

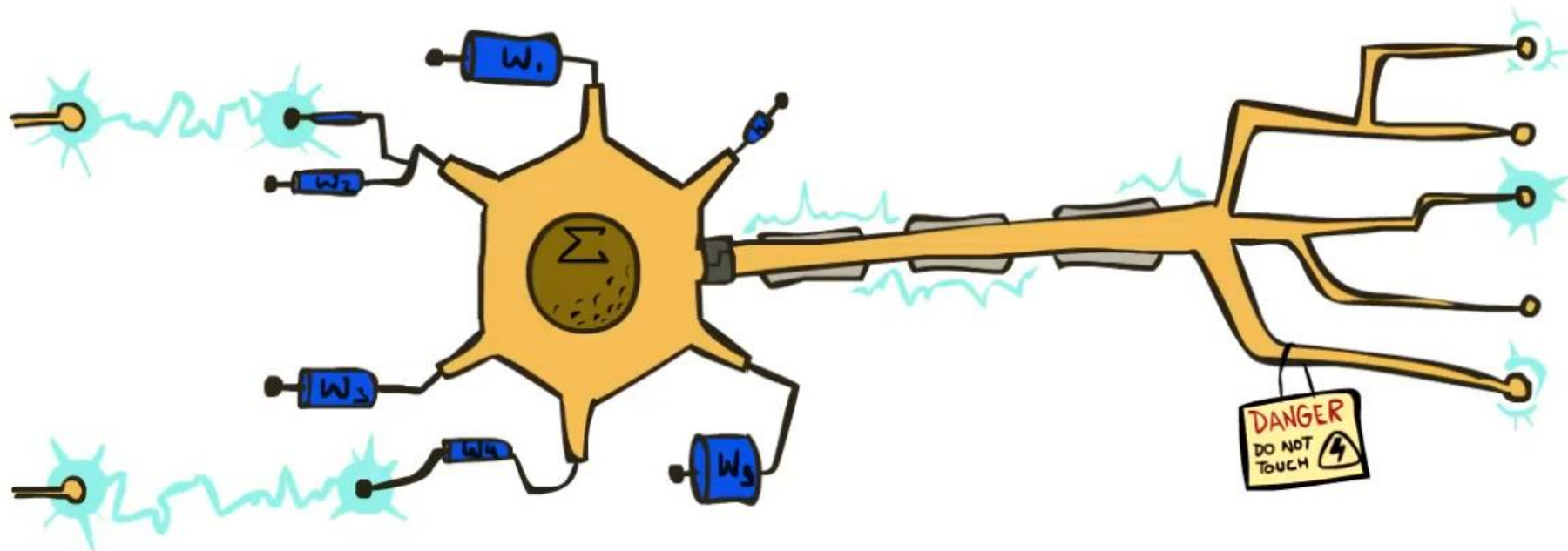
پرسپترون‌ها

سید ناصر رضوی n.razavi@tabrizu.ac.ir

۱۳۹۵

پرسپترون‌ها

۲



کلاس‌بندی مبتنی بر خطا

۳



خطاها و چگونگی مقابله با آنها

□ دو مثال از خطا.

Dear GlobalSCAPE Customer,
GlobalSCAPE has partnered with ScanSoft to offer you the latest version of OmniPage Pro, for just \$99.99* - the regular list price is \$499! The most common question we've received about this offer is - Is this genuine? We would like to assure you that this offer is authorized by ScanSoft, is genuine and valid. You can get the . . .

. . . To receive your \$30 Amazon.com promotional certificate, click through to
<http://www.amazon.com/apparel>
and see the prominent link for the \$30 offer. All details are there. We hope you enjoyed receiving this message. However, if you'd rather not receive future e-mails announcing new store launches, please click . . .

خطاها و چگونگی مقابله با آنها

❑ **مشکل.** هنوز در صندوق شما هرزنامه وجود دارد!

❑ **نیاز به ویژگی‌های بیشتر** - کلمات به تنهایی کافی نیستند.

