

فرادرس

فراتر از یک کلاس درس
www.faradars.org

آموزش یادگیری ماشین (Machine Learning) (تئوری - عملی) - بخش دوم

درس سوم: یادگیری جمعی

مدرس:

فرشید شیرافکن

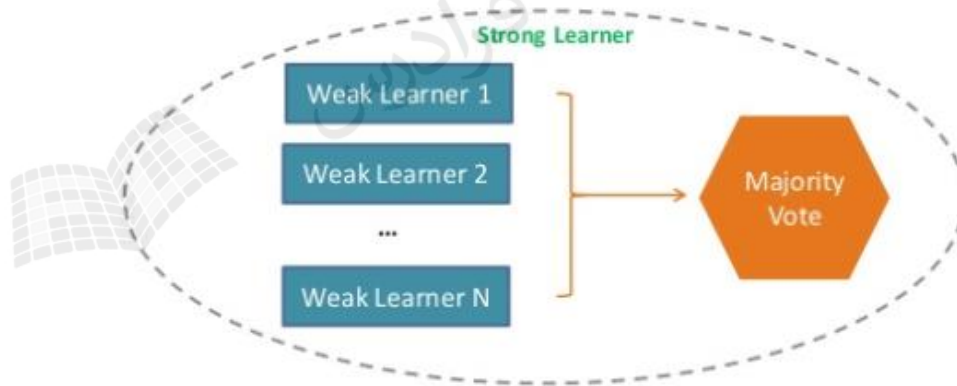
دانشجوی دکترای بیو انفورماتیک

دانشگاه تهران

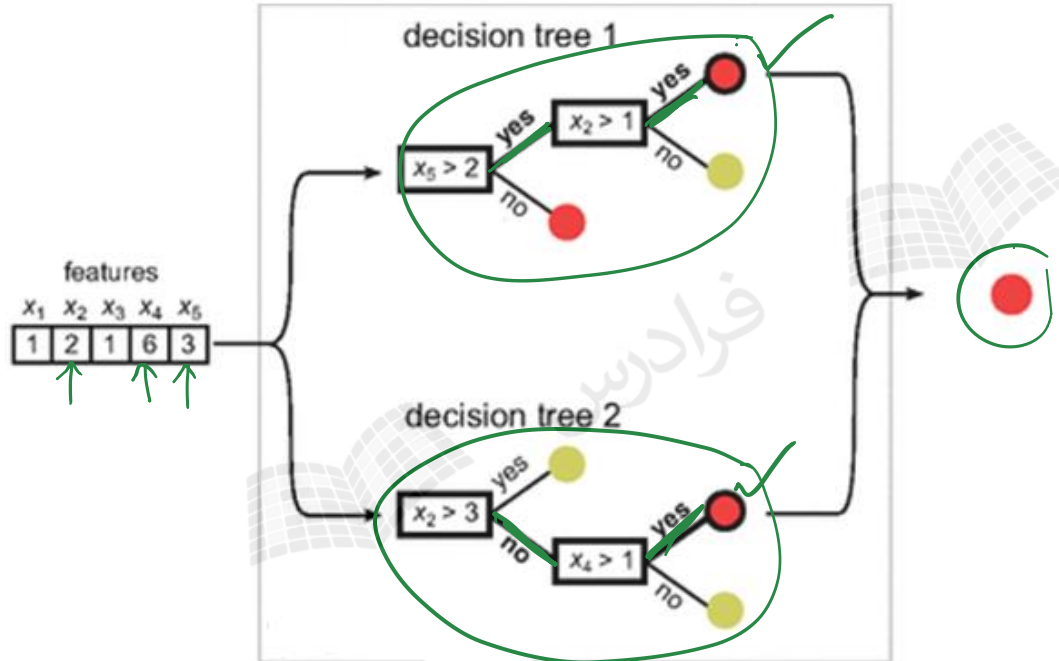
یادگیر جمعی

ترکیب چند یادگیر ضعیف برای تولید یک یادگیر قوی.

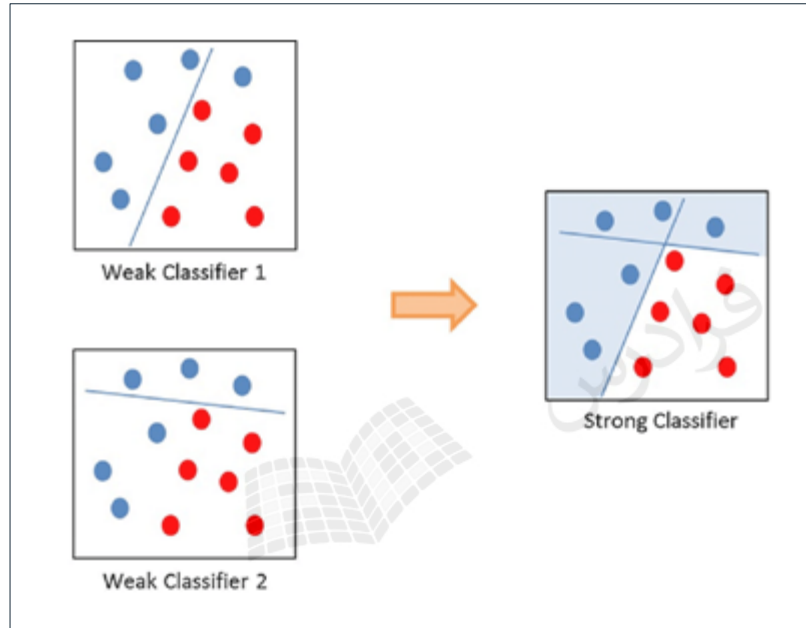
یادگیر ضعیف (weak learners): دسته بندی که فقط کمی از دسته بند تصادفی بهتر است.



