上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

学生创新中心 Student Innovation Center

集成学习 Ensemble Learning

学生创新中心：肖雄子彦

SJTU

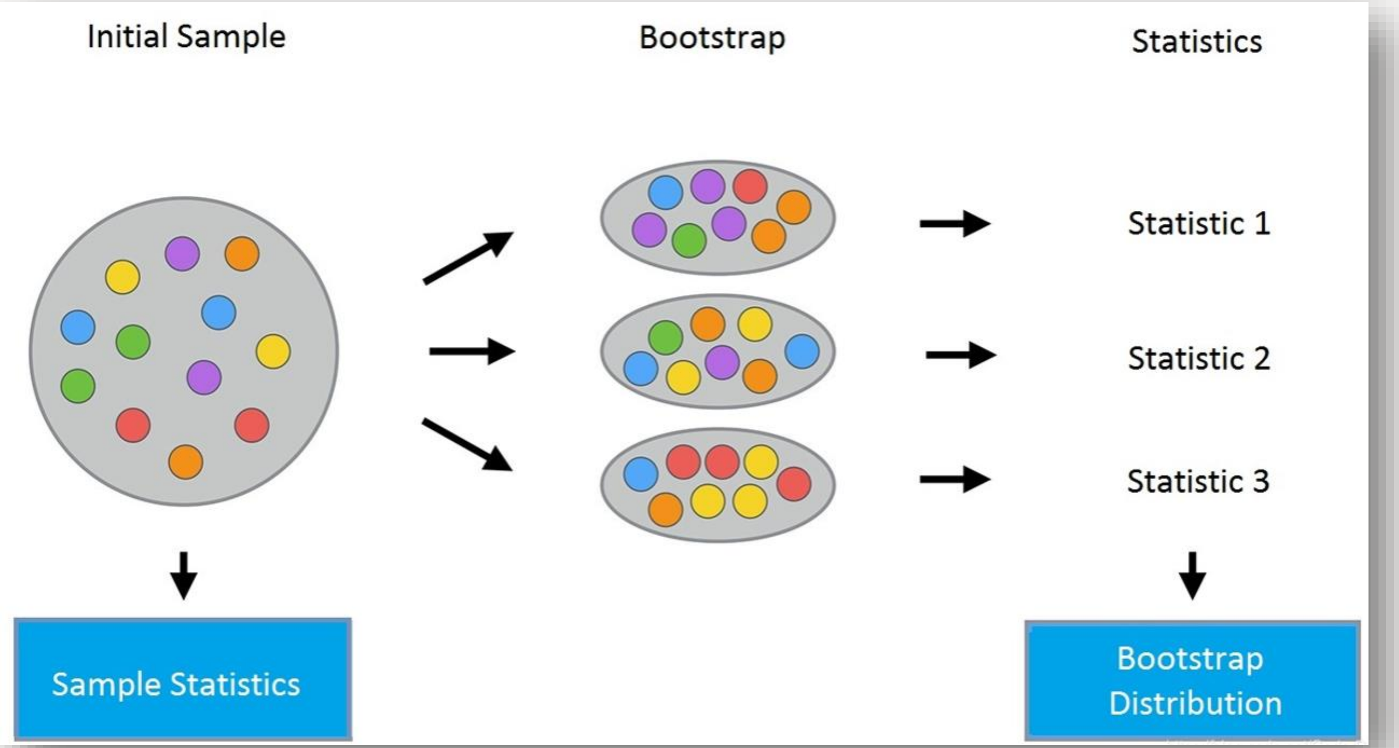