❑ آیا شما قبلاً نامه‌ای به فرستنده این نامه فرستاده‌اید؟

❑ آیا این نامه به یک میلیون شخص دیگر نیز ارسال شده است؟

❑ آیا اطلاعات فرستنده سازگار است؟

❑ آیا تمام متن نامه با حروف بزرگ نوشته شده است؟

❑ آیا آدرس‌های درون نامه شما را به مکان‌های درستی هدایت می‌کنند؟

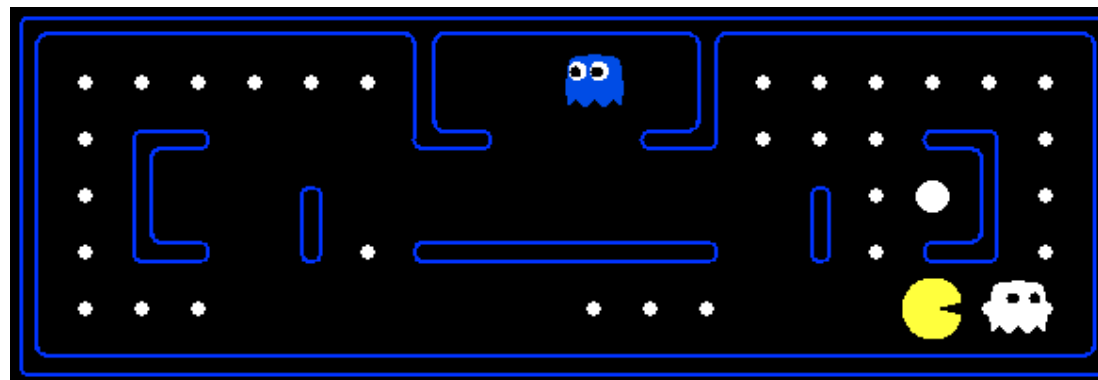
❑ آیا نامه خطاب به شما نوشته شده است؟ یعنی از نام شما استفاده شده است؟

❑ در شبکه‌های بیز ساده می‌توان انواع مختلفی از ویژگی‌ها را به کار برد، اما به نظر می‌رسد بهترین عملکرد این مدل‌ها زمانی است که ویژگی‌ها همگن هستند. (فقط کلمات)

جستجوی وب

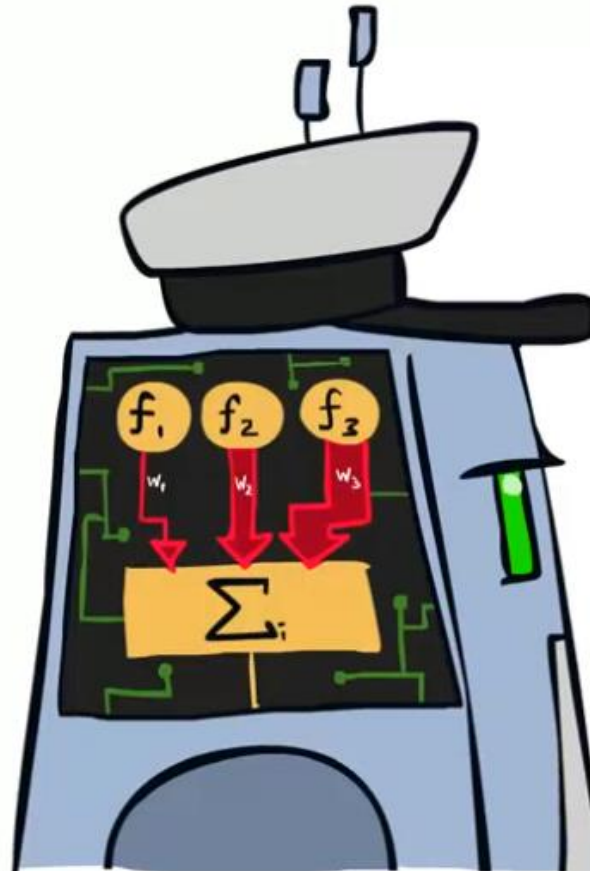


مسائل تصمیم‌گیری



کلاس بندی خطی

۷



بردار ویژگی

۸

x

$f(x)$

y

Hello,
Do you want free printer
cartridges? Why pay more
when you can get them
ABSOLUTELY FREE! Just



# of free	: 2
YOUR_NAME	: 0
MISSPELED	: 2
FROM_FRIEND	: 0
...	



SPAM

2



PIXEL (7 , 12)	: 1
PIXEL (7, 13)	: 0
...	
NUM_LOOPS	: 1
...	

