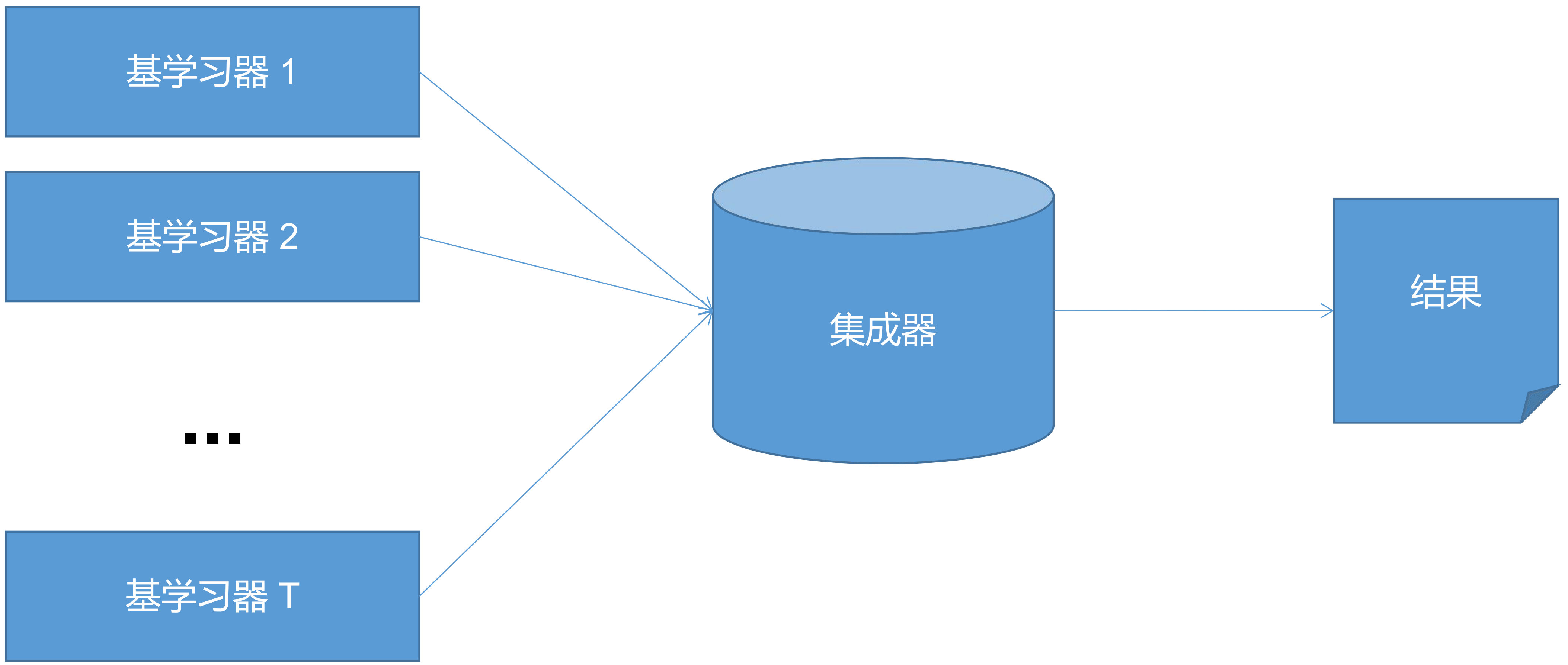
集成学习 （Ensemble Learning）

基础概念

集成学习



集成的效果

	测试例1	测试例2	测试例3
h_1	√	√	×
h_2	×	√	√
h_3	√	×	√
集成	√	√	√

(a) 集成提升性能

	测试例1	测试例2	测试例3
h_1	√	√	×
h_2	√	√	×
h_3	√	√	×
