LinearRegression(copy\_X=True, fit\_intercept=True, n\_jobs=1, normalize=False)

Zone 1 0.63

Zone 2 0.19

Zone 3 0.60

Zone 4 0.92

Zone 5 0.95

Zone 6 0.88

Zone 7 0.75

Zone 8 0.85

Zone 9 0.68

Zone 10 0.76

Mean R2 for London: 0.720815917662

RandomForestRegressor(bootstrap=True, criterion='mse', max\_depth=19,

max\_features='auto', max\_leaf\_nodes=None,

min\_impurity\_decrease=0.0, min\_impurity\_split=None,

min\_samples\_leaf=1, min\_samples\_split=2,

min\_weight\_fraction\_leaf=0.0, n\_estimators=21, n\_jobs=1,

oob\_score=False, random\_state=None, verbose=0, warm\_start=False)

Zone 1 0.52

Zone 2 0.72

Zone 3 0.80

Zone 4 0.90

Zone 5 0.97

Zone 6 0.90

Zone 7 0.85

Zone 8 0.84

Zone 9 0.82

Zone 10 0.67

Mean R2 for London: 0.797005857026

GradientBoostingRegressor(alpha=0.9, criterion='friedman\_mse', init=None,

learning\_rate=0.1, loss='ls', max\_depth=3, max\_features=None,

max\_leaf\_nodes=None, min\_impurity\_decrease=0.0,

min\_impurity\_split=None, min\_samples\_leaf=1,

min\_samples\_split=2, min\_weight\_fraction\_leaf=0.0,

n\_estimators=100, presort='auto', random\_state=None,

subsample=1.0, verbose=0, warm\_start=False)

