

机器学习与量化交易实战

第七讲



看大家写的作业，读后感如图：



①

数据的分析

eg. Class balanced?

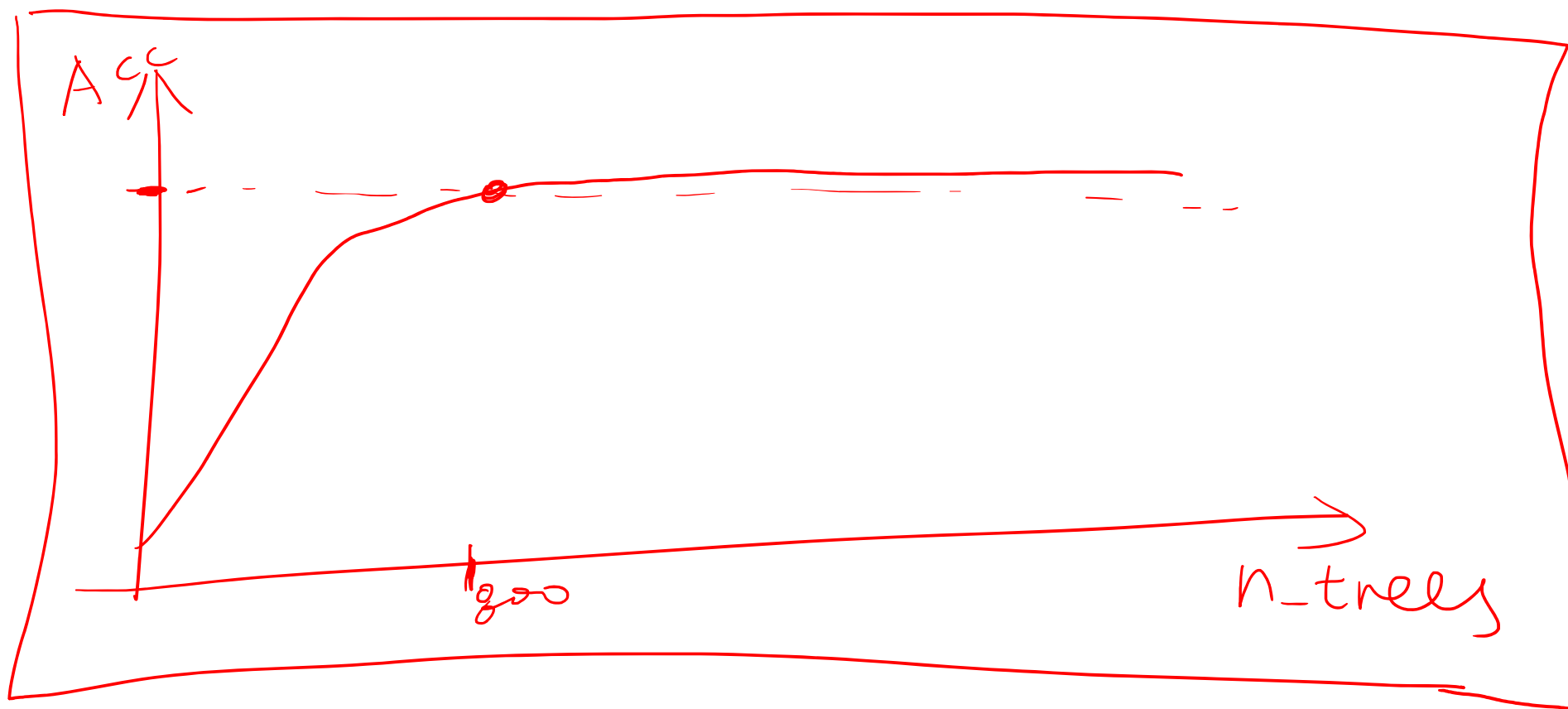
- Statistics

- Visualization 可视化.

②

→ 有目的性的选择 C/R

RF.



① 不要 17mm 4 松.

② 多 想.

③ Prap . Space .

$$R^2: 0.06x$$

$$R^2: 0.07+$$

投. FEATURES. ←

NLP

长线		短线		自
→	t_1			0

招 55

ENSEMBLE 集成学习

Ass-~~save~~ SAVING METHOD.

✓ ① Feature Extraction.

✓ ② Modeling [M_1, M_2, \dots, M_w]

✓ ③ Ensemble! ←

~~④ IS - -~~

- Stacking / blending / Voting.

EASY
IMPORTANT

- AdaBoost.

± 1

VOTING

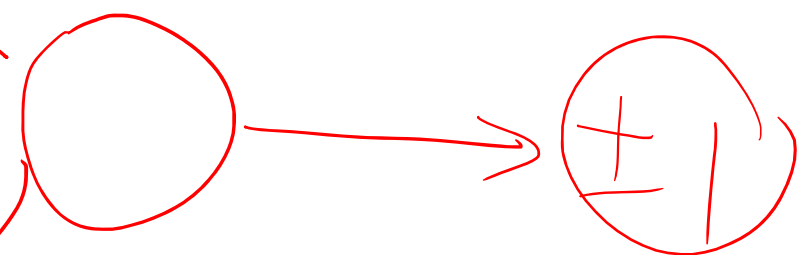
HARD VOTING

X

①

<u>SUAA</u>	h_1	70	+1
<u>RF</u>	h_2	67	+1
<u>LR</u>	h_3	83%	-1
	\vdots		
	h_{10}	73%	+1

