* Breiman (2001) is credited as being the first to introduce the Random Forest (RF) algorithm
* In some cases, bagging and aggregating (a step for creating an RF) can worsen the expected misclassification rate
* Oshiro (2012)’s paper suggests that there is an upper limit on the number of trees, after which the performance gain stops being significant.
* This paper introduces an R package, OOBCurve, that can be used to examine the performance of an RF
* For most datasets, increasing the number of trees continues to improve RF performance, even if only marginally. For a few datasets, performance could hit a minimum, then increase, then plateau.
  + The classical performance measures of Log Loss and Sum of Squared Errors do not capture this possible peculiarity. Only Area Under the Curve (AUC) is able to reveal it.

**Section 2: Introducing the Random Forest**

* An ensemble learning technique that brings together *T* decision trees
  + A label prediction is obtained by aggregating the predictions made by all the trees:

Average prediction in the case of Regression

* + - ; y\_t is the value predicted by an individual tree

Majority classification in the case of Classification

* + - ; the probability the instance is of class C is a count of how many trees predict C divided by the number of trees
* This reduces the variance experienced by just a single Dtree
* There are variants to the original RF algorithm introduced by Breiman that attempt to address some of its flaws such as variable selection bias.
* Errors are measured with the usual equations:

SSE = for Regression

* + The error of individual trees can also be quantified with this equation. It is not limited to measuring just the entire RF’s prediction.

LL = for Binary Classification

* Some new measurements introduced by the paper:

Average error =

* + For binary classification
  + Can be estimated as ?

Brier score =

* + For binary classification
  + Actual label minus the probability of that label being predicted?
* See top of pg. 6 for AUC equation
* See pg. 6 for Multiclass Classification equations
* Rousseeuw (1984) proposes taking the median error to limit the effect of outliers
* Out-of-Bag (OOB) estimations involve measuring the error on an instance of the dataset that was not used to create the tree it is testing.

**Section 3: Theorems about the RF**

* This section goes over the derivations of each expected-error function

**Section 4: Description of the tests**

**Section 5: Recommendations**

* The authors recommend not relying on error rates to judge when the algorithm has converged on an optimal solution. Instead, they recommend using AUC, the Brier score, or the log loss for classification problems.
* They seem to suggest 128 trees is usually the ideal amount to grow?
* They recommend that training should stop if the most recently created tree does not improve performance by some hyper-parameter determined by the programmer.
  + This parameter should be set with the tradeoff of accuracy and computation time in mind.