第5章 集成学习

Exercise 5.1

以下不属于bagging的特点是（）

A、有放回抽样多个子集

B、训练多个分类器

C、最终结果为每个学习器加权后的线性组合

D、可以减少过拟合

Exercise 5.2

在Baggging集成学习中，多样性是通过（）实现的。

A、数据样本扰动

B、输入属性扰动

C、输出表示扰动

D、算法参数扰动

Exercise 5.3

有关集成学习下列说法正确的是（）

A、基本模型应尽量来自于同一算法，通过改变训练数据和参数，得到不同的基本模型

B、通常来讲，基本模型之间相关性应该低一些

C、集成的基本模型的数量越多，集成模型的效果就越好

D、bagging boosting 时常用的集成学习的方法

Exercise 5.4

Consider three decision stump hypotheses from $\mathbb{R}$ to $\{-1, +1\}$: $g_1(x) = sign(1-x), g_2(x) = sign(1+x), g_3(x) = -1$. When mixing the three hypotheses uniformly, what is the resulting $G(x)$?

A、$2[\![|x| \leq 1]\!] - 1$
B、$2[\![|x| \geq 1]\!] - 1$
C、$2[\![|x| \leq -1]\!] - 1$
D、$2[\![x \geq 1]\!] - 1$

Exercise 5.5

Consider three decision stump hypotheses from $\mathbb{R}$ to $\{-1, +1\}$:
$g_1(x) = sign(1 - x), g_2(x) = sign(1 + x), g_3(x) = -1$. When $x = 0$,
what is the resulting $\Phi(x) = (g_1(x), g_2(x), g_3(x))$ used in the returned
hypothesis of linear/any blending?
A、 $(+1, +1, +1)$
B、 $(+1, +1, -1)$
C、 $(+1, -1, -1)$
D、 $(-1; -1; -1)$

Exercise 5.6

For four examples with $u_n^{(1)} = \frac{1}{4}$ for all examples. If g_1 predicts the first
example wrongly but all the other three examples correctly. After the
'optimal' re-weighting, what is $u_1^{(2)}/u_2^{(2)}$?
A、 4
B、 3
C、 1/3
D、 1/4

Exercise 5.7

According to $\alpha_t = ln(\lambda_t)$, and $\lambda_t = \sqrt{\frac{1-\epsilon_t}{\epsilon_t}}$, when would $\alpha_t > 0$?
A、 $\epsilon_t < \frac{1}{2}$
B、 $\epsilon_t > \frac{1}{2}$
C、 $\epsilon_t \neq 1$
D、 $\epsilon_t \neq 0$

Exercise 5.8

比较支持向量机、AdaBoost，逻辑回归模型的的学习策略和算法。