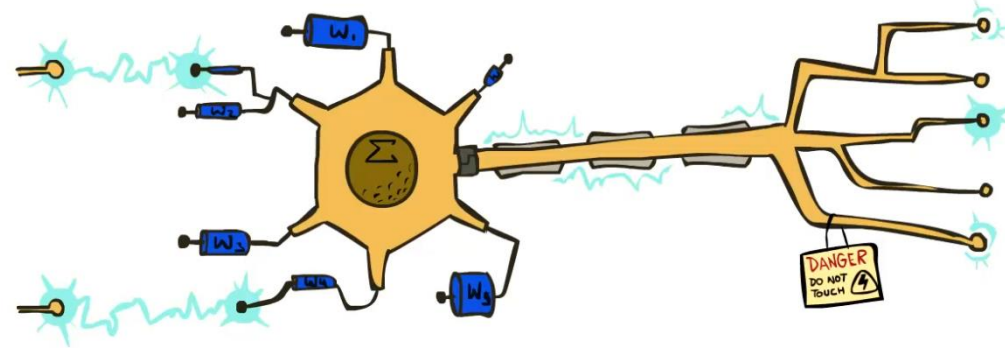
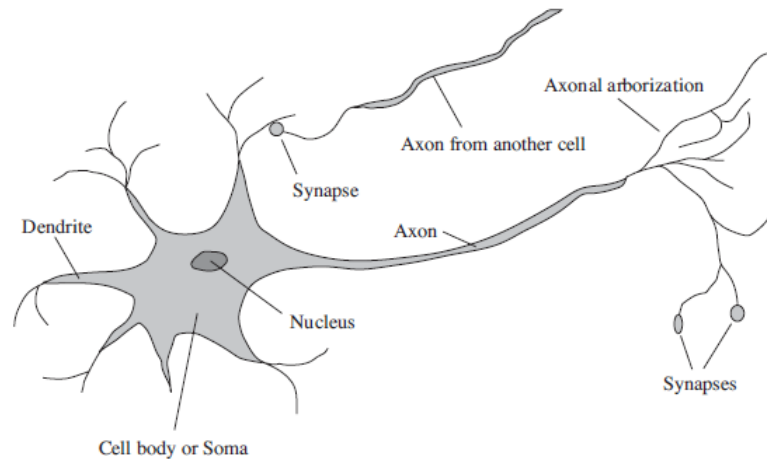


"2"

کمی زیست‌شناسی (به زبان ساده)

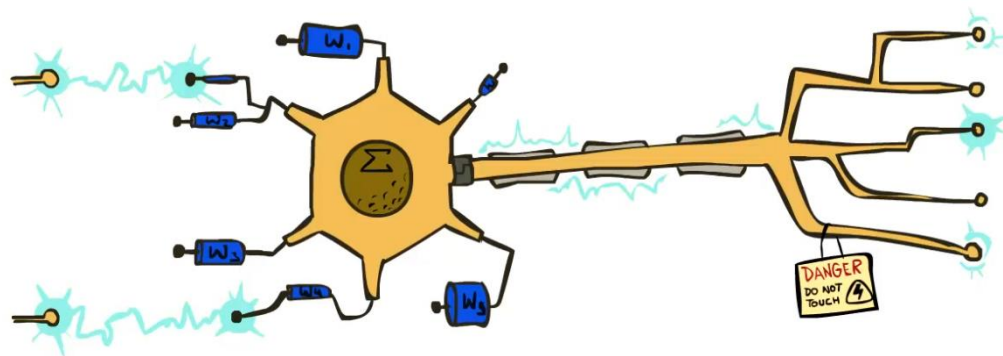
۹

□ یک الگو برداری بسیار ضعیف. نورون‌های انسان!



کلاس بندی خطی

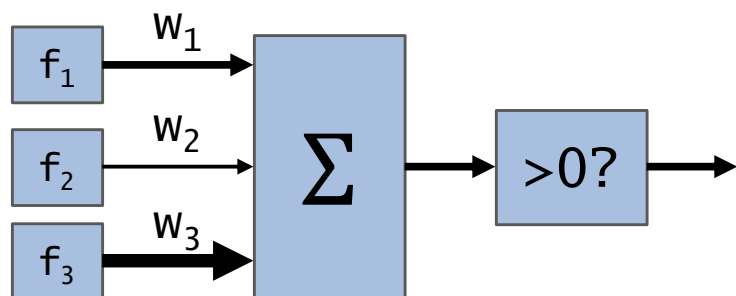
۱۰



- ورودی‌ها. مقدار ویژگی‌ها.
- وزن‌ها. هر ویژگی یک وزن دارد.
- فعالیت. مجموع وزن دار مقدار ویژگی‌ها.