مثال



مثال



دیدگاه‌ها

Bagging: Bootstrap aggregating • ✓

Boosting •



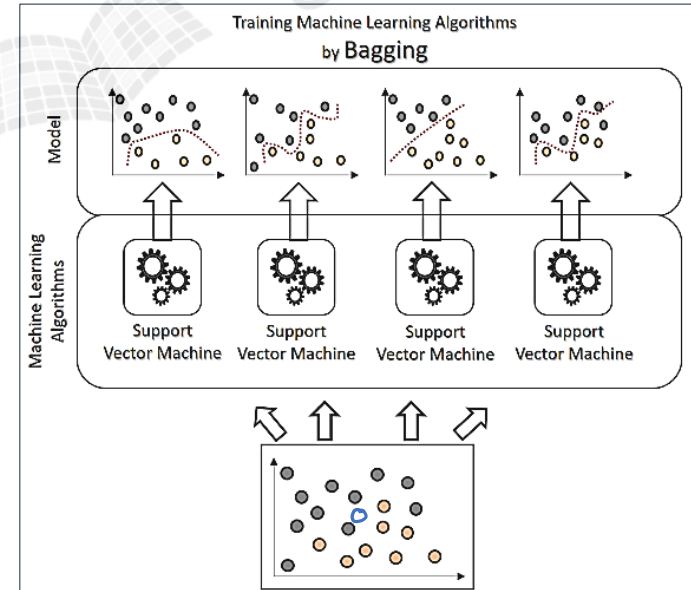
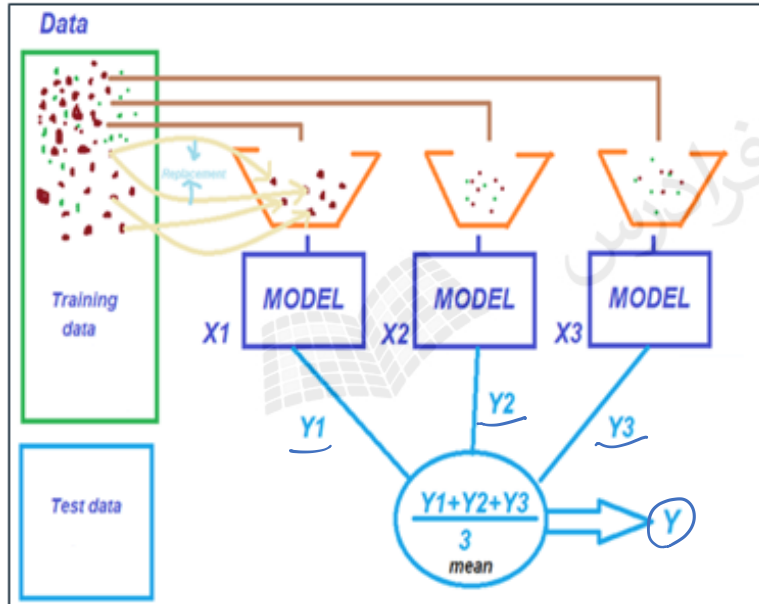
: Bootstrap ←

نمونه برداری تصادفی با جایگزینی
از نمونه های اصلی با همان اندازه.

Bagging

$$t = 3$$

Bagging is a machine learning ensemble algorithm designed to improve the accuracy of machine learning algorithms used in classification and regression.



$$t = 4$$

svm

الگوریتم Bagging

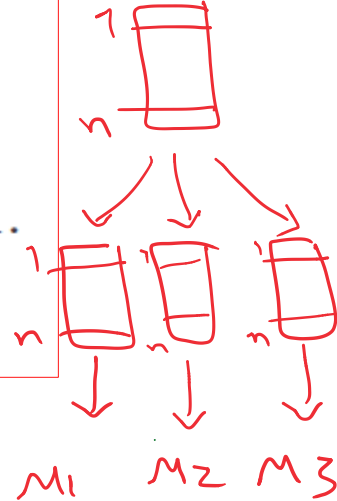
$t = 3$

model generation

Let n be the number of instances in the training data.

→ For each of t iterations:

- Sample n instances with replacement from training data.
- Apply the learning algorithm to the sample.
- Store the resulting model.



classification

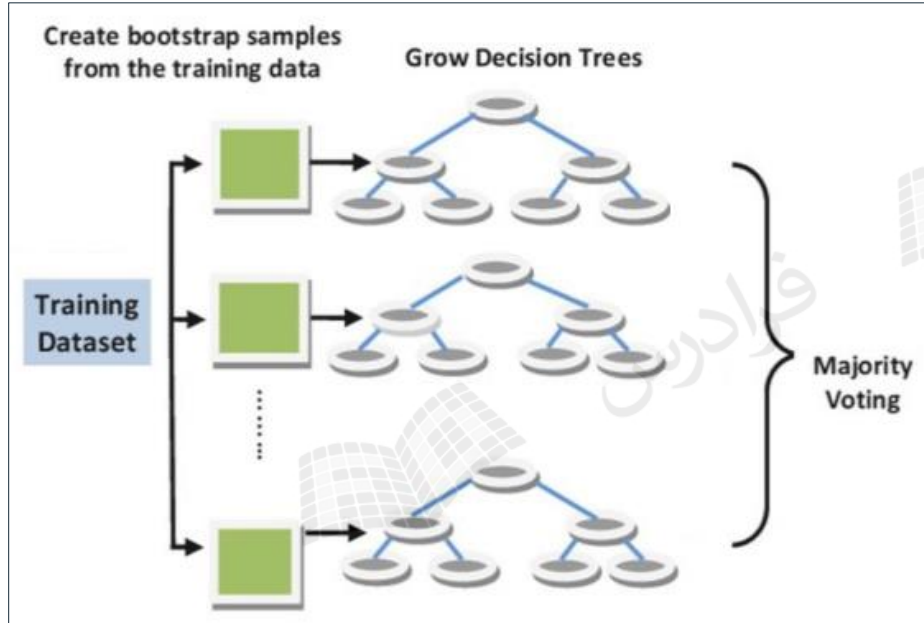
→ For each of the t models:

Predict class of instance using model.

