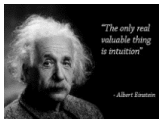
Random Forest Models



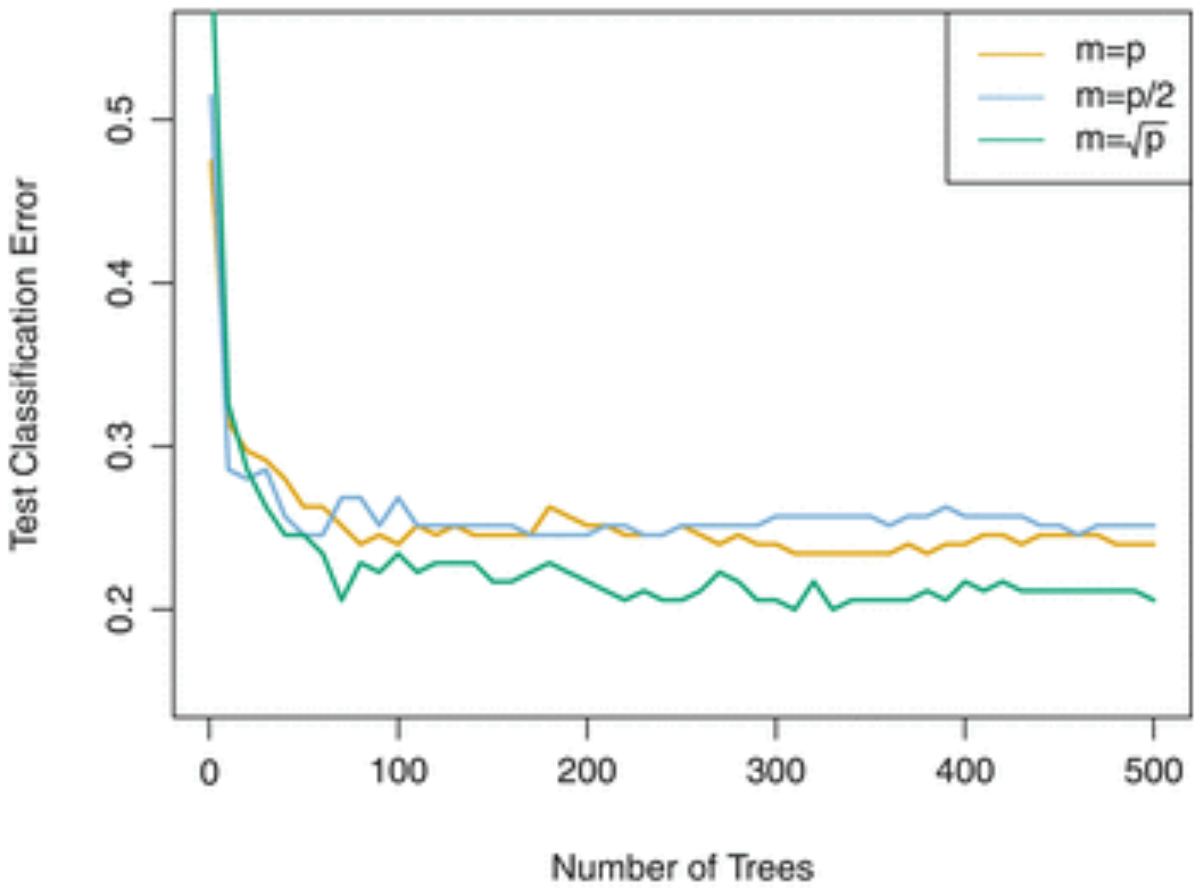
Random Forests Models: Intuition

- **Bagging** helps reduce variance, but the **models** generated each time have some **correlation** because the models use the **same predictors**, so the generated models will be somewhat **similar**
- For example, if there is one **strong predictor, Bagging** will often **select** this predictor at the **top** of the **tree** every time, thus generating somewhat similar models, so the **contribution** to **variance reduction** will be **small**.
- **Random Forests** overcomes this problem by also keeping the **predictors** included in **each model** iteration **random**
- That is if there are P **predictors** in the model, **Random Forests** will only use M **predictors**, with $M < P$ in each model, but these M predictors will be **different** and **randomly** selected each time.
- The number of M predictors to select is a **tuning** parameter
- **Random Forest** models are particularly useful when the number of predictors P is **large**
- As M **approaches** P Bagging and Random become more **similar**

Random Forest Illustration: Varying M

Observations

1. For $M = P$ is identical to **Bagging**
2. As the **number** of sampled **trees** increase, **error** ↓ sharply **initially** and then flattens
3. As M becomes **smaller** relative to P the test classification **error** ↓ (up to a point; needs to be tested with cross-validation)



R Tips

`randomForest(){randomForest}` → Function used to fit various Random Forest models (please note the cap F)

```
rand.fit=randomForest(y~x1+x2+etc.,data=dataName,
                      mtry=5,importance=TRUE)
```
→ `mtry=5` tells Random Forest to use 5 predictors; if p is smaller than the full set of predictors in the model, then `mtry=p` generates a Random Forest model; by default this method computes 500 trees.

```
rand.fit.25=randomForest(y~x1+x2+etc.,data=dataName,
                      mtry=5,importance=TRUE,
                      ntree=25)
```
→ Use the `ntree=25` attribute to change the default number of trees to generate (to 25, for example)

KOGOD SCHOOL
of
BUSINESS