ضرب داخلی

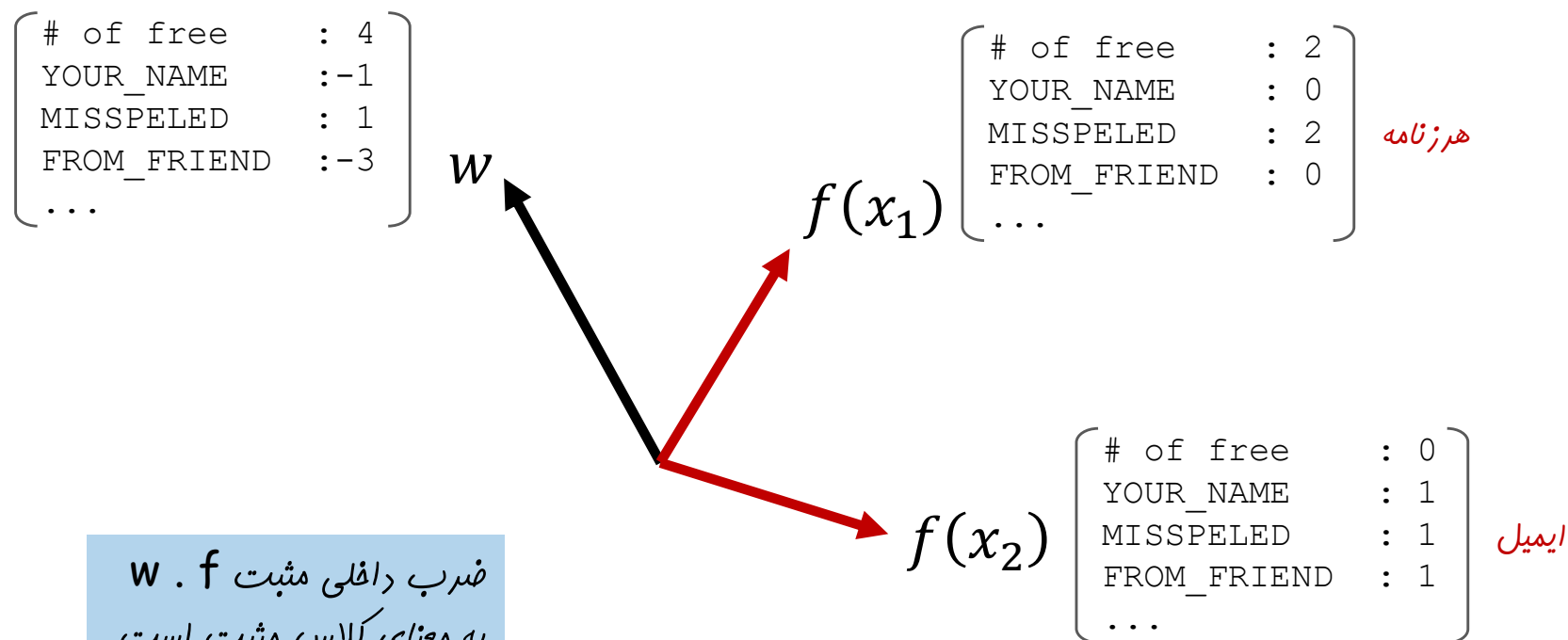
$$activation_w(x) = \sum_i w_i \cdot f_i(x) = w \cdot f(x)$$



□ خروجی.