Return class that has been predicted most often.

c_1 c_2 c_1
└──────────┘
 c_1

جنگل تصادفی (Random Forest)



- Bagging بر روی درخت‌های تصمیم.

- تصادفی بودن این روش به مرحله انتخاب ویژگی مربوط می‌شود.

- هم برای رگرسیون و هم برای دسته‌بندی قابل استفاده است.

مثال

وزن \rightarrow گزینی / نهی \rightarrow گزینی / نهی \rightarrow در حد حق / نهی

	F1	F2	F3	F4	class
1	NO	NO	NO	125	NO
2	YES	YES	YES	180	YES
3	YES	YES	NO	210	NO
4	YES	NO	YES	167	YES

بیماری قلبی

دیتای اصلی

$$d = 4$$

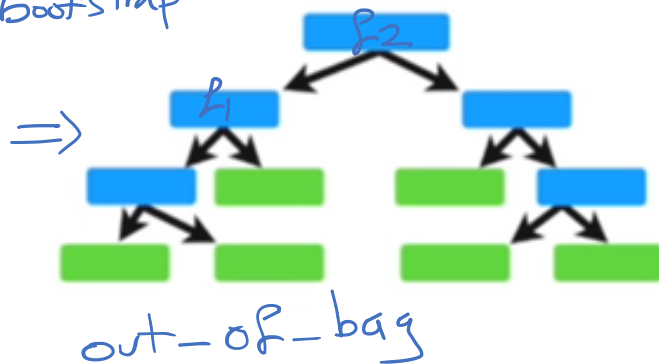
$$m = \sqrt{d} = 2$$

f_2, f_3 تصادفی

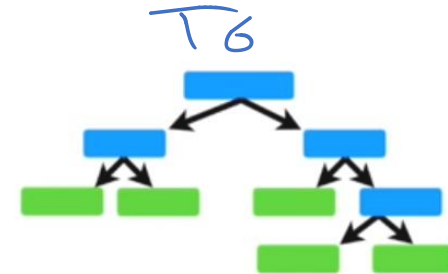
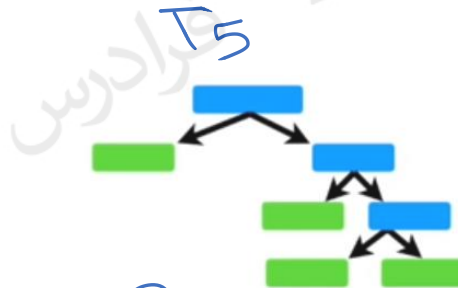
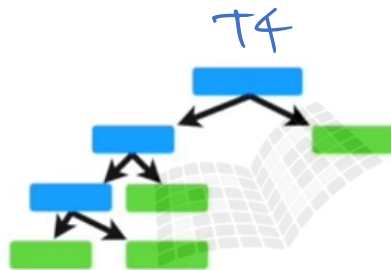
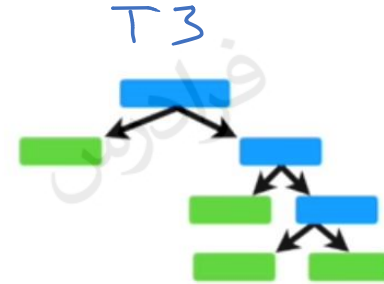
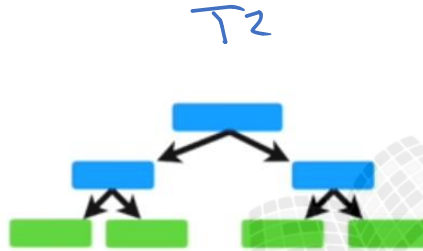
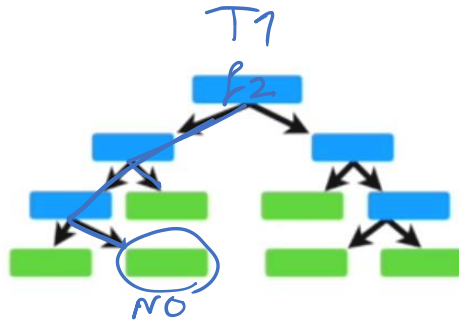
f_1, f_4

	F1	F2	F3	F4	class
2	YES	YES	YES	180	YES
1	NO	NO	NO	125	NO
4	YES	NO	YES	167	YES
4	YES	NO	YES	167	YES

دیتای bootstrap

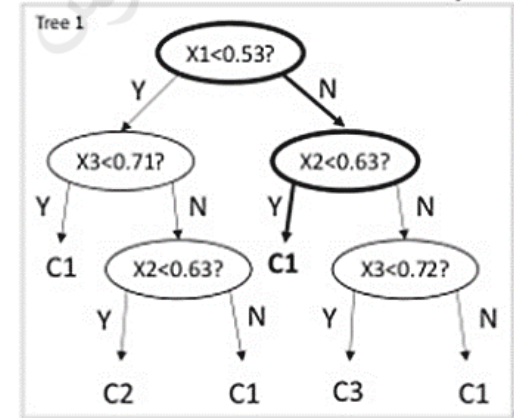
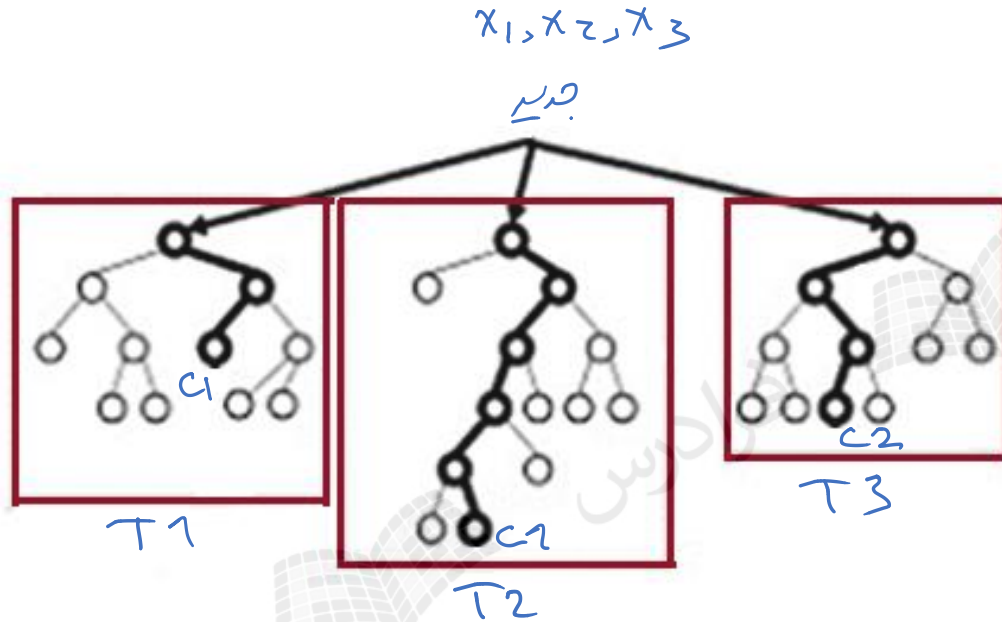


