附加问题 1 证明：集成学习器的泛化错误率随着个体学习器数量增加而越来越小？

符号	定义
S（x）	泛化错误率
m	总体样本变量样本容量
a	分类错误样本个数
h_i	个体学习器（弱学习器）
f（x）	分类准确的真实函数
W	弱学习器的泛化错误率
x	样本表示
H（x）	集成分类学习器
T	弱学习器个数

泛化错误率是集成学习器在一个样本变量上的误差，$S(x)$的定义公式可为：

$$S(x)=P(x\neq f(x))=\frac{a}{m}(a\leq m)$$

假设集成分类学习器中个体学习器（弱学习器）的分别$h_1,h_2,\cdots,h_T$组成，在分类问题中每个弱学习分类器的错误率为w，则每个弱学习分类器的泛化错误率定义为：

$$P(h_i(x)\neq f(x))=w$$

假设 *w* 之间的相互独立，即弱学习器之间误差相互独立，设定当超过 k 个集成分类器正确则可以认定集成分类学习器最终分类准确，即：$H(x)=sign(\sum_{i=1}^{T}h_i(x))$

由伯努利分布可知，假设正面朝上的概率是 p,反面朝上的概率则为 1-p，投掷 n 次，正面朝上次数的期望值为 np, n 次中正面朝上的次数 $H(n)$存在以下不等式：

$$P(H(n)\leq R)=\sum_{i=0}^{R}\binom{n}{i}p^i(1-p)^{n-i}$$

直觉上，如果我们有更多的样本则样本单独期望应该越来越接近总体期望，这可由**伯努利分布的特例—Hoeffding 不等式**可知当随机变量相互独立时，

$$\begin{cases}\bar{X}=\dfrac{X_1+\cdots+X_n}{n}\\ \forall t>0\\ P(E(\bar{X})-\bar{X}\geq t)\leq\sqrt{-\dfrac{2n^2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}}\end{cases}$$

当样本数量逐渐变多时，不等式的越来越接近 0，所以样本期望越来越接近总体期望。即最终的集成分类学习器的泛化错误率可以定义为：

$$P(H(x)\neq f(x))=\sum_{i=0}^{k}\binom{T}{i}(1-w)^i w^{T-i}\leq\sqrt{-\frac{1}{2}T(1-w)^2}$$

由此可知，集成学习中个体分类器数目不断增大，集成学习的错误率呈指数级下降，最终趋近于 0。

附加问题 2 证明：分类问题中，个体学习器是弱学习器，弱学习器之间的差异性越大，集成效果越来越好？

解：假设集成学习器中个体学习器（弱学习器）的分别 $h_1, h_2, \cdots, h_T$ 组成，在分类问题中每个弱学习分类器的错误率为 w，则每个弱学习分类器的泛化错误率定义为：

$$P(h_i(x) \neq f(x)) = w$$

加权平均法的集成学习器满足：

$$H(x) = \sum_{i=1}^{T} a_i h_i(x)$$

通过加权平均法结合产生的集成来完成分类学习任务，对于样本变量，定义个体学习器之间的差异性可以为：

$$C(h_i | x) = (h_i(x) - H(x))^2$$

则集成学习器的差异性是：

$$\bar{C}(h | x) = \sum_{i=1}^{T} a_i C(h_i | x) = \sum_{i=1}^{T} (h_i(x) - H(x))^2$$

分类任务中，最常用的是正确率（分类正确的样本占总样本的比例）和错误率（分类错误样本占总样本的比例），对于样本集 $D\{(x_1, y_1), \cdots, (x_i, y_i)\}$ ，其中 y_i 是真实值，f 是学习到的分类学习器。正确率和错误率分别定义为：

$$acc(f : D) = \frac{1}{m} \sum_{i=1}^{m} I[f(x_i) = y_i] \qquad E(f : D) = \frac{1}{m} \sum_{i=1}^{m} I[f(x_i) \neq y_i]$$

将分类结果划分为 TP,FP,TN,FN 四种情形，可以得到真正率（TPR），假正率（FPR），公式如下：

$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{TN + FP}$$

由问题一可以证明，个体学习器增加，错误率却反而越来越小。猜想应该是分类过程中，差异性会影响假正率的值，模型多次计算，ROC 曲线越接近左上角（1,1）点，将 ROC 的曲线面积取为 AUC 值，可以得到越接近 1 则说明模型性能越好。