目录 / CONTENTS

01

Bagging

Bagging

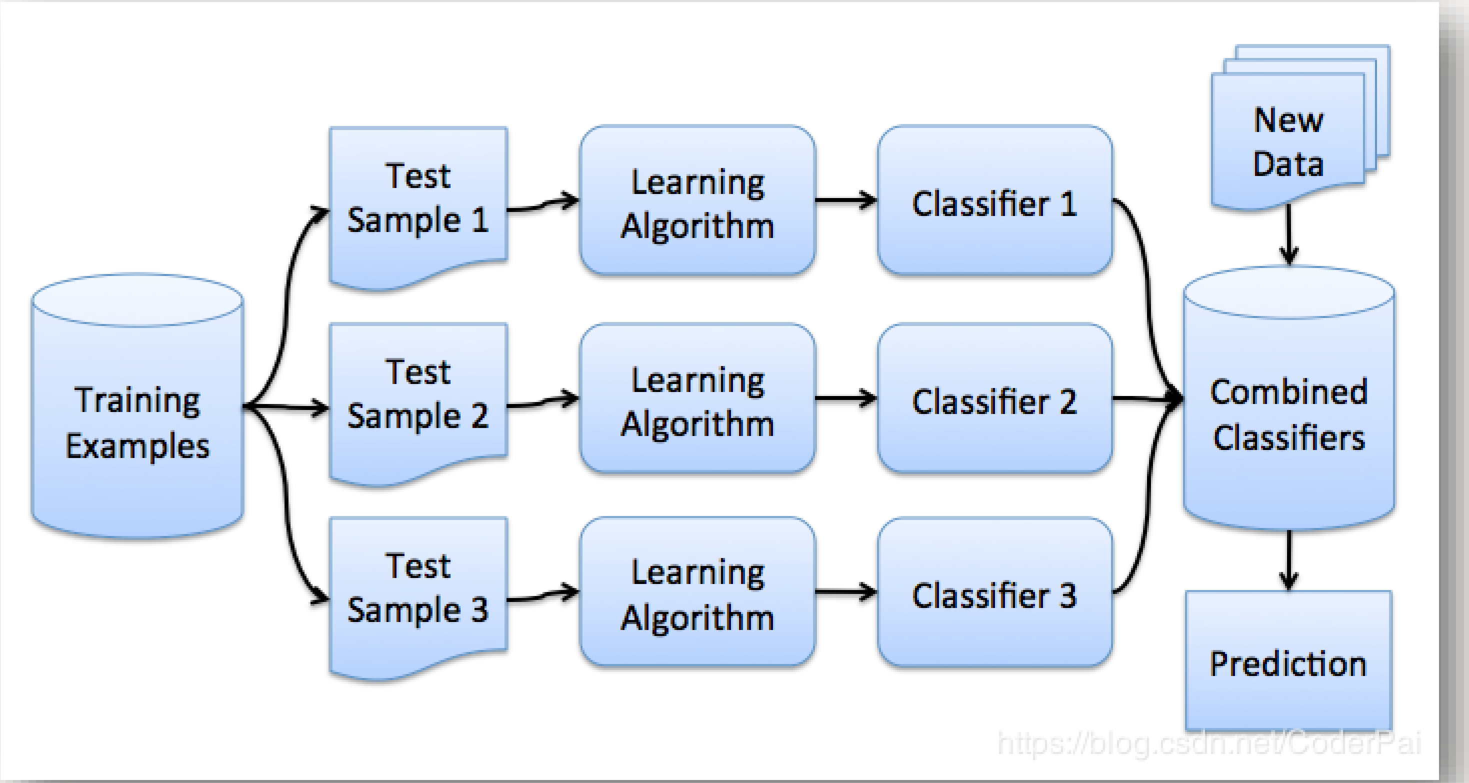
Bagging 也称为 Bootstrap aggregation，是最早，也是最基本的集成技术之一。
陪审团：一个专家的决策是有限的，那就多听几个人的意见。