f_1
 f_2
 f_3
 f_4
 class



$m = \sqrt{2} = 2 \rightarrow \text{out-of-bag error}$
 $3 \rightarrow //$

مثال



اکثریت $\rightarrow c_1$

کاربردهای جنگل تصادفی

بازار بورس : شناسایی رفتار بورس در آینده.

بانکداری : شناسایی مشتریان که قصد کلاهبرداری از بانک را دارند.

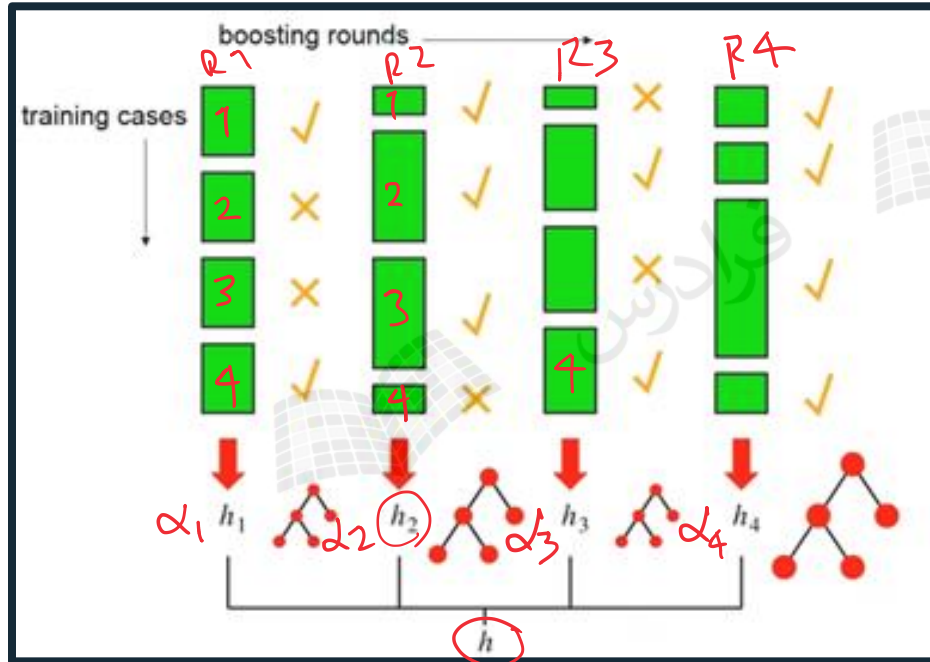
پزشکی : شناسایی ترکیب صحیحی از مولفه‌ها و تحلیل تاریخچه پزشکی بیمار، برای شناسایی بیماری او .

تجارت الکترونیک : شناسایی اینکه مشتریان یک محصول را دوست داشته‌اند یا خیر.

Boosting

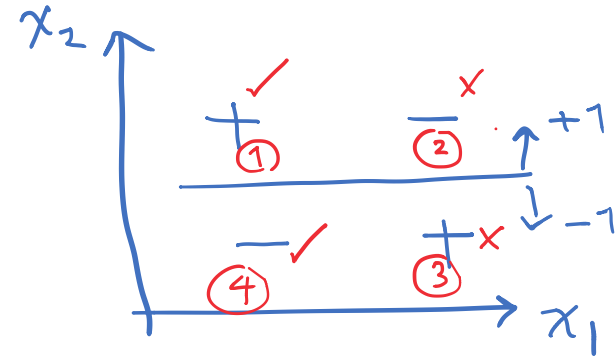
Boosting

دسته‌بند کلی از جمع وزن‌دار دسته‌بندهای پایه ساخته می‌شود.




$$H_m(x) = \alpha_1 h_1(x) + \dots + \alpha_m h_m(x)$$

$$\hat{y} = \text{sign}(H_m(x))$$



$$\frac{1}{4}$$

AdaBoost

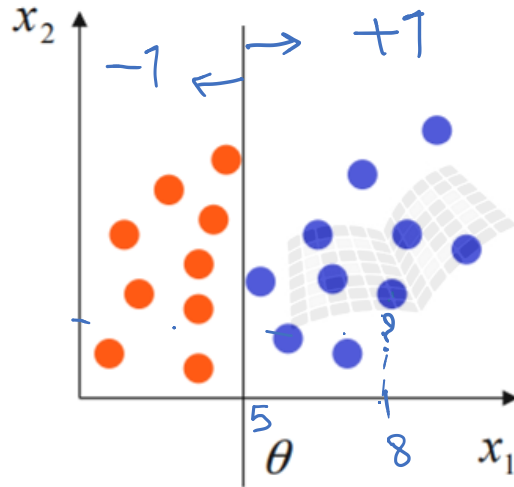
- هدف: بهینه کردن یک تابع هزینه
- به داده‌هایی که به اشتباه دسته‌بندی شده اند، وزن بیشتر داده می‌شود.
-  در حالی که دسته‌بند پیچیده‌ای را یاد می‌گیرد، به سرریز مقاوم است.
- دسته‌بند پایه: درخت تصمیم با عمق محدود یا شبکه عصبی چند لایه یا decision stumps.