Zone 1 0.48

Zone 2 0.67

Zone 3 0.84

Zone 4 0.89

Zone 5 0.96

Zone 6 0.89

Zone 7 0.86

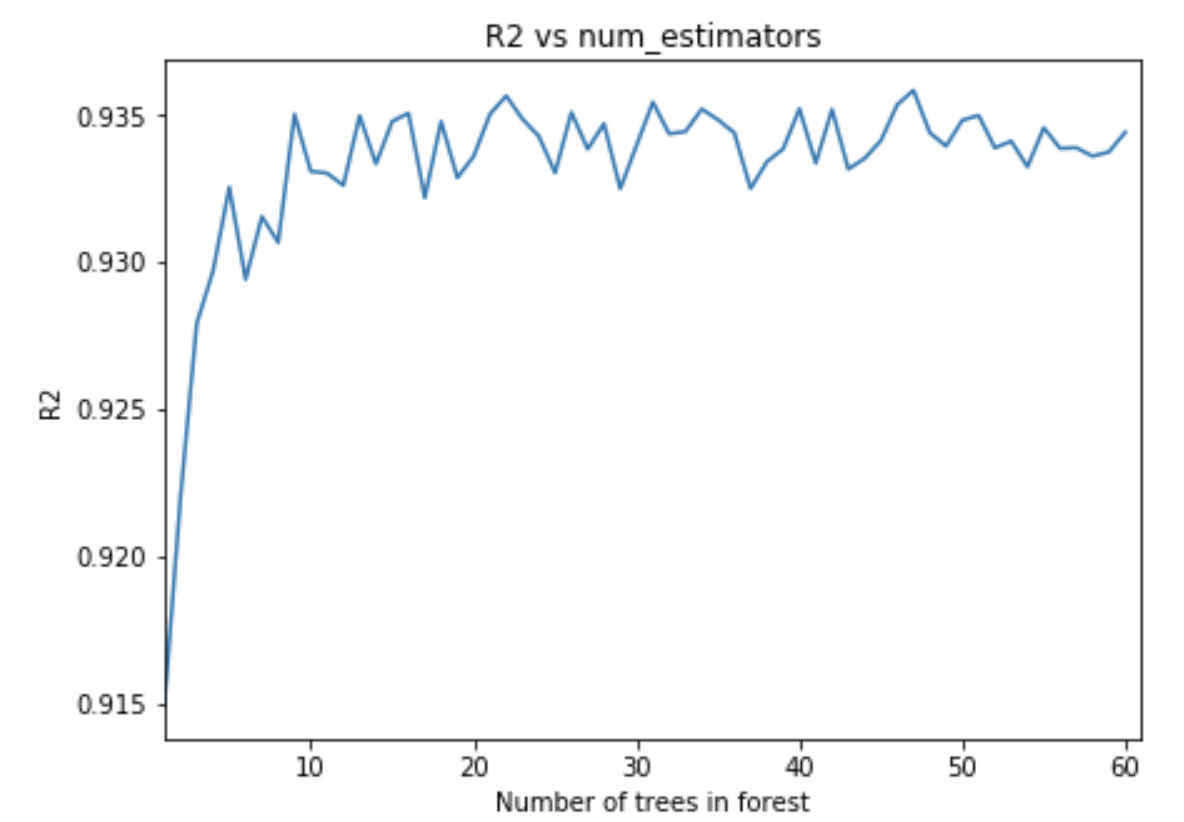
Zone 8 0.80

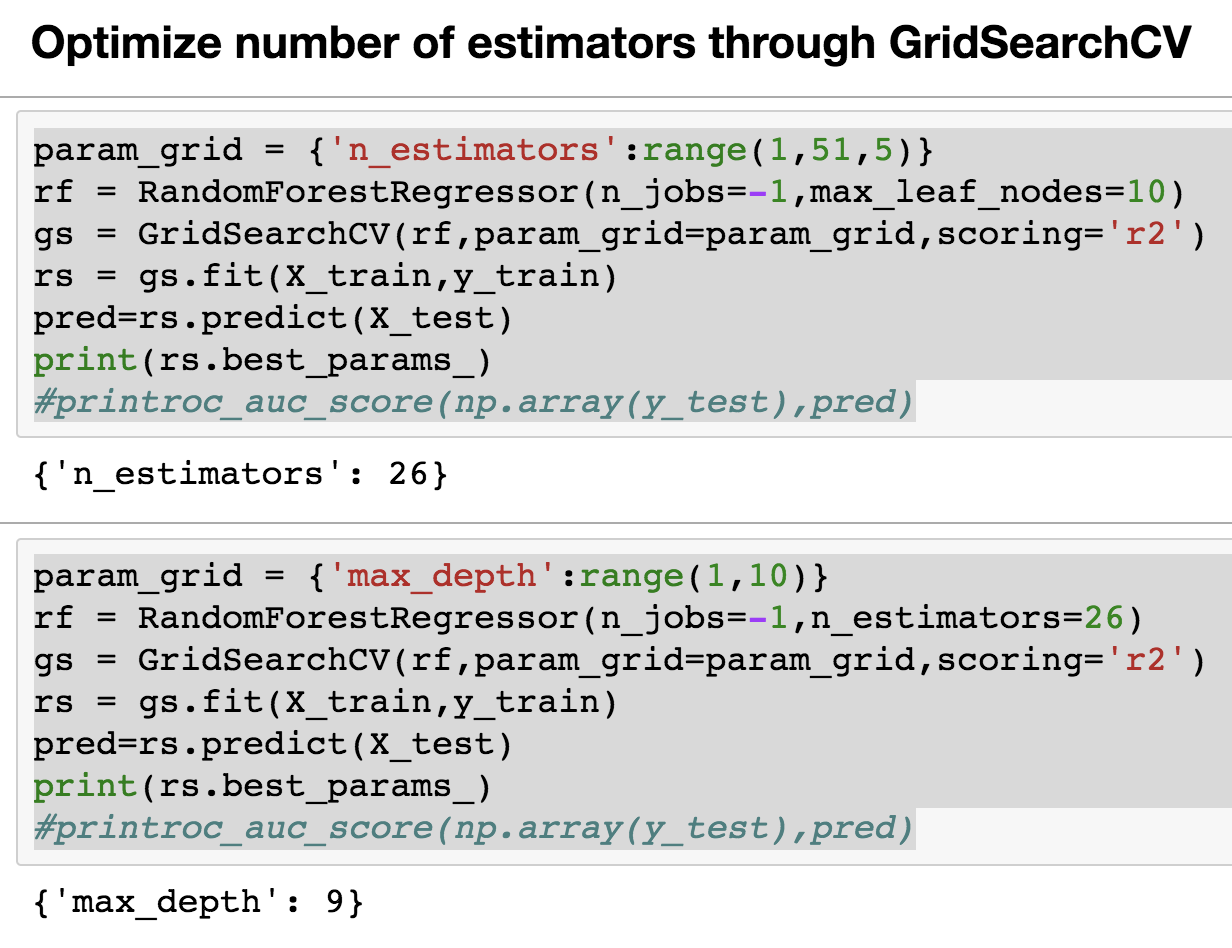
Zone 9 0.80

Zone 10 0.74

Mean R2 for London: 0.791467835667

#### Build random forest regressors with between 1 and 50 trees. See how the R2 changes.



****

**Model options**

**1. Linear regression (OLS, ridge, lasso)**

"Ridge and Lasso regression are powerful techniques generally used for creating parsimonious models in presence of a ‘large’ number of features."

*Parameters*:

* ridge: Objective = RSS + α \* (sum of square of coefficients)
* lasso: Objective = RSS + α \* (sum of absolute value of coefficients)

So we need the optimal alpha for either of these.

**2. SVM regression**

* linear kernel
* RBF (gaussian) kernel

Parameters:

* linear needs C (term to penalize observations across the boundary lines)
* RBF needs C and gamma. ("If gamma is large, then variance is small implying the support vector does not have wide-spread influence. Technically speaking, large gamma leads to high bias and low variance models, and vice-versa.")

**3. Ensemble tree methods**

3.1: Random forest Parameters

* number of trees
* max depth

Advantage: visualize feature importance in the tree structure

3.2 Gradient boosted trees regression

* min child weight (ie. minimum observations for a new leaf)
* max iterations (keep it as high as your computation budget allows)
* max depth

Similar to random forest but the trees learn off each other rather than being constructed independently.