随机森林算法：

Step1：create a bootstappped dataset

我们从需要预测的数据中，有放回的随机取约 60%的数据，形成我们的 bootstrapped dataset

Step2: Create a decesion tree using the bootstraped dataset. But when it come to the feature to be selected in each node, we random choose n feature to be compared.The n can be 1,2,3,.....
We later will learn how to determine the best n to build the forest.

Step3: Go back to step 1 and repeat.That is,making a new bootstrapped data set and form a tree in step 2's method.
We can build up to a hundred trees using this method,so this is why it is called a forest.

step 4:When we do predicting,we run the sample though all the trees,and vote to determine the final result.

我们如何评价的我们的随机森林模型呢？通常来说，因为我们的 bootstrapping 取样，每次大概会有三分之一的数据不会被取到，这些数据叫 out of bag dataset.用这些 out of bag dataset 来测试我们的随机森林模型。

概念：the porprotion of out-of-bag sample that were incorrectly classified is the 'out-of-bag error'

我们用 out of bag error 的大小来判断我们模型的准确程度。

最后我们再来看一下 n 的选取问题，最好的方式是 n 取遍所有值，比较不同 n 下的 out of bag error,选最小的那一个。