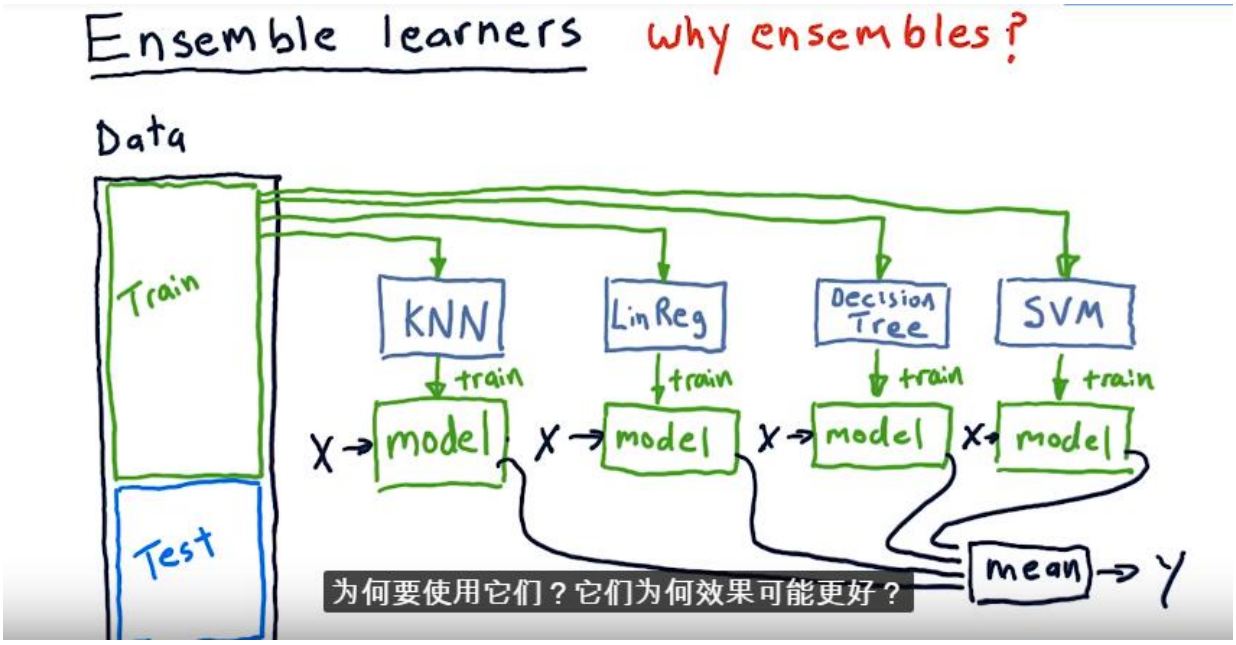
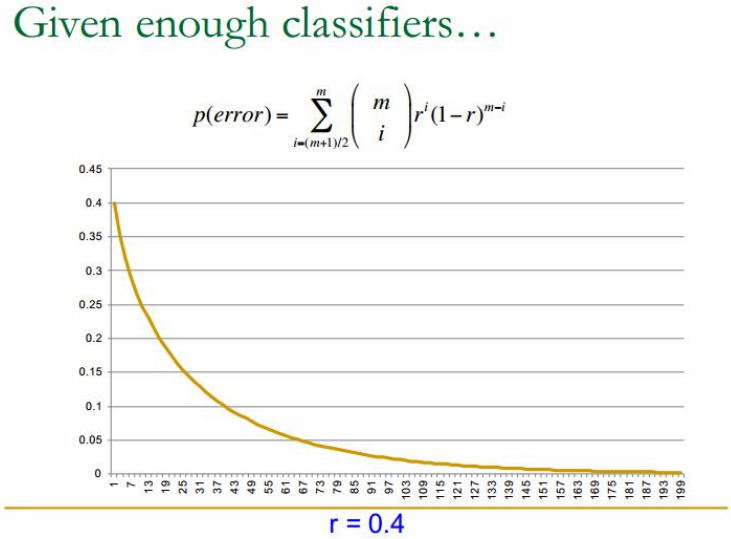


Ensemble learning(集成学习)：
训练不同的算法（分类器 classifier），将 X 导入这些分类器，然后通过投票或者是均值的方法来预测 Y

The benefit of ensemble learning:假设每个不同的分类器的准确率都有 60%的话，我们训练出了 3 个分类器，别为 logistic regression，KNN，Decesion tree.在这种情况下，单个分类器的准确程度只有 60%，但是如果把这些分类器结合起来，采用投票的方式，多数服从少数，则会提升预测的概率：
此时的预测准确率为：

$$0.6\times0.6\times0.6 + C_3^2\times0.6\times0.6\times0.4$$
$$= 0.648$$

可以看到，预测的准确率提升了。由此，我们可以看到，只要每个分类器的准确率大于 0.5，且分类器尽可能地多，准确率就可以提得很高。
PPT 中给出了错误率的计算公式和图



可以看到，在准确率 0.6，错误率 0.4 的情况下，有 60 多个分类器就可以把准确率升到 95%以上。

缺点：
①我们上面计算集成准确概率时，用的概率是伯努利分布，要求每个事件互相独立才能那么算。
如果分类器互相不独立，比如说都是线性模型，那么是不会提升准确率的。
②准确率是我们根据测试集结果得出的，并不是绝对的概率。