①随机均匀的从训练集中选择N个元素，作为训练样本。
②重复有放回，抽取多次，得到多个训练样本。
③用不同的样本分别训练一个不同分类器。
④把结果平均起来。

Bagging

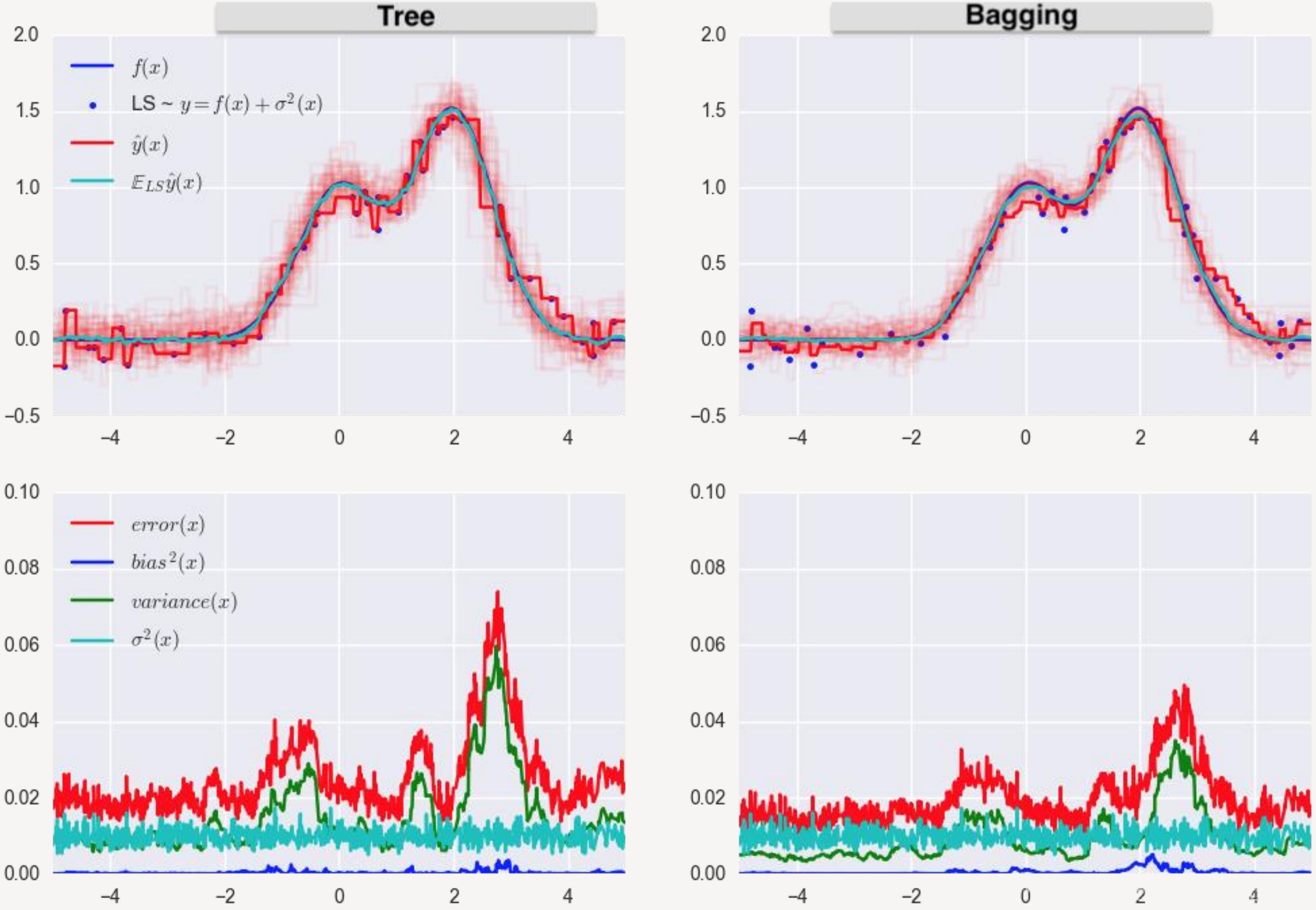
对于每一个抽样的Test Sample，都去训练一个分类器。
最终的分类器是平均所有这些单独分类器的输出。在分类算法中，称之为投票。



Bagging特点

- 减少Variance：当我们在不同样本数据上进行训练时，考虑各个分类器的结果，增强泛化能力。
- 降低过拟合：Bagging 用于容易过拟合的模型，不要在个别数据上学的“太好”。