$$\text{sign}(\sum h_i)$$



$$w_i = \frac{1}{N}$$

Blending

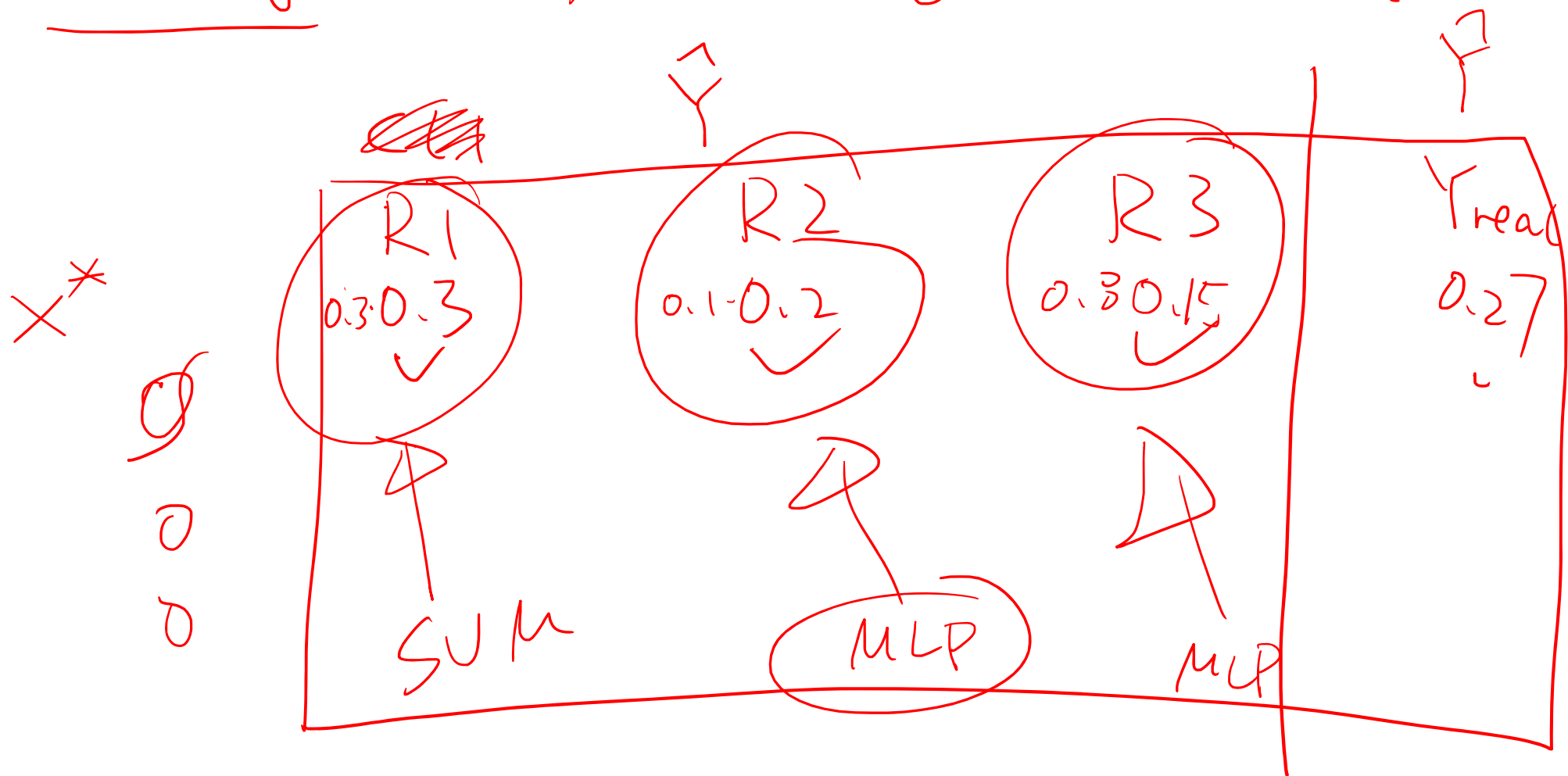
$$H: f(\alpha_1 h_1 + \alpha_2 h_2 + \dots + \alpha_n h_n)$$

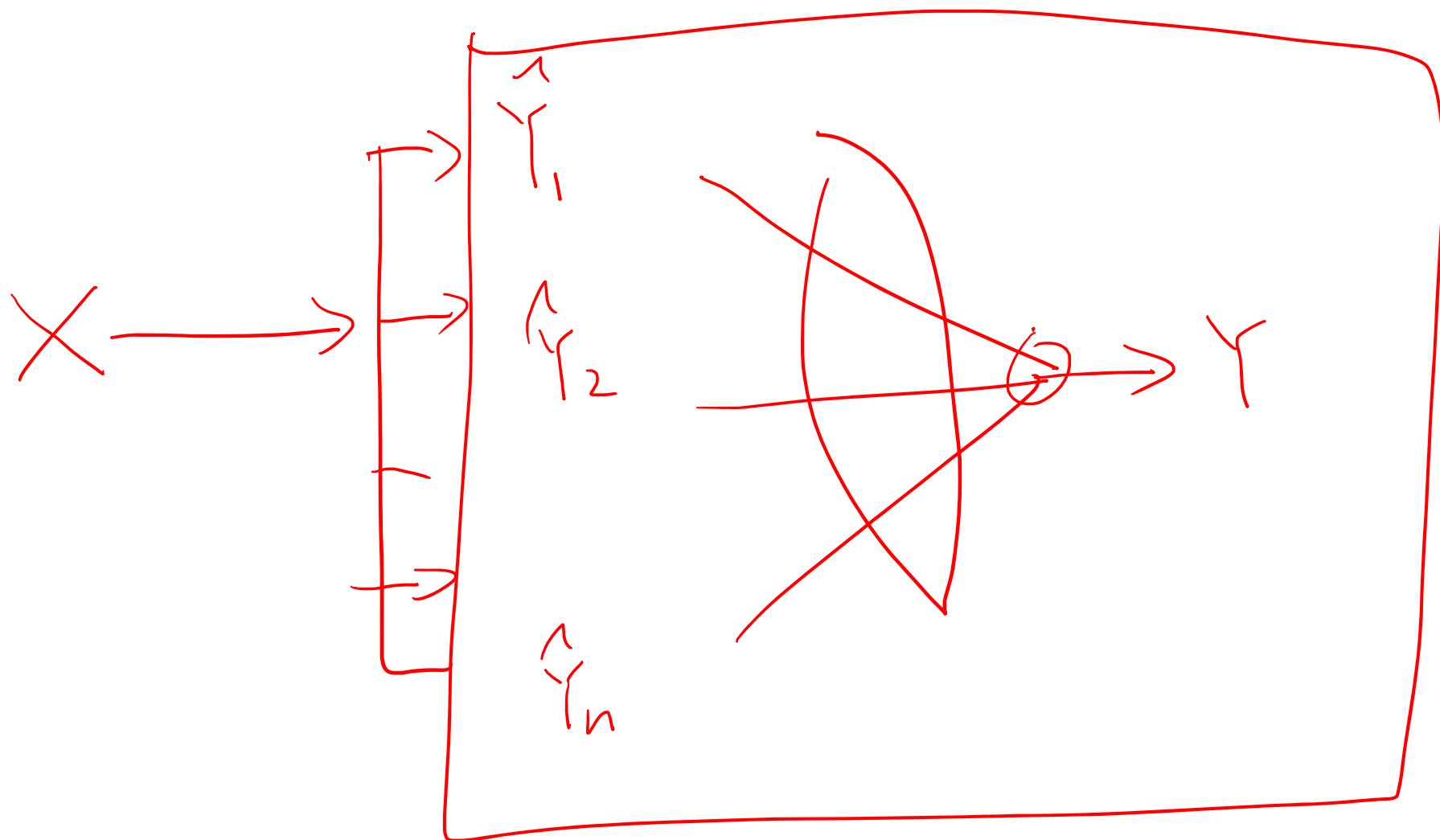
Annotations: α_1 (0.7) points to SUM; α_2 (0.4) points to RF; α_n (0.3) points to LR.