Decision stump

A **decision stump** is a model consisting of a one-level [decision tree](#).

a decision tree with one internal node which is immediately connected to the terminal nodes .

decision stump makes a prediction based on the value of just a single input feature.

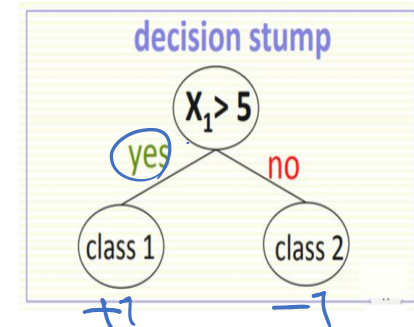


$$h(\underline{x}; j, \theta) = \begin{cases} +1 & x_j > \theta \\ -1 & \text{else} \end{cases}$$

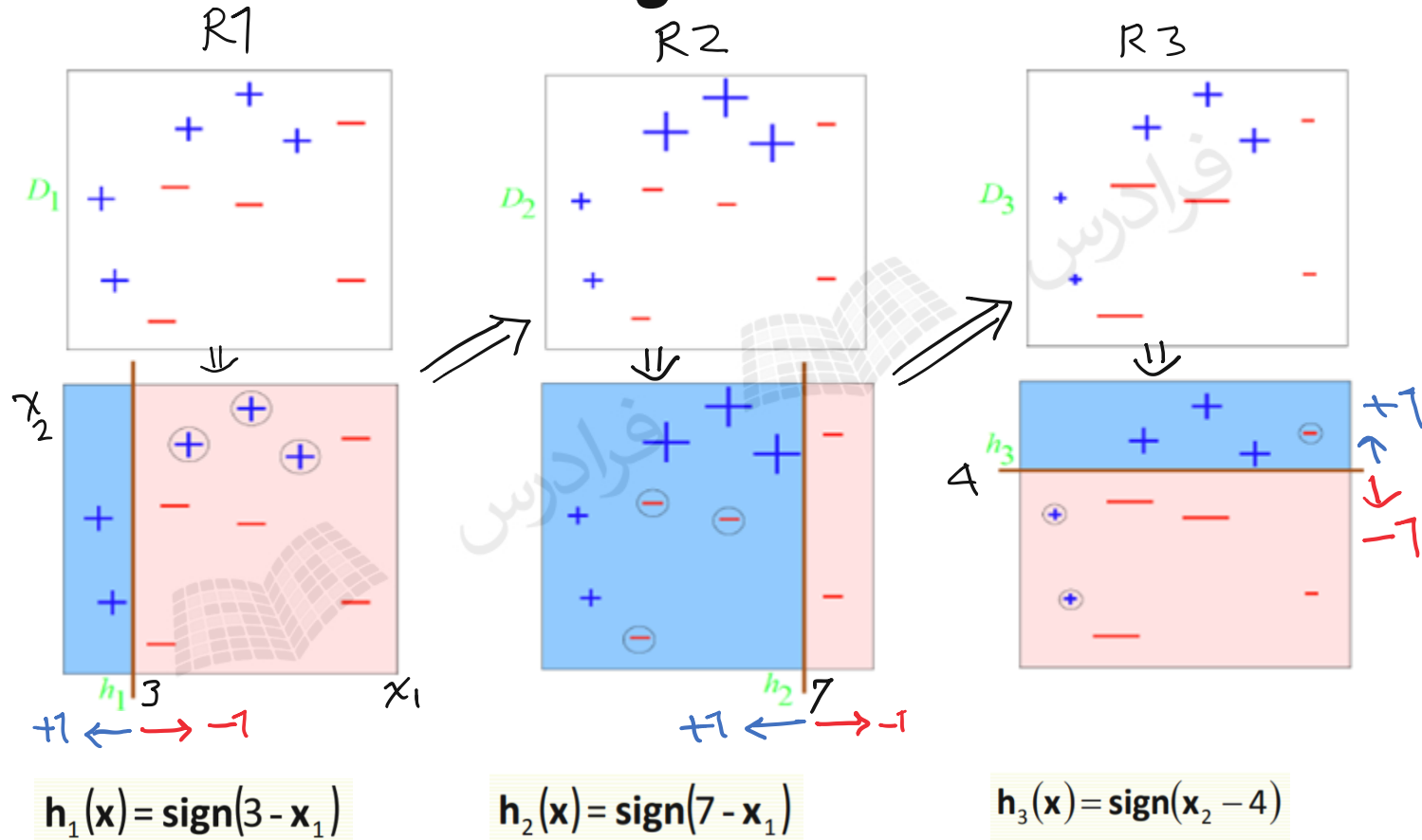
$$j=1$$

$$x_1 > 5$$

$$x = \begin{cases} 8 = x_1 \rightarrow +1 \\ 3 = x_2 \end{cases}$$



مثال



ترکیب سه دسته‌بند:

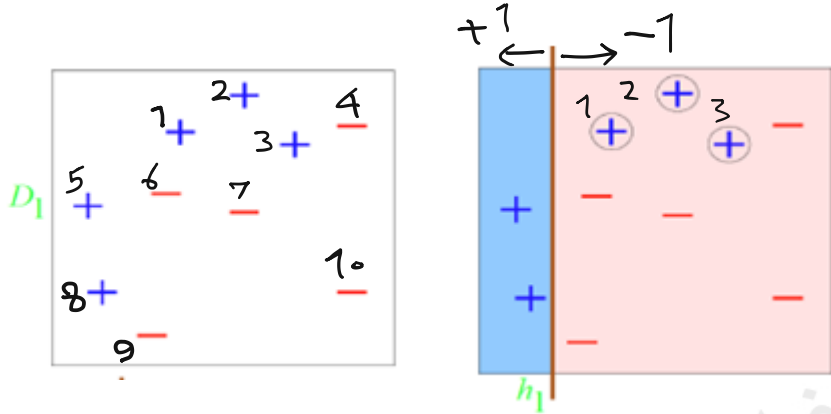
$$\mathbf{H}_{\text{final}} = \text{sign} \left(0.42 \begin{array}{|c|} \hline \text{Blue} \\ \hline \end{array} + 0.65 \begin{array}{|c|} \hline \text{Blue} \\ \hline \end{array} + 0.92 \begin{array}{|c|} \hline \text{Blue} \\ \hline \end{array} \right) = 4$$

$$H_f = \text{sign} \left(\alpha_1 h_1(x) + \alpha_2 h_2(x) + \alpha_3 h_3(x) \right)$$

$$X = \begin{cases} x_1 = 5 \\ x_2 = 2 \end{cases}$$

$$\text{sign}(-0.42 + 0.65 - 0.92) = -1$$

دور اول



① $J_1 = \frac{1}{10} + \frac{1}{10} + \frac{1}{10} = \frac{3}{10}$ مجموع وزن
 داده های اشتباه دسته بندی شده اند.

② $\epsilon_1 = \frac{J_1}{\text{مجموع وزن}} = \frac{\frac{3}{10}}{1} = 0.3$

③ $\alpha_1 = \frac{1}{2} \ln\left(\frac{1-\epsilon_1}{\epsilon_1}\right) \approx 0.42$

1	2	3	4	5	6	7	8	9	10
<u>0.10</u>	<u>0.10</u>	<u>0.10</u>	0.10	0.10	0.10	0.10	0.10	0.10	0.10
0.15	0.15	0.15	0.07	0.07	0.07	0.07	0.07	0.07	0.07
0.17	0.17	0.17	0.07	0.07	0.07	0.07	0.07	0.07	0.07

وزن

وزن جدید

$+0.42$

$0.1 \times e$

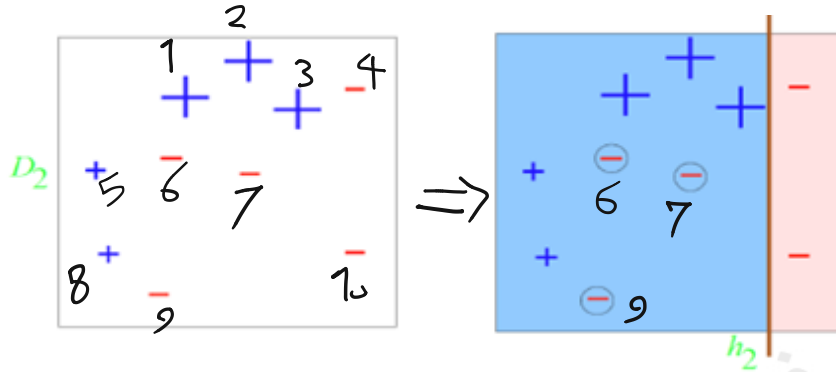
$= 0.15$

$+ \alpha_1$
وزن $\times e$

$- \alpha_1$
وزن $\times e$

$0.1 \times e^{-0.42} \approx 0.07$

دور دوم



$$\textcircled{1} J_2 = 3(0.07) = 0.21$$

$$\textcircled{2} \epsilon_2 = \frac{J_2}{\text{مجموع وزن}} \approx 0.21$$

$$\textcircled{3} \alpha_2 = \frac{1}{2} \ln\left(\frac{1 - \epsilon_2}{\epsilon_2}\right) \approx 0.65$$

وزن $\textcircled{4}$

$$\begin{cases} 0.17 \times e^{-0.65} \approx 0.09 \\ 0.07 \times e^{-0.65} \approx 0.04 \end{cases}$$

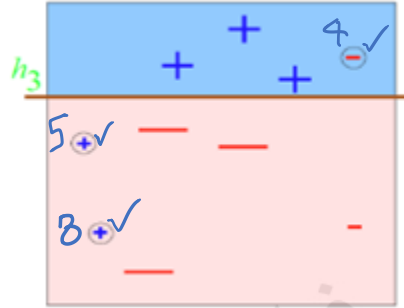
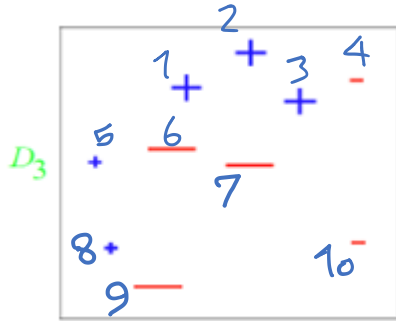
$$\text{نتیجه} \rightarrow 0.07 \times e^{+0.65} \approx 0.14$$

وزن
وزن جدید
نرمال

1	2	3	4	5	6	7	8	9	10
0.17	0.17	0.17	0.07	0.07	<u>0.07</u>	<u>0.07</u>	0.07	<u>0.07</u>	0.07
0.09	0.09	0.09	0.04	0.04	0.14	0.14	0.04	0.14	0.04
0.11	0.11	0.11	0.05	0.05	0.17	0.17	0.05	0.17	0.05

$$Z_2 = 0.82$$

دور سوم



$$\textcircled{1} J_3 = 3(0.05) = 0.15$$

$$\textcircled{2} \epsilon_3 = \frac{0.15}{\text{مجموع وزن}} \approx 0.14$$

$$\textcircled{3} \alpha_3 = \frac{1}{2} \ln\left(\frac{1-\epsilon_3}{\epsilon_3}\right) \approx 0.92$$

