Decision Tree

Min bias – huge variance

Low mean sqred error

Tree depth – how many questions you’re aloowed to ask

Min # of observations in a leaf

(min leaf node)

splits have to be numeric

In classification Gini determines the tree in regression ir is low mean sqrd error

Disadvantage:

Not great for unbalanced data

Not good with small sets, hard to split

I think my project will be linar regression…which sucks. The one day I missed

Ensembling

Pretty cool, like the wisdom of the commons or democracy in action

Only time when overfitting is good, gives weights to the experts