集成	√	√	×

(b) 集成不起作用

	测试例1	测试例2	测试例3
h_1	√	×	×
h_2	×	√	×
h_3	×	×	√
集成	×	×	×

(c) 集成起负作用

好而不同

集成方法

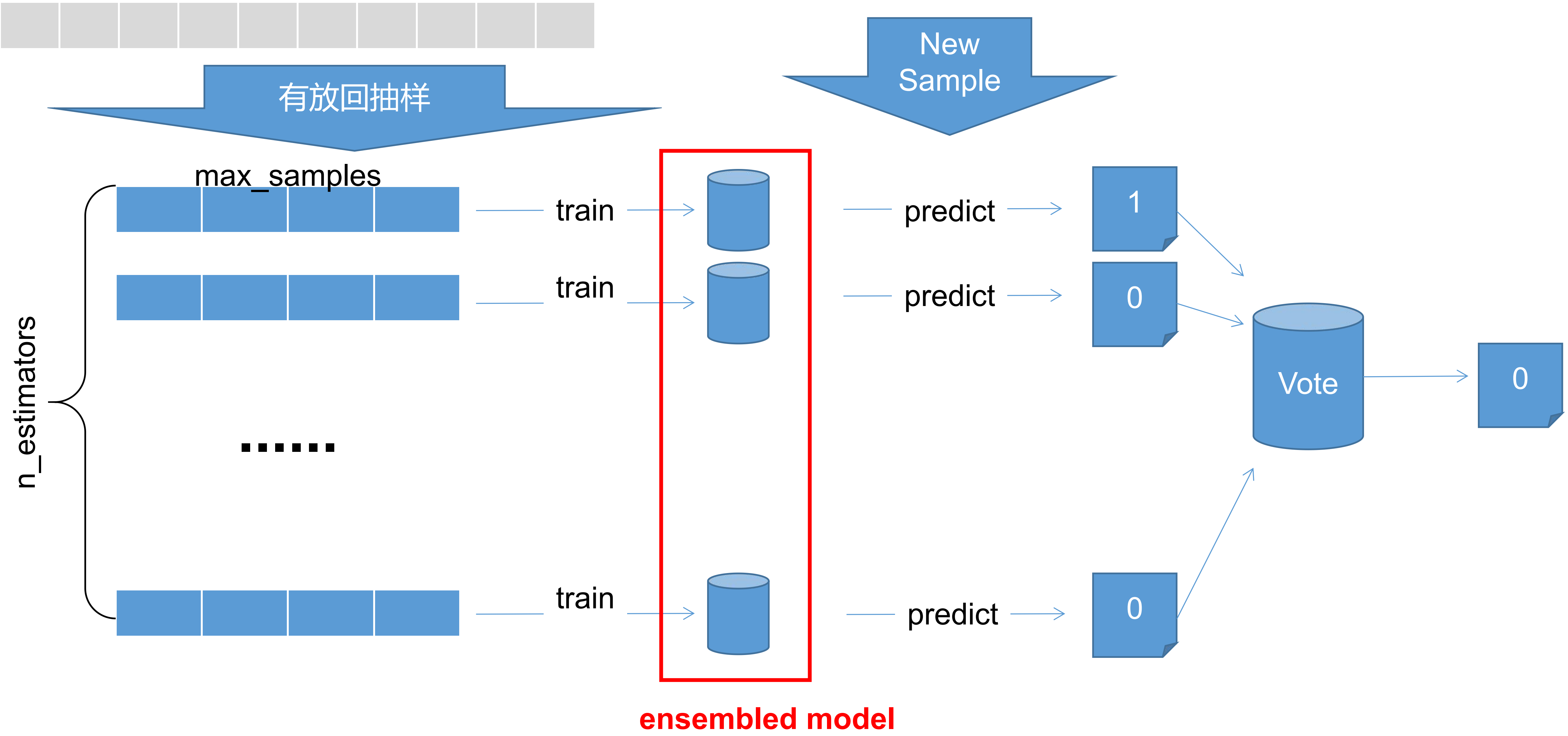
Bagging

Stacking

Boosting

Bagging (Bootstrap Aggregating)