④ وزن :

	1	2	3	4	5	6	7	8	9	10
وزن	0.11	0.11	0.11	<u>0.05</u>	<u>0.05</u>	0.17	0.17	<u>0.05</u>	0.17	0.05
وزن جدید	0.04	0.04	0.04	0.11	0.11	0.07	0.07	0.11	0.07	0.02

$$0.05 \times e^{+0.92} \approx 0.17$$

$$\begin{cases} 0.11 \times e^{-0.92} \approx 0.04 \\ 0.17 \times e^{-0.92} \approx 0.07 \\ 0.05 \times e^{-0.92} \approx 0.02 \end{cases}$$

الگوریتم AdaBoost

Given: $(x_1, y_1), \dots, (x_m, y_m)$ where $x_i \in X, y_i \in Y = \{-1, +1\}$

Initialize $D_1(i) = 1/m$. ←

For $t = 1, \dots, T$:

$T=3$

- Train weak learner using distribution D_t .
- Get weak hypothesis $h_t : X \rightarrow \{-1, +1\}$ with error

$$\epsilon_t = \Pr_{i \sim D_t} [h_t(x_i) \neq y_i].$$

- Choose $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)$.
- Update:

$$\begin{aligned} D_{t+1}(i) &= \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{if } h_t(x_i) \neq y_i \end{cases} \rightarrow \text{اسبه} \\ &= \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t} \end{aligned}$$

where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

Output the final hypothesis:

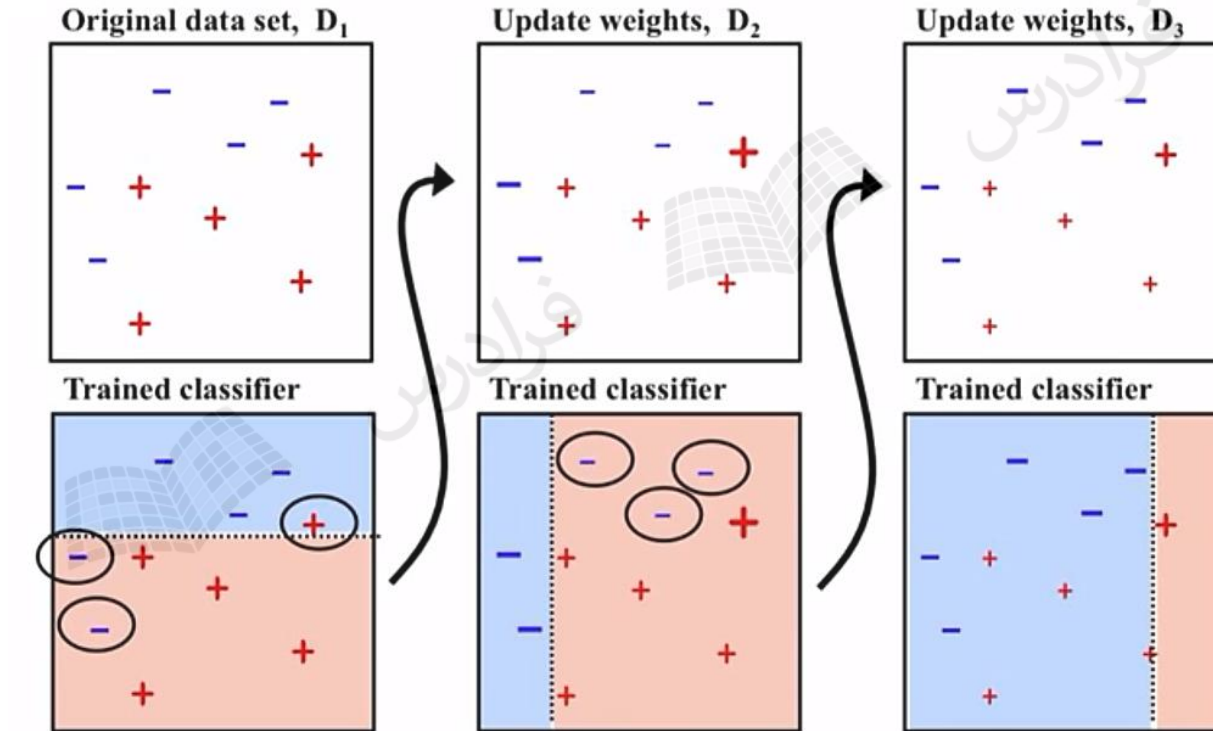
$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right).$$

$$\begin{aligned} m=10 \quad D_1(i) &= \frac{1}{m} \\ h_1 \rightarrow \begin{cases} \epsilon_1 \\ \alpha_1 = \frac{1}{2} \ln \left(\frac{1 - \epsilon_1}{\epsilon_1} \right) \\ D_2(i) = \frac{D_1(i) e^{\pm \alpha_1}}{Z} \end{cases} \end{aligned}$$

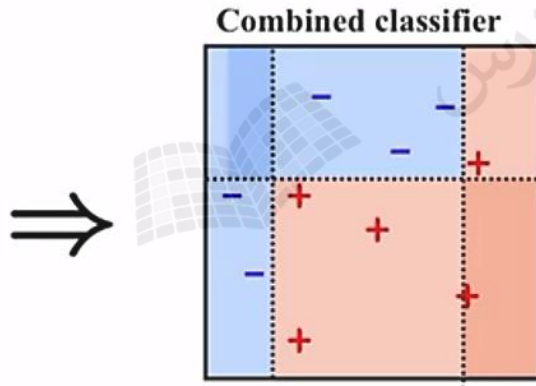
$$h_2 \rightarrow \begin{cases} \epsilon_2 \\ \alpha_2 \\ D_3(i) \end{cases} \quad h_3$$