Bagging



决策树和Bagging
训练的比较

Bagging可以解决
Variance比较大的问题

目录 / CONTENTS

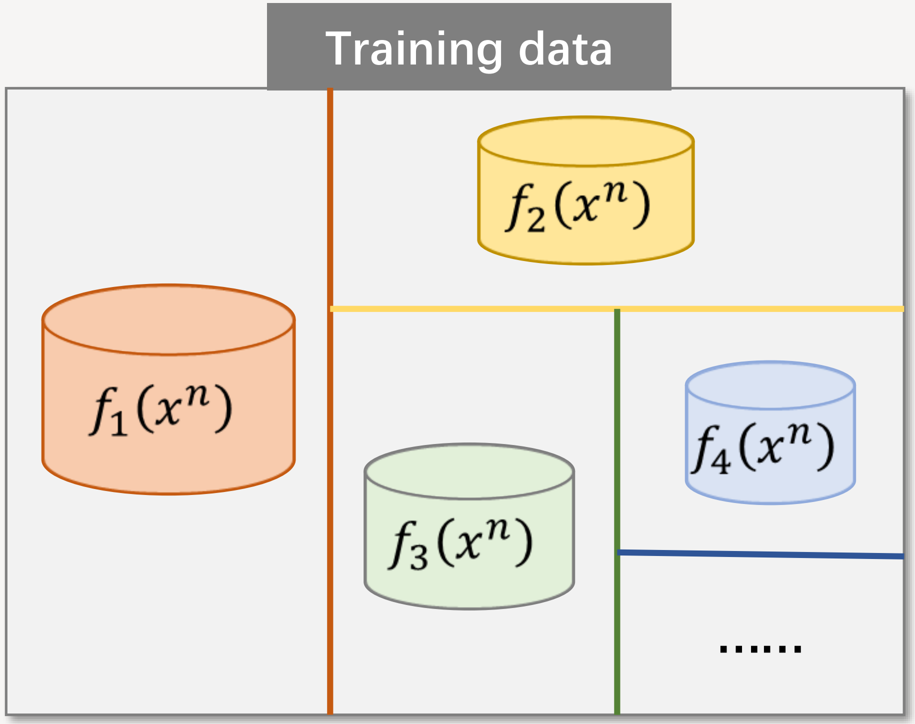
02

Boosting

AdaBoost

提高模型性能的好手——AdaBoost

- 把很多弱分类器的组合起来能变成一个强分类器（让Error从很高降到很低）
- 每一个生成的新的弱分类器，需要和上一个弱分类器性能上有所互补
- 在Training data上进行有序的训练，后面的分类器训练建立在前面分类器的基础上



每一次新的 $f_{n+1}(x^n)$ 所用的训练集和前一次 $f_n(x^n)$ 所用的训练集相同又不同

Same	Difference
无需Re-sample 不必像Bagging一样，制造不同的training data。(重新随机抽样)	需要Re-weight 当进行下一次训练时，需要改变上一次training data的权重。

AdaBoost

For Example

分对的可以不用太关心，分错的需要加强训练。

Why?

这个权重要怎么体现出来？

$(x^1, \hat{y}^1)$ $w_1 = 1$ → 0.7

$(x^2, \hat{y}^2)$ $w_2 = 1$ → 1.8

$(x^3, \hat{y}^3)$ $w_3 = 1$ → 0.3

$$L(\theta) = \sum_n l(f(x^n), \hat{y}^n) \Rightarrow L(\theta) = \sum_n w_n l(f(x^n), \hat{y}^n)$$

Data → $f_1(x^n)$ → “New” Data → $f_2(x^n)$

fail

AdaBoost算法思想：

- 模型训练时：提高那些被前一轮弱分类器错误分类的样本的权值，降低那些被正确分类的样本的权值。
- 集成所有模型时：采用加权表决法——分类误差率小的弱分类器权值高，在表决中起较大的作用；分类误差率大的弱分类器的权值小，在表决中起较小的作用。

AdaBoost

AdaBoost算法思想：

- 模型训练时：提高那些被前一轮弱分类器错误分类的样本的权值，降低那些被正确分类的样本的权值。
- 集成所有模型时：采用加权表决法，分类误差率小的弱分类器权值高，在表决中起较大的作用；分类误差率大的弱分类器的权值小，在表决中起较小的作用。

假设有第一个弱分类器的错误率ε_1 $\varepsilon_1 < 0.5$

$$\varepsilon_1 = \frac{\Sigma_n W_n^1 \boxed{\delta(f_1(x^n) \neq \hat{y}^n)}}{Z_1} \qquad Z_1 = \sum_n W_n^1$$

$$W_n^1 \downarrow W_n^2 \Rightarrow \boxed{\frac{\Sigma_n W_n^2 \delta(f_1(x^n) \neq \hat{y}^n)}{Z_2} = 0.5}$$