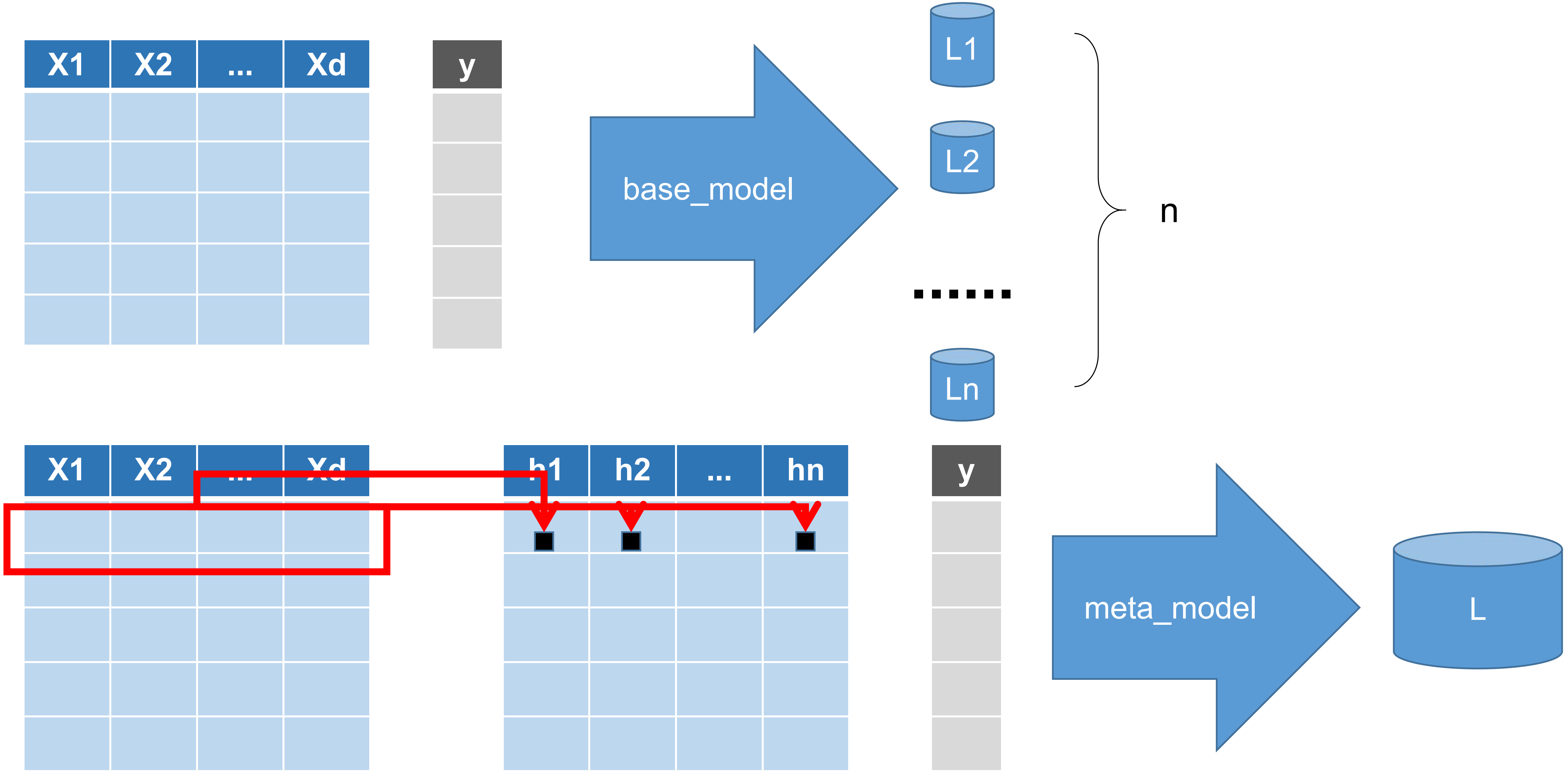
Bagging



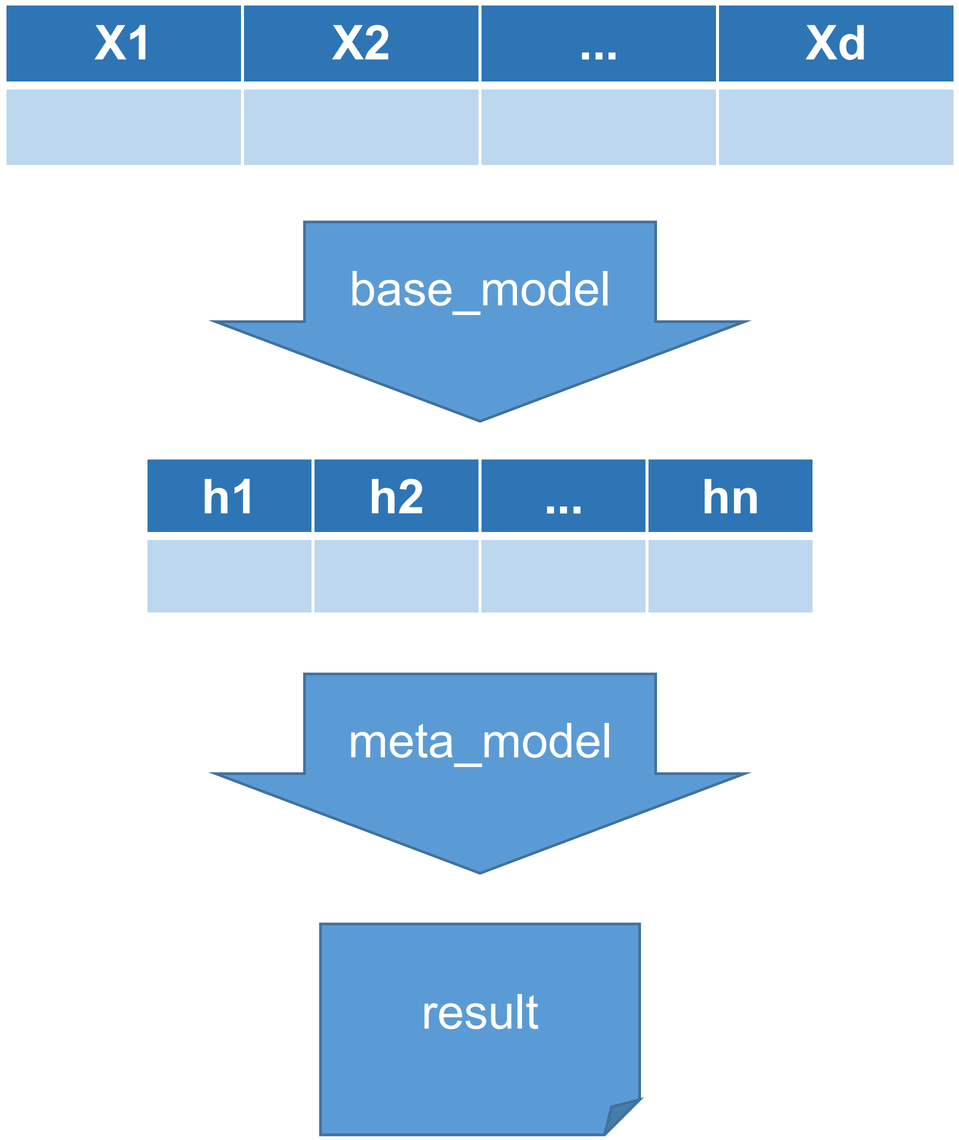
有放回抽样
New Sample
max_samples
n_estimators
train
train
......
train
predict
predict
predict
1
0
0
Vote
0
ensembled model