$$H(x) = \text{sign} (\alpha_1 h_1 + \alpha_2 h_2 + \alpha_3 h_3)$$

مثال

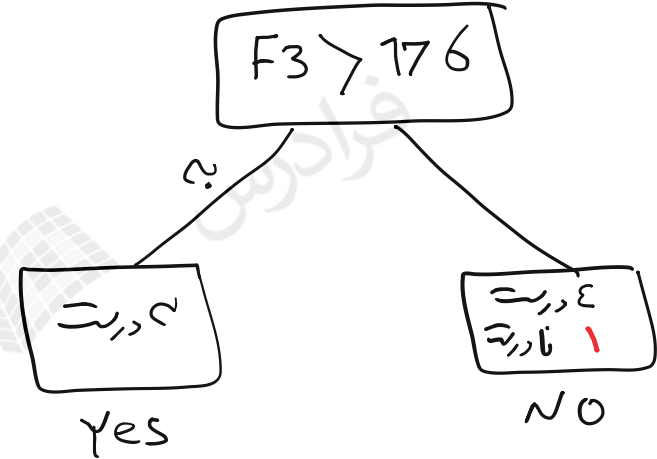


$$.33 * \begin{array}{|c|} \hline \text{Blue} \\ \hline \text{Orange} \\ \hline \end{array} + .57 * \begin{array}{|c|} \hline \text{Blue} \\ \hline \text{Orange} \\ \hline \end{array} + .42 * \begin{array}{|c|} \hline \text{Blue} \\ \hline \text{Orange} \\ \hline \end{array} \geq 0$$



مثال

	X			Y	
	F1	F2	F3	class	
1	Yes	Yes	205	Yes +1	$\frac{1}{8} \times e^{-\alpha_1}$
2	No	Yes	180	Yes	$\frac{1}{8} \times e^{-\alpha_1}$
3	Yes	No	210	Yes	$\frac{1}{8} \times e^{-\alpha_1}$
4 ← السبأه	Yes	Yes	167	Yes	$\frac{1}{8} \times e^{+\alpha_1}$
5	No	Yes	156	No -1	$\frac{1}{8} \times e^{-\alpha_1}$
6	No	Yes	125	No	$\frac{1}{8} \times e^{-\alpha_1}$
7	Yes	No	168	No	$\frac{1}{8} \times e^{-\alpha_1}$
8	Yes	Yes	172	No	$\frac{1}{8} \times e^{-\alpha_1}$

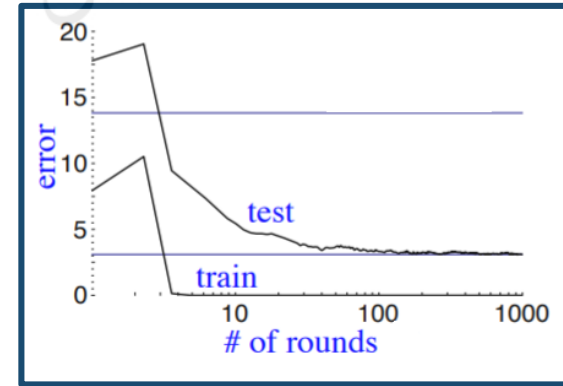
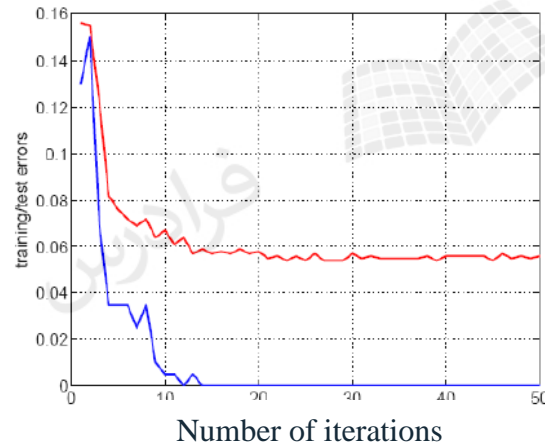
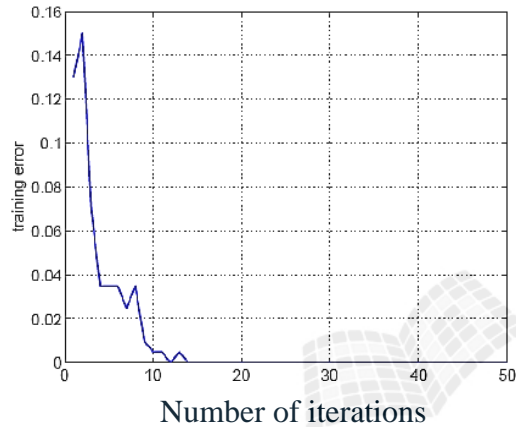


$$\epsilon_1 = \frac{1}{8} = \frac{1}{8}$$

$$\alpha_1 = \frac{1}{2} \ln \left(\frac{1 - \epsilon_1}{\epsilon_1} \right)$$

$$\alpha_1 = \frac{1}{2} \ln 7$$

خطای آموزش



Loss
J

Generalization Error Bounds

$$error_{true}(H) \leq error_{train}(H) + \tilde{O} \left(\sqrt{\frac{Td}{m}} \right)$$

T : تعداد دور

d : VC

یادگیر ضعیف

m : تعداد نمونه‌ها

آزمایشی

$$error_{true}(H) \leq \hat{\Pr} \left[\underbrace{\text{margin}_f(x, y)}_{\text{margin}} \leq \theta \right] + \tilde{O} \left(\sqrt{\frac{d}{m\theta^2}} \right)$$

این اسلایدها بر مبنای نکات مطرح شده در فرادرس
«آموزش یادگیری ماشین (Machine Learning) (تئوری - عملی) - بخش دوم»
تهیه شده است.

برای کسب اطلاعات بیشتر در مورد این آموزش به لینک زیر مراجعه نمایید.

faradars.org/fvdm94062