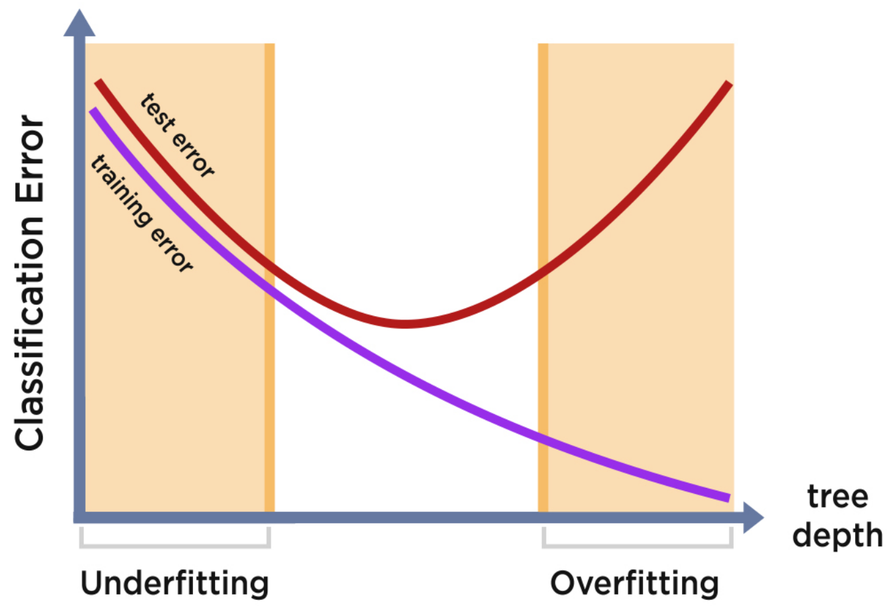
Underfitting and Overfitting

There are two equally problematic cases which can arise when training a classifier on a data set: underfitting and overfitting. Each case relates to the degree to which the data in the training set is extrapolated to apply to unknown data. If the classifier isn't able to distinguish important aspects of the data, it will fail to classify new data accurately. On the other hand, if the classifier learns the idiosyncrasies that are particular to only the training dataset, it will fail to generalize to new data.

Underfitting: The classifier learned on the training set is not expressive enough to accurately classify training data. In this case, both the training error and the test error will be high, as the classifier does not account for relevant information present in the training set.

Overfitting: The classifier learned on the training set is too specific and cannot be used to accurately infer anything about unseen data. Although training error continues to decrease over time, test error will begin to increase again as the classifier begins to make decisions based on patterns which exist only in the training set and not in the broader distribution.

> An underfit model has not captured the predictive features present in the data and will perform poorly on new data.
>
> An overfit model has captured the idiosyncrasies of the training data and will not generalize to unseen data.
>
> We can use plots of training and test error to assess the fit of our model.



You are training an image classifier to recognize pictures of birds. You notice that you can tune a hyperparameter in your model to give you nearly 100% accuracy on the training set (i.e., every image in the training set is correctly classified). When you look at the validation set, which contains images the classifier has never seen before, you notice that the classifier achieves 30% accuracy. What's happening?

- The model is performing as well as it can and you should deploy it.
- ✔ You are likely overfitting to the training set. The classifier is learning idiosyncrasies specific to the training data and cannot generalize to new examples.
- You are likely underfitting to the training set. You need to continue training the classifier on the training data in order to improve the accuracy on validation (testing) data.
- This is theoretically impossible; your code must have a bug.

1/1