#### 集合模型面试问题汇总

1. 请简要描述bagging and boosting.
2. bagging：

如果我们的模型有某种过拟合问题，比如决策树，可以考虑采用bagging方法。

bagging的思想就是有放回地、选择不同特征的采样，以训练出不同的模型。再由这些模型投票得到最终结果。这应该也是随机森林的思想。

1. boosting：

我们有一个弱分类器，欠拟合。比如深度很浅的决策树。我们每次训练模型，会有预测正确的和预测错误的。我们比较关心那些预测错误的样本，所以我们将增加错误样本的权重。就这样迭代下去，产生不同的模型，这些模型对应的权重也有所差异。最终通过投票得到最终预测。

注：决策树训练过程中一个字段为feature\_importances\_ 特征重要性。