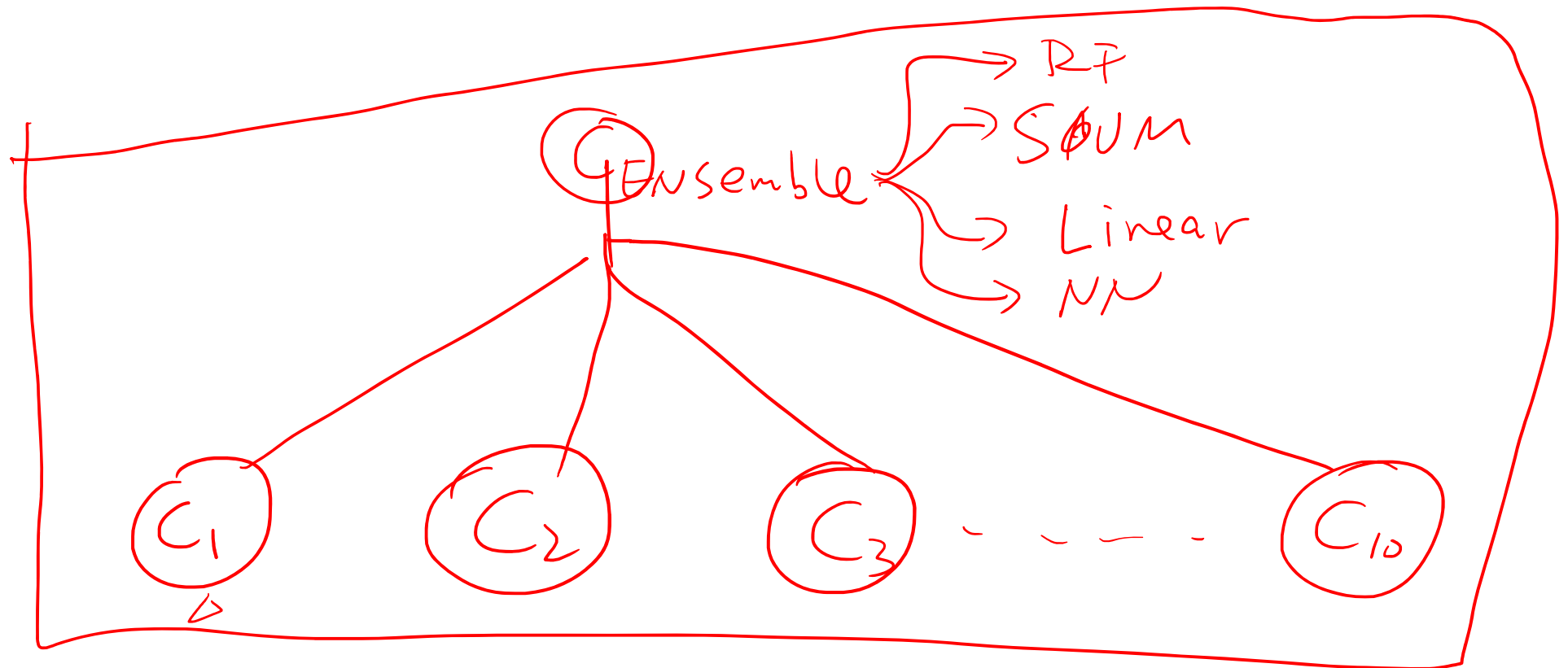
Q: How To Define weights ~~weight~~ for clfs