Stacking

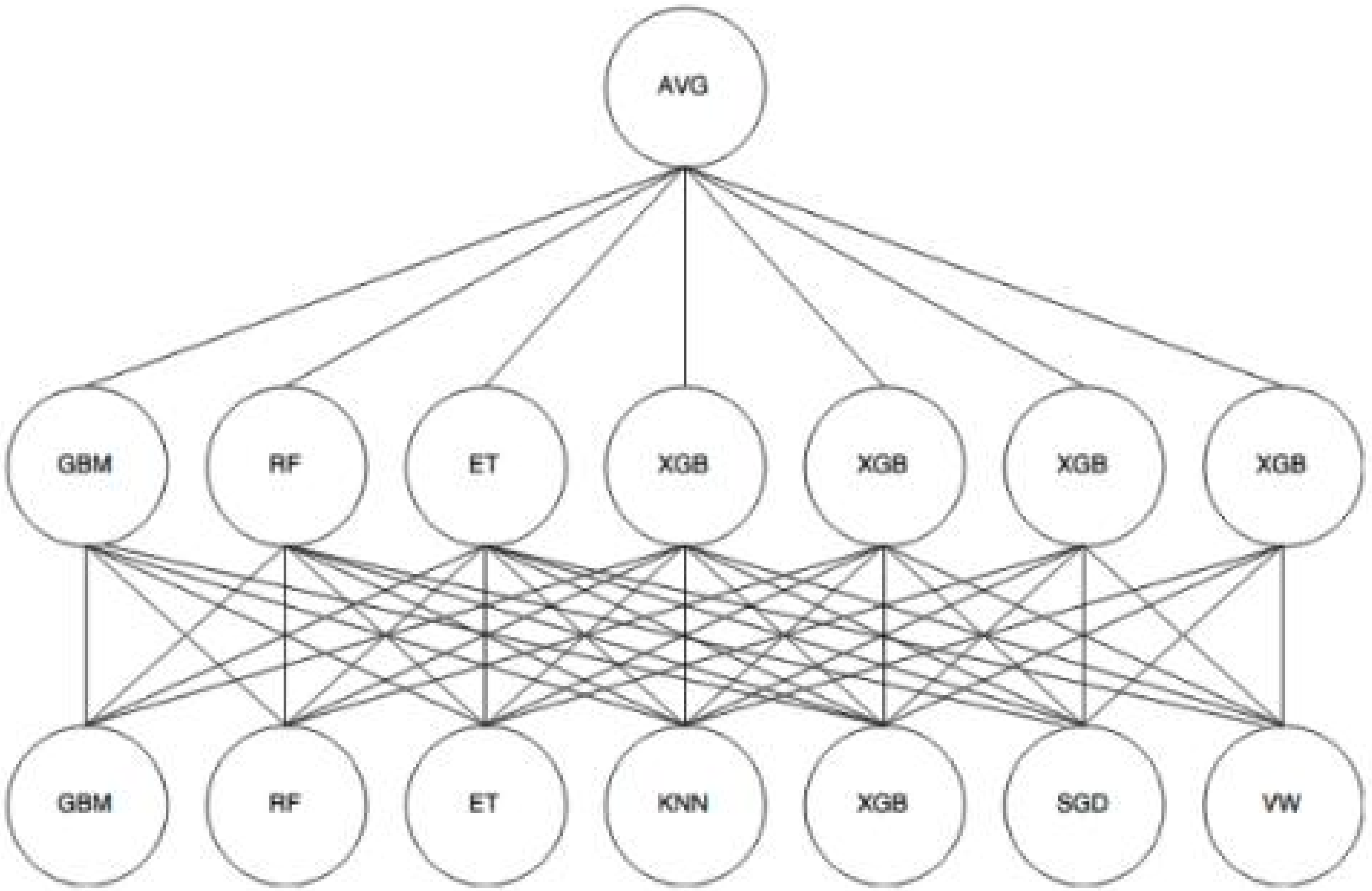
Stacking



Stacking

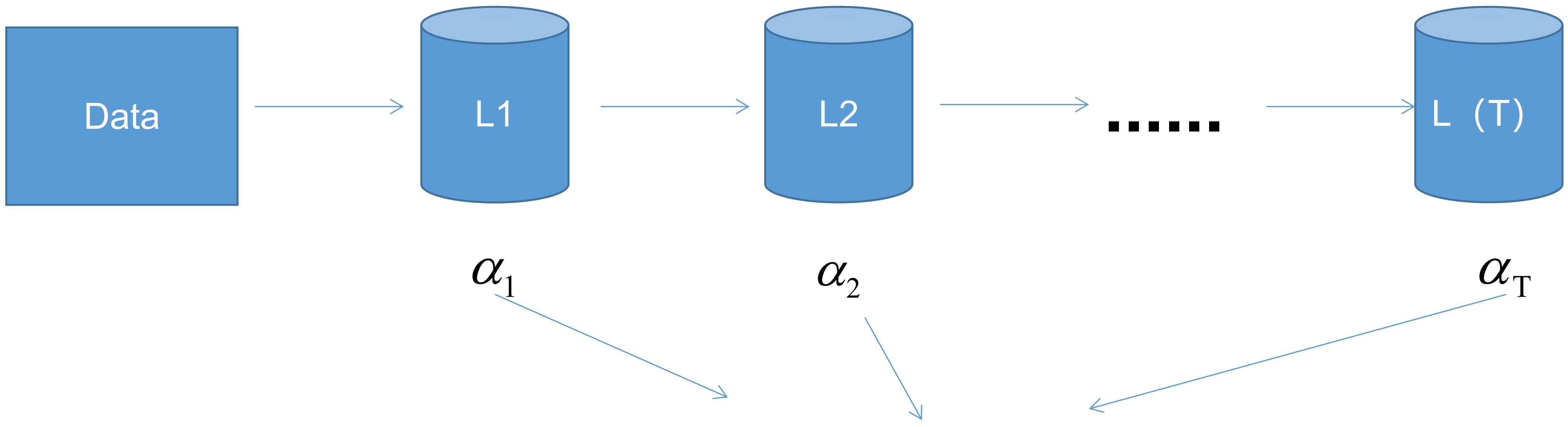


Stacking



Boosting

Boosting



$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

$$h_t(x) \in \{-1,1\}$$

AdaBoost

D	X1	X2	...	Xd	y	D1	D2	D3
(x1,y1)						0.2	0.5	0.28
(x2,y2)						0.2	0.125	0.07
...						0.2	0.125	0.07
...						0.2	0.125	0.08
(xm,ym)						0.2	0.125	0.5