在新权重的data set下训练 $f_2(x^n)$

	$(x^1, \hat{y}^1)$	$(x^2, \hat{y}^2)$ ✖	$(x^3, \hat{y}^3)$
	$w_1 = 1$	$w_2 = 1$	$w_3 = 1$
$f_1(x^n)$, $\varepsilon_1 = 0.33$	↓ 减小	↓ 增大	↓ 减小
	$w_1 = 1/\sqrt{2}$	$w_2 = \sqrt{2}$	$w_3 = 1/\sqrt{2}$
			0.5

$f_2(x^n)$ $\varepsilon_1 < 0.5$

……

AdaBoost

- 如果 $f_1(x^n) \neq \hat{y}^n$ ：增加权重
- 如果 $f_1(x^n) = \hat{y}^n$ ：减少权重

$W_n^1 \xrightarrow{\times D^1} W_n^2$

$W_n^1 \xrightarrow{\div D^1} W_n^2$

$$\varepsilon_1 = \frac{\Sigma_n W_n^1 \delta(f_1(x^n) \neq \hat{y}^n)}{Z_1}$$

D^1 怎么求得？

$$0.5 = \frac{W_n^1 D^1}{\sum_{f_1(x^n)\neq\hat{y}^n} W_n^1 D^1 + \sum_{f_1(x^n)=\hat{y}^n} W_n^1 / D^1}$$

$$\frac{\boxed{\Sigma_n W_n^2 \delta(f_1(x^n) \neq \hat{y}^n)}}{\boxed{Z_2}} = 0.5$$

$$\sum_n W_n^2$$

$$\sum_{f_1(x^n)\neq\hat{y}^n} W_n^1 D^1 = \sum_{f_1(x^n)=\hat{y}^n} W_n^1 / D^1 \Rightarrow D^1 \underbrace{\sum_{f_1(x^n)\neq\hat{y}^n} W_n^1}_{\varepsilon_1 Z_1} = \frac{1}{D^1} \underbrace{\sum_{f_1(x^n)=\hat{y}^n} W_n^1}_{(1-\varepsilon_1) Z_1} \Rightarrow D^1 = \sqrt{(1-\varepsilon_1)/\varepsilon_1}$$

AdaBoost

$$\{(x^1,\hat{y}^1),(x^2,\hat{y}^2),(x^3,\hat{y}^3)\ \dots\dots\}$$
$$W_1^1 \qquad W_2^1 \qquad W_3^1$$

$$\varepsilon_1=\frac{\Sigma_n W_n^1\delta(f_1(x^n)\neq\hat{y}^n)}{Z_1}$$

- 如果 $f_1(x^n)\neq\hat{y}^n$ ：增加权重
- 如果 $f_1(x^n)=\hat{y}^n$ ：减少权重

$$D^1=\sqrt{(1-\varepsilon_1)/\varepsilon_1}$$

$$W_n^1 \xrightarrow{\times D^1} W_n^2 \qquad W_n^1 \xrightarrow{\div D^1} W_n^2$$

$$\alpha^1=ln\sqrt{(1-\varepsilon_1)/\varepsilon_1}$$

$$W_n^2=W_n^1\times D^1=W_n^1\times\exp(\alpha^1) \qquad W_n^2=W_n^1\div D^1=W_n^1\times\exp(-\alpha^1)$$

$$W_n^2=W_n^1\times\exp(\boxed{-\hat{y}^n f_1(x^n)}\alpha^1)$$

$\hat{y}=\pm 1$

$f_2(x^n)\quad f_3(x^n)\quad \dots\dots\quad f_t(x^n)$

$$H(x)=sign(\sum_{i=1}^{t}f_i(x^n)) \qquad H(x)=sign(\sum_{i=1}^{t}\alpha^i f_i(x^n))$$