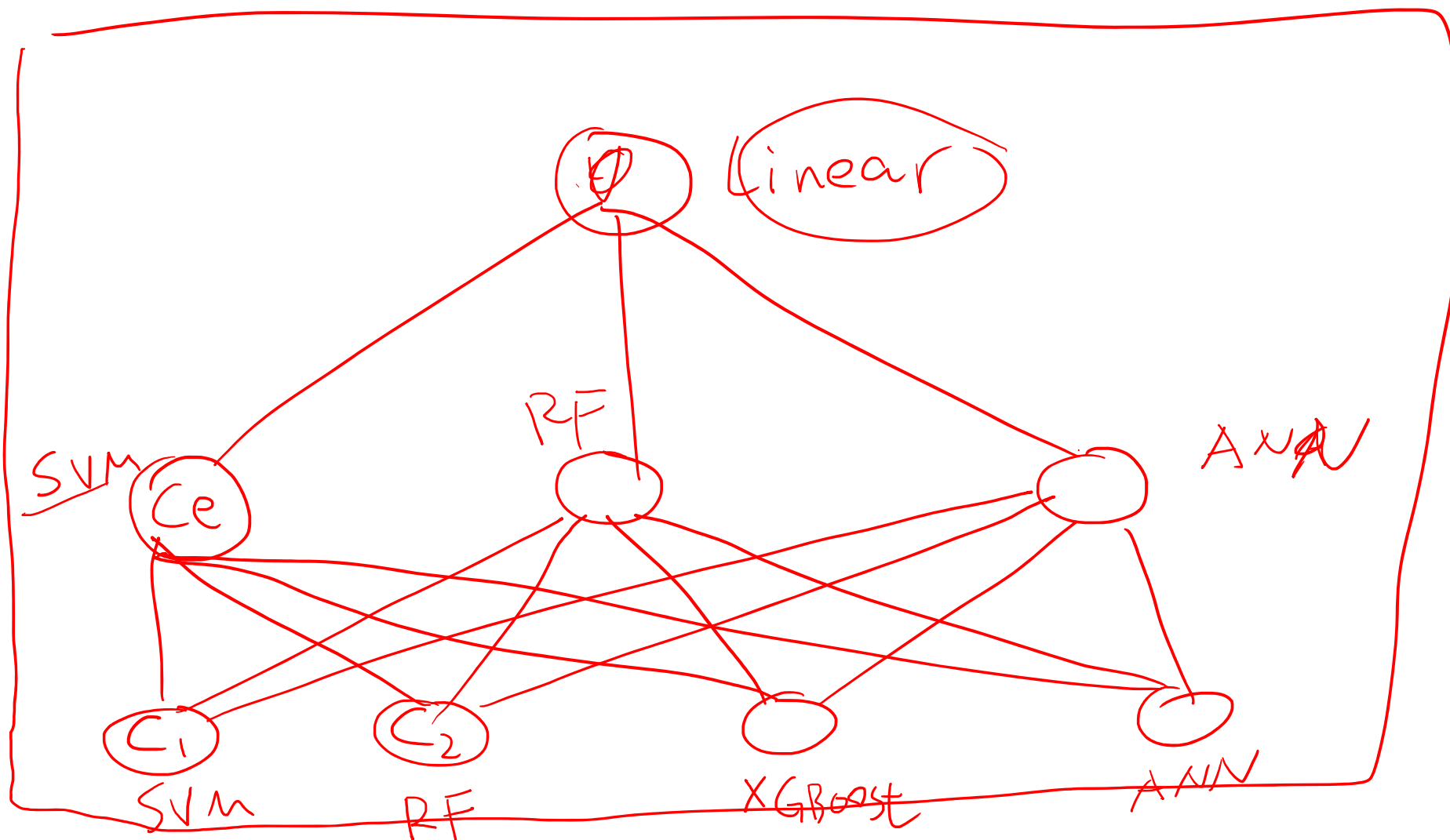
① $\{z_i\}$	Acc	w
\nearrow	70%	Acc / \bar{Acc}_i
\nearrow SUM		
LR	63%	

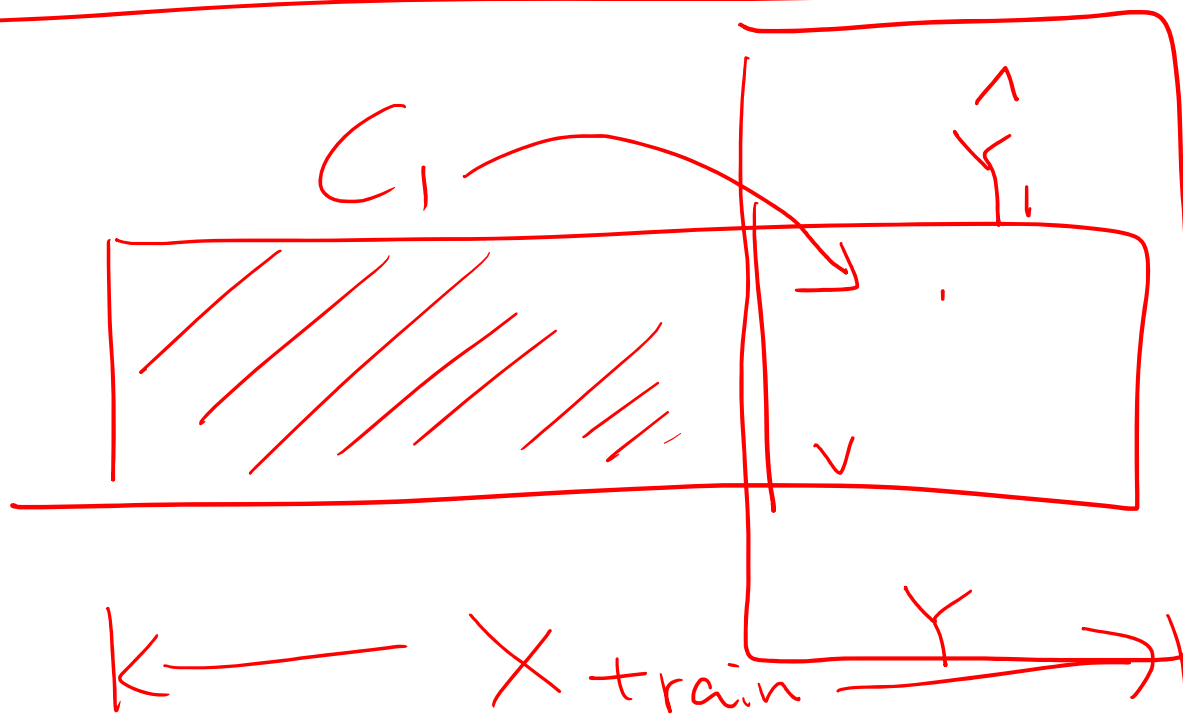
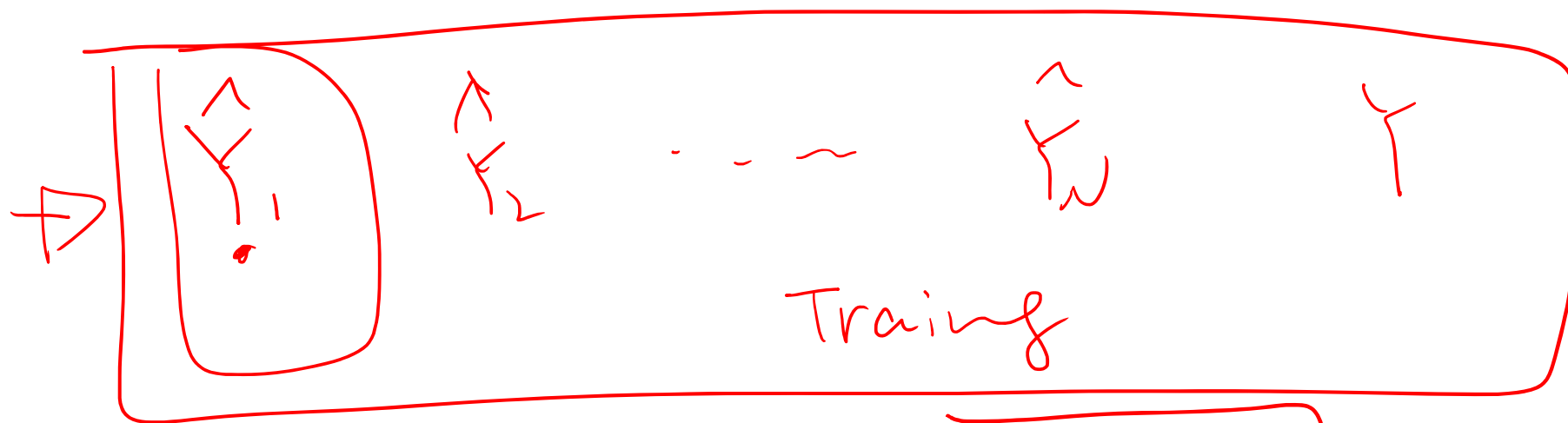
Blending: 用机器学习的方法来做。