$$h_t = L(D, D_t)$$

$$\varepsilon_t = E_{x\sim D_t}\left[h_t(X) \neq y\right]$$

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$$

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} \exp(-\alpha_t), & h_t(x_i) = y_i \\ \exp(\alpha_t), & h_t(x_i) \neq y_i \end{cases}$$

权重 / 分布 (Optional)

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} \exp(-\alpha_t), & h_t(x_i) = y_i \\ \exp(\alpha_t), & h_t(x_i) \neq y_i \end{cases}$$
$$= \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

$$D_{T+1}(i)$$
$$= \frac{1}{m}\frac{\exp(-\alpha_1 y_i h_1(x_i))}{Z_1} \cdots \frac{\exp(-\alpha_T y_i h_T(x_i))}{Z_T}$$
$$= \frac{\exp\left(-y_i \sum_{t=1}^{T} \alpha_t h_t(x_i)\right)}{m\prod_{t=1}^{T} Z_t}$$

$$Z_t = \sum_{i=1}^{m} D_t(i)\exp(-\alpha_t y_i h_t(x_i))$$

损失函数 (Optional)

$$H(x_i) \neq y_i$$

$$sign\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right) \neq y_i$$

$$y_i \times sign\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right) \leq 0$$

$$y_i\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right) \leq 0$$

$$\exp\left(-y_i\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right)\right) \geq 1$$

$$I\{H(x_i) \neq y_i\} \leq \exp\left(-y_i\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right)\right)$$

$$\frac{1}{m}\sum_{i=1}^{m} I\{H(x_i) \neq y_i\}$$

$$\leq \frac{1}{m}\sum_{i=1}^{m} \exp\left(-y_i\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right)\right)$$

$$= \sum_{i=1}^{m}\left(\left(\prod_{t=1}^{T} Z_t\right) D_{T+1}(i)\right)$$

$$= \prod_{t=1}^{T} Z_t$$

求解 alpha (Optional)

$$\min \prod_{t=1}^{T} Z_t \xrightarrow{\text{greedy}} \min Z_t$$

$$\boxed{\begin{aligned} Z_t &= \sum_{i=1}^{m} D_t(i)\exp\left(-\alpha_t y_i h_t(x_i)\right) \\ &= E_{x\sim D_t}\left[\exp\left(-\alpha_t y_i h_t(x)\right)\right] \\ &= \exp(-\alpha_t)P\left(y_i = h_t(x)\right) + \exp(\alpha_t)P\left(y_i \neq h_t(x)\right) \\ &= \exp(-\alpha_t)(1-\varepsilon_t) + \exp(\alpha_t)\varepsilon_t \end{aligned}}$$

$$\frac{\partial Z_t}{\partial \alpha_t} = 0$$

$$\boxed{\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)}$$

Error Bound (Optional)

$$
\begin{aligned}
&Z_t \\
&= \exp(-\alpha_t)(1-\varepsilon_t) + \exp(\alpha_t)\varepsilon_t \\
&= \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}(1-\varepsilon_t) + \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}\varepsilon_t \\
&= 2\sqrt{\varepsilon_t(1-\varepsilon_t)}
\end{aligned}
$$

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$$

$$
\begin{aligned}
r_t &= \sum_{i=1}^{m} D_t(i) y_i h_t(x_i) \\
&= E[y_i h_t(x)] \\
&= 1 \bullet P(y_i = h_t(x)) + (-1) \bullet P(y_i \neq h_t(x)) \\
&= 1 - 2\varepsilon_t \\
&\rightarrow \varepsilon_t = \frac{1-r_t}{2}
\end{aligned}
$$

$$\frac{1}{m} I\{H(x_i) \neq y_i\} \leq \prod_{t=1}^{T} Z_t = \prod_{t=1}^{T} \sqrt{1-r_t^2}$$

AdaBoost 的优缺点

优点	缺点
容易实现 不容易出现过拟合 参数不多 理论基础好	alpha 是局部最优 对噪音敏感 **base model 不能为线性模型（logistic regression）**