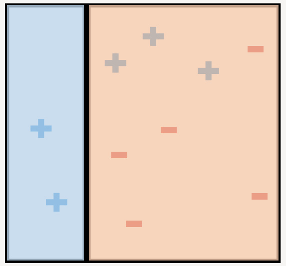
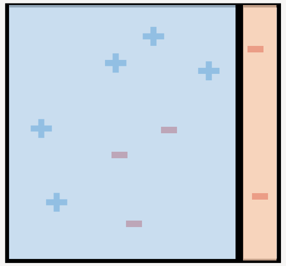
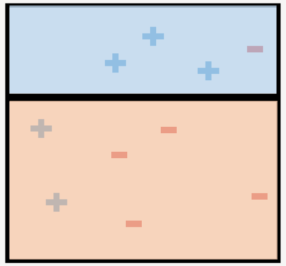
$$\begin{cases}\alpha^1=ln\sqrt{(1-\varepsilon_1)/\varepsilon_1}\\ \varepsilon_1=0.2 \qquad \varepsilon_1=0.4\end{cases}$$

$$\alpha^1=0.69 \qquad \alpha^1=0.2$$

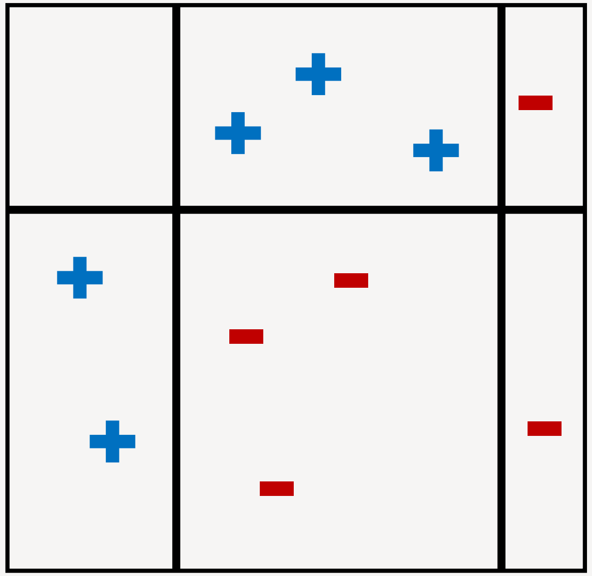
误差小的模型
话语权更大

AdaBoost

For Example

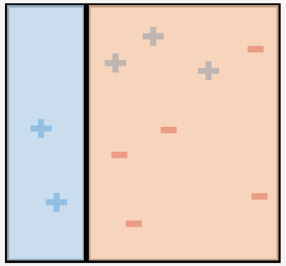
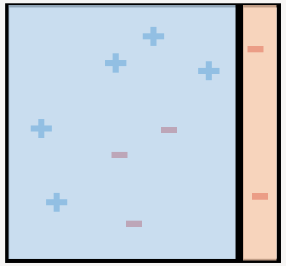
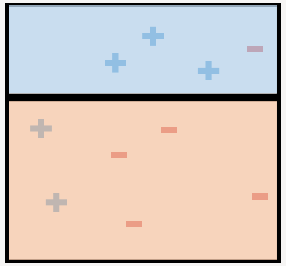
sign (0.42  + 0.66  + 0.95 )

$$H(x) = sign(\sum_{i=1}^{t} \alpha^{i} f_i(x^n))$$

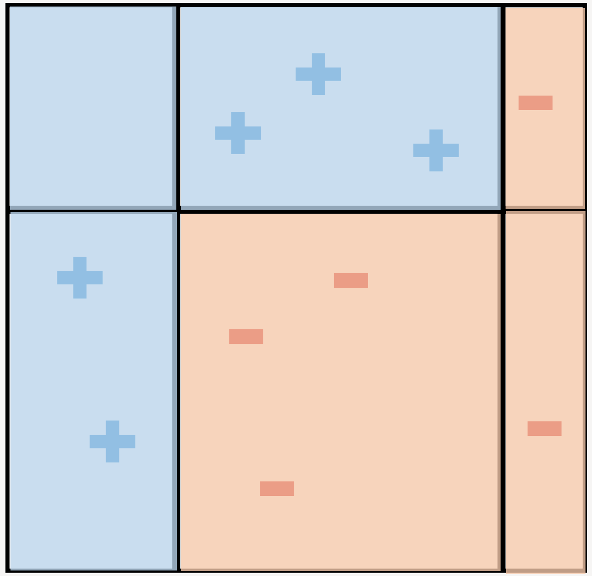


AdaBoost

For Example

$sign\ ($ 0.42  + 0.66  + 0.95  $)$

$$H(x) = sign(\sum_{i=1}^{t} \alpha^{\mathrm{i}} f_i(x^n))$$



· Thanks ·