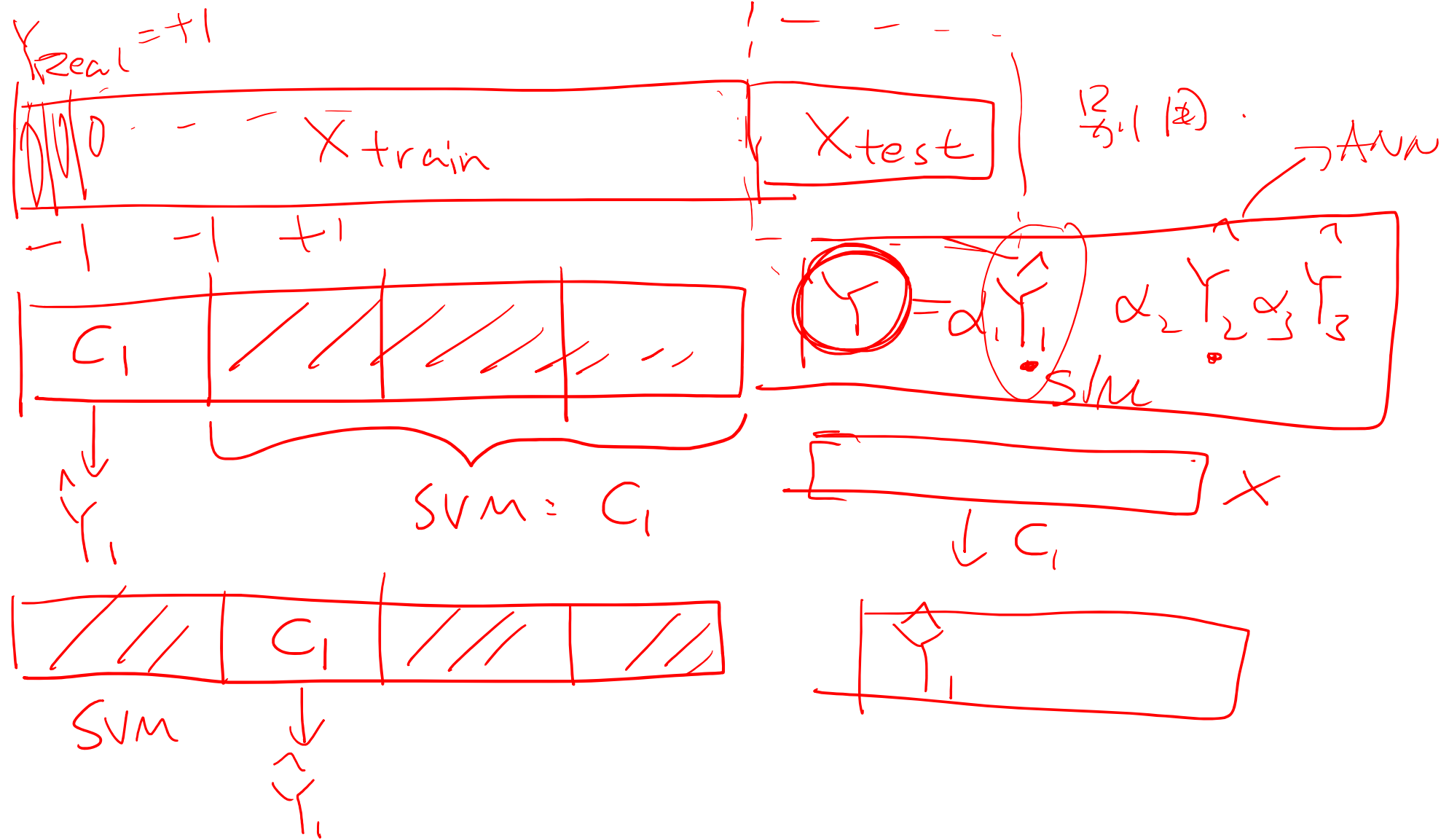






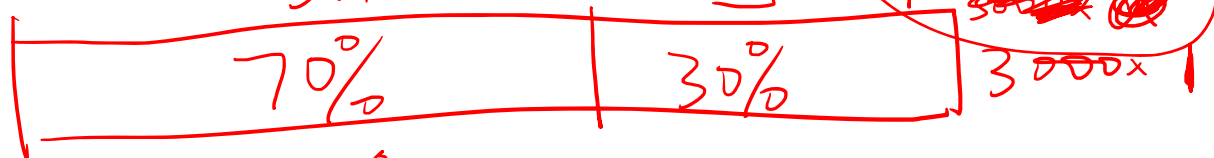




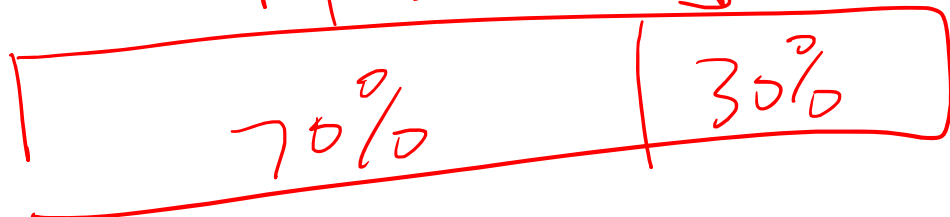


X_{train}

SVM



RF

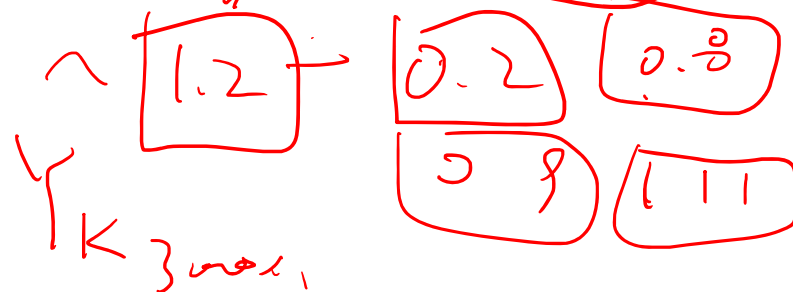
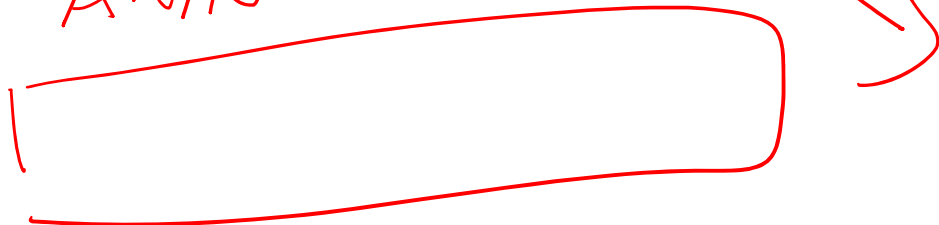


\hat{Y}_2 3000x

\hat{Y} 3000x

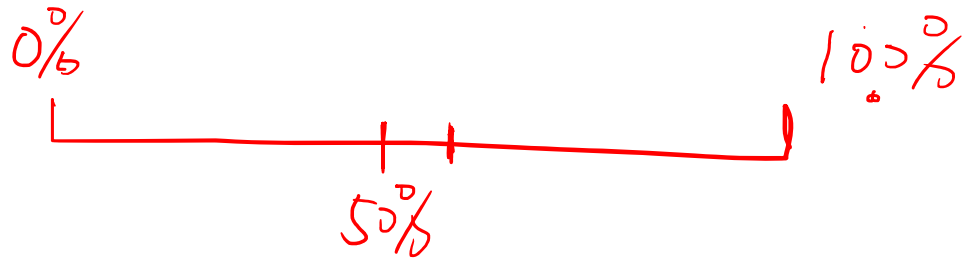
$$\hat{Y} = \alpha_1 \hat{Y}_1 + \alpha_2 \hat{Y}_2 + \dots + \alpha_n \hat{Y}_n$$

ANN



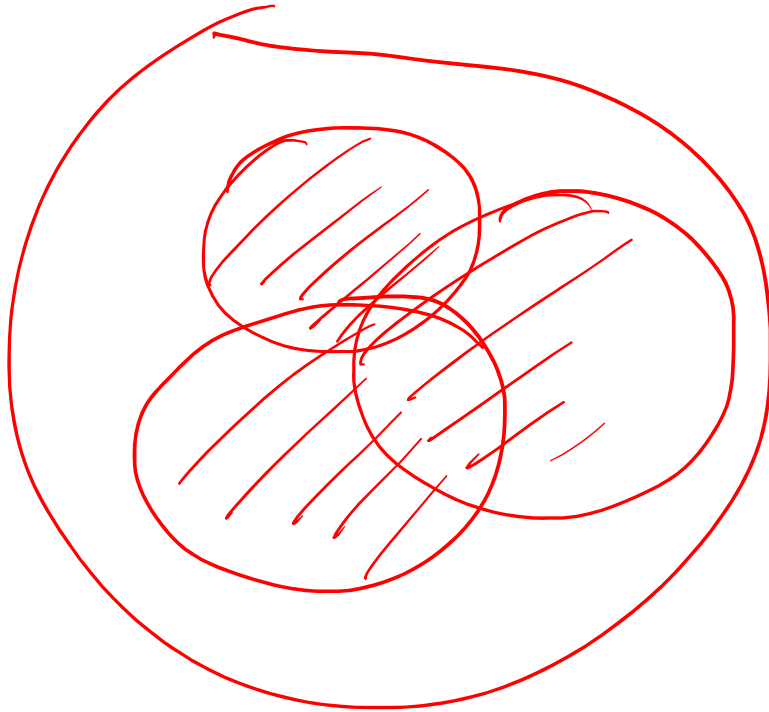
adaBoost :

$h_i \in [-1, +1]$



$$H = \text{Sign} (h_1 + h_2 + \dots + h_n)$$

A diagram illustrating the combination of weak classifiers. A large circle labeled 'H' is on the left. A line connects it to a large, irregular blob on the right. Inside this blob are three smaller circles, each with diagonal hatching. Arrows point from these circles to labels: 'h1' points to the top-left circle, 'h2' points to the top-right circle, and 'h3' points to the bottom-left circle.



