**Gradient Boosting**

**Boosting Algorithm: When you take small models and make one big model.**

**The loss function for gradient boosting is just (actual – predicted), which is also known as pseudo-residual**

**Note: For this specific algorithm, it is generally best to use decision trees**

* **The catch with gradient boosting is that once you get the residual error, you train your model on those residuals. In other words, you train the model to try and predict how much error it will make.**

**Once you train model 2 on the residuals of model 1, the newer prediction will be the prediction of model 1 +(lr \* model 2). The learning rate is important so that it ensures that you reach the correct answer more and more accurately in a slow manner. Otherwise, overfitting can occur.**

**Essentially, you change model 1’s prediction based on residuals. Bigger residual means more that you change model 1’s prediction. Changing model 1’s prediction ensures that residuals decrease**

**The main goal is the add more and more models, such that the residual decreases more and more but in a slow manner, and you get closer to the more accurate prediction**

**Making your predictions and adjusting that prediction by residual error is bound to decrease the next residual values by nature**