0.1 Random Forests, an Ensemble Method

The random forest (Breiman, 2001) is an ensemble approach that can also be thought of as a form of nearest neighbor predictor.

Ensembles are a divide-and-conquer approach used to improve performance. The main principle behind ensemble methods is that a group of weak learners can come together to form a strong learner. The figure below (taken from here) provides an example. Each classifier, individually, is a weak learner, while all the classifiers taken together are a strong learner.

The data to be modeled are the blue circles. We assume that they represent some underlying function plus noise. Each individual learner is shown as a gray curve. Each gray curve (a weak learner) is a fair approximation to the underlying data. The red curve (the ensemble strong learner) can be seen to be a much better approximation to the underlying data.

I'm running a random forest model using R's caret package, and running varImp on the returned object gives me the averaged variable importance across the number of bootstrap iterations. However, I would rather assess variable importance for each iteration. Is this possible using the caret package?

Reproducible example:

```
library(caret)
mod <- train(Species ~ ., data = iris,
         method = "cforest",
         controls = cforest_unbiased(ntree = 10))
varImp(mod)
```

returns:

```
cforest variable importance
Overall
Petal.Width  100.0000
Petal.Length  86.6279
Sepal.Length   0.5814
Sepal.Width    0.0000
```

what I'm interested in is rather a list of length=number of bootstrap resamples with variable importance for each iteration. This might be possible using some combination of returnResamp = "all" and a custom summaryFunction but I'm not wise enough to know.