① 对样本加权.

$$\text{Loss} = \frac{1}{N} (\alpha_i \hat{y}_{(i)} - y)^2$$

$$\min \text{Loss}(\theta)$$

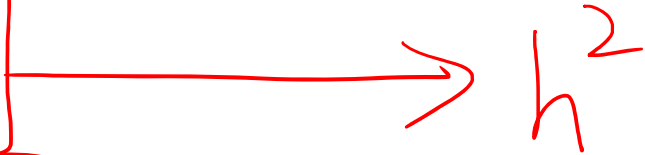
~~DATA~~

DATA



h^1

DATA



h^2

h^1 错. x_i 加权

强化

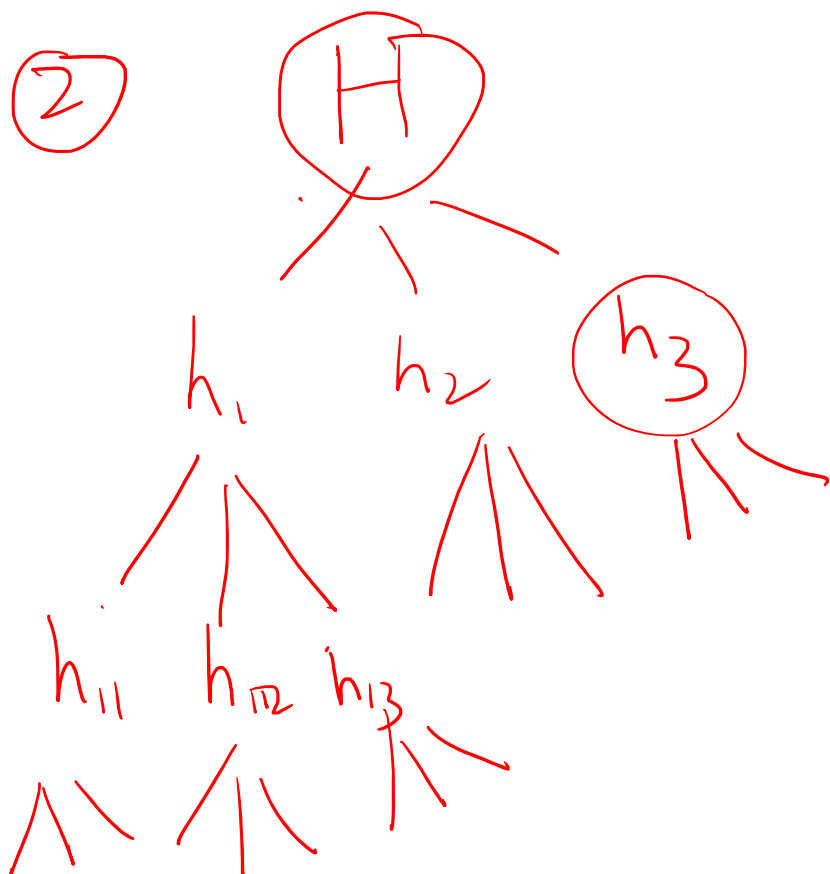
$H = (h_1, h_2, h_3)$

DATA



h^3

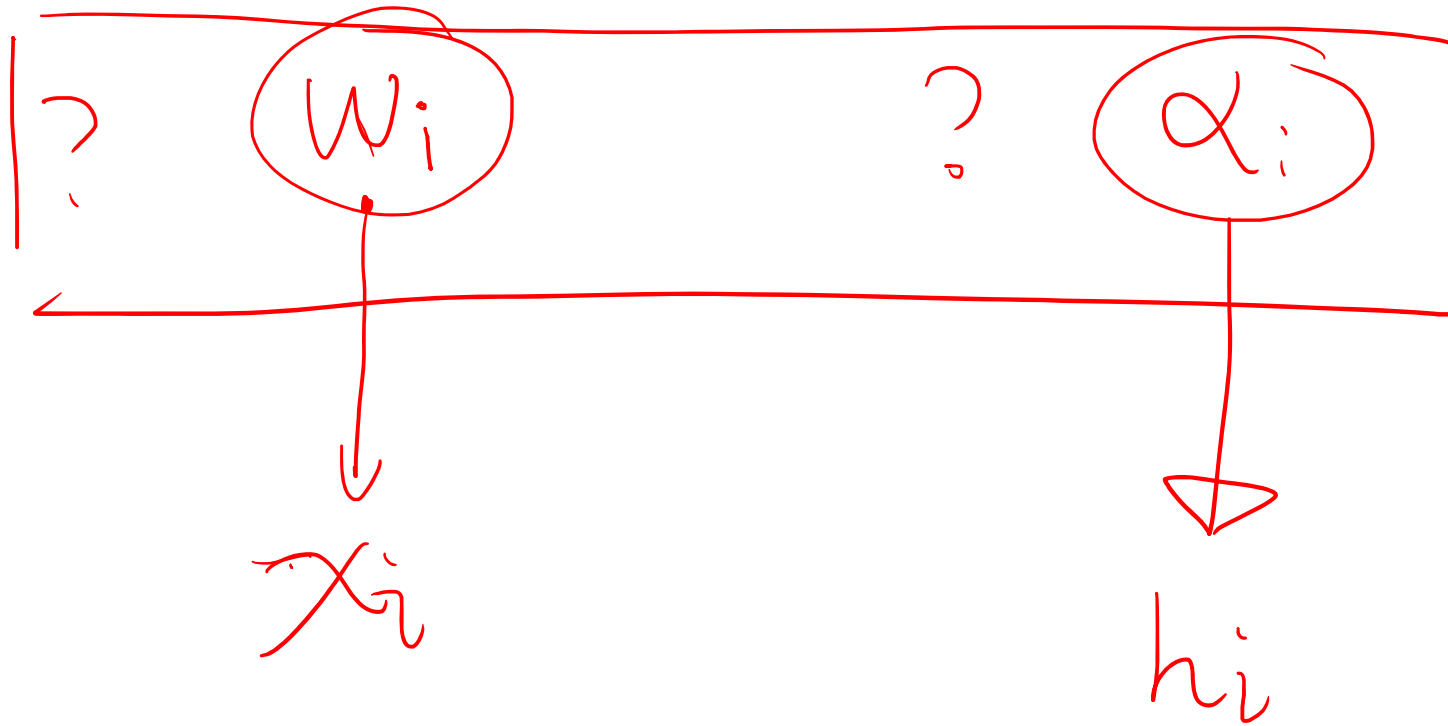
$h_1 \neq h_2$



$$E = \sum \frac{1}{N}$$

WRONG

$$H(x) = \text{sign}(\alpha_1 h_1 + \alpha_2 h_2 + \dots + \alpha_n h_n) \quad \pm 1$$



x_1, y_1
 x_2, y_2
 \vdots
 x_n, y_n

$$w_i = \frac{1}{N}$$

sklearn $e^t = \text{loss}$

② Pick BEST h^t
 minimize ϵ_o^t

?

α^t

?

③ calculate w^{t+1}

sign
 $H = (\sum \alpha_i h_i)$

± 1

$$w_i^{t+1} = \frac{w_i^t}{\sum_{j=1}^n e^{-\alpha^t h^t(x_j) \cdot y_j}} \cdot e^{-\alpha^t h^t(x) \cdot y_i}$$

Normalization



Loss

ERROR BOUND

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \epsilon^t}{\epsilon^t}$$



Loss

$$Z = \frac{\sqrt{\epsilon^t}}{\sqrt{1-\epsilon^t}} \underbrace{\sum W_i}_{\substack{\text{correct} \\ / 1-\epsilon}} + \frac{\sqrt{1-\epsilon}}{\sqrt{\epsilon}} \underbrace{\sum W_i}_{\substack{\text{wrong} \\ \checkmark \epsilon}}$$

$$W = \frac{W_i}{2}$$

$$\sqrt{\frac{1-\epsilon}{\epsilon}}$$

$$Z = 2\sqrt{\epsilon^t(1-\epsilon^t)}$$

$$w_i^{t+1} = \frac{w_i^t}{2} \cdot \frac{1}{1-\epsilon}$$

$$\sum w_i = \frac{1}{2}$$

correct

$$\sum w_i = \frac{1}{2}$$

WRONG

$$\frac{w_i^t}{2} \cdot \frac{1}{1-\epsilon}$$

X	V	
✓ $x_i: \frac{1}{2 \times 80}$	$\frac{1}{2}$	80
XX: $\frac{1}{2} \cdot \frac{1}{20}$	$\frac{1}{2}$	20
X	X	100

$$\frac{1}{2} \cdot \frac{1}{1-\epsilon} \sum w_i^t = \frac{1}{2}$$

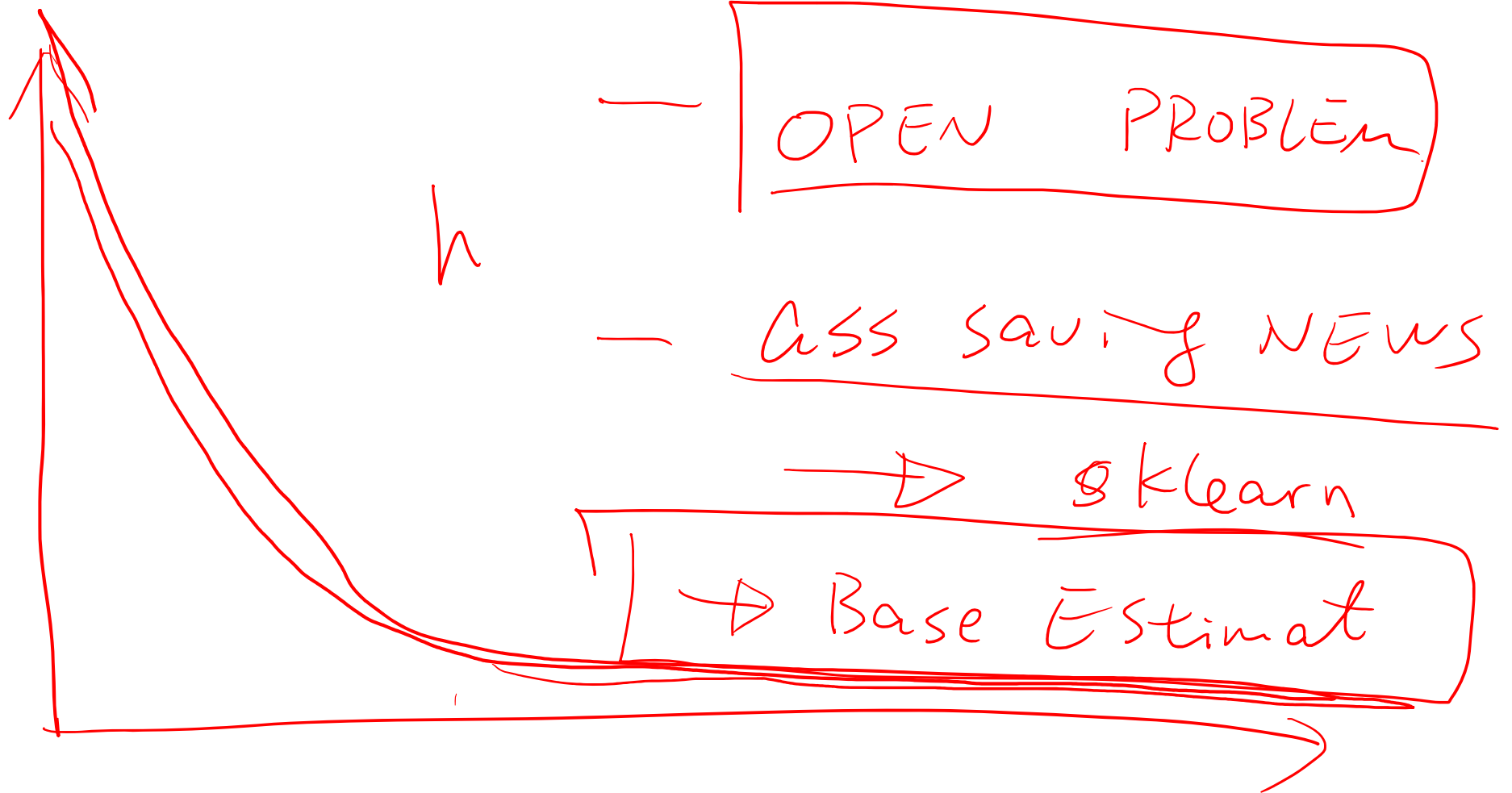
correct

→ 1-ε

~~THANK GOD STONE~~

$$\boxed{\sum_{\text{correct}} W^t} = \boxed{1 - \epsilon^t}$$

—



SKLEARN

BAGGING

ensemble

↓
adaboost

API
USERGUIDE

example