- اگر مقدار فعالیت مثبت باشد، خروجی +۱
- اگر مقدار فعالیت منفی باشد، خروجی -۱

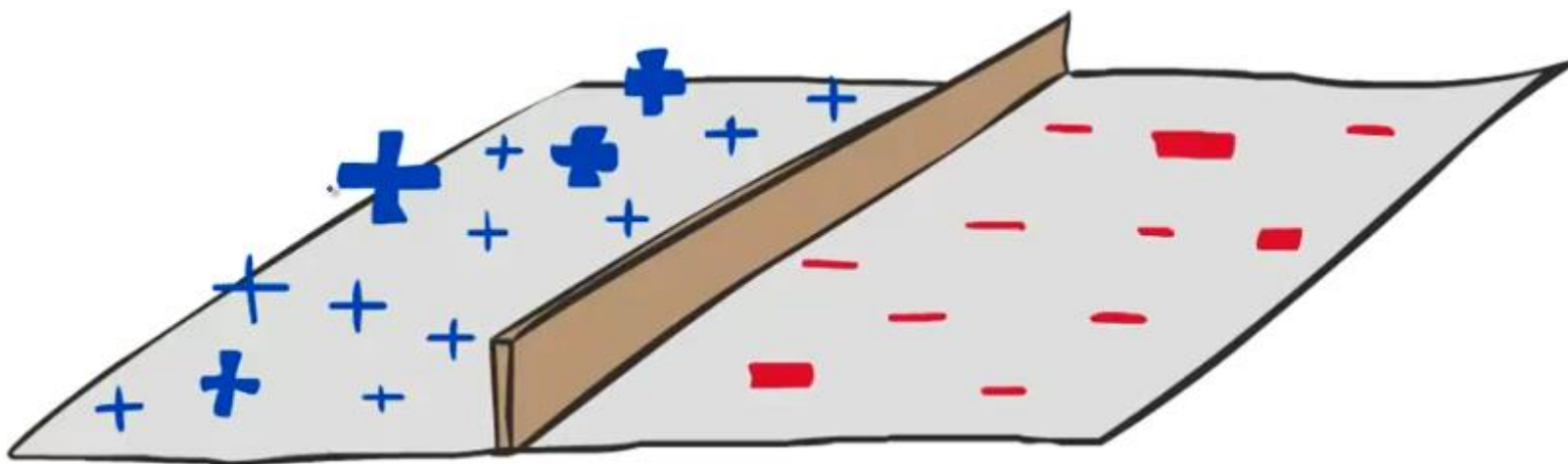
- حالت دودویی. مقایسه‌ی ویژگی‌ها با یک بردار وزن.
- یادگیری. تنظیم مقدار بردار وزن با توجه به نمونه‌های آموزشی.



ضرب داخلی مثبت $w \cdot f$
به معنای کلاس مثبت است.

قواعد تصمیم‌گیری

۱۲



قاعده‌ی تصمیم‌گیری دودویی

۱۳

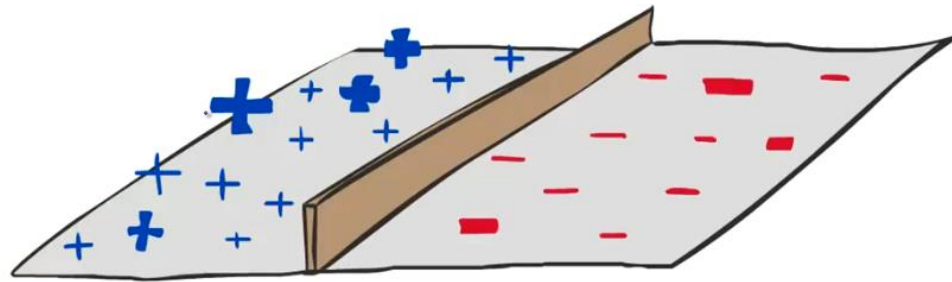
□ در فضای بردارهای ویژگی.

□ هر نمونه بیانگر یک نقطه است.

□ هر بردار وزن بیانگر یک ابرصفحه است.

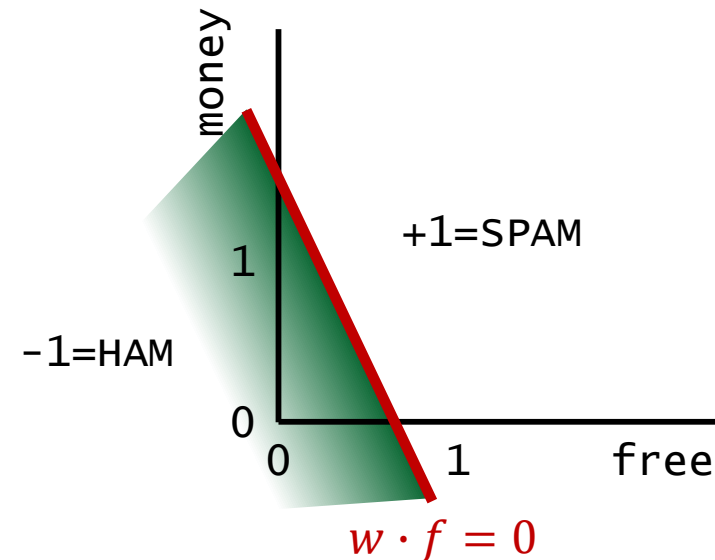
□ یک طرف ابرصفحه به کلاس مثبت اشاره دارد.

□ طرف دیگر آن به کلاس منفی اشاره دارد.



w

BIAS	: -3
free	: 4
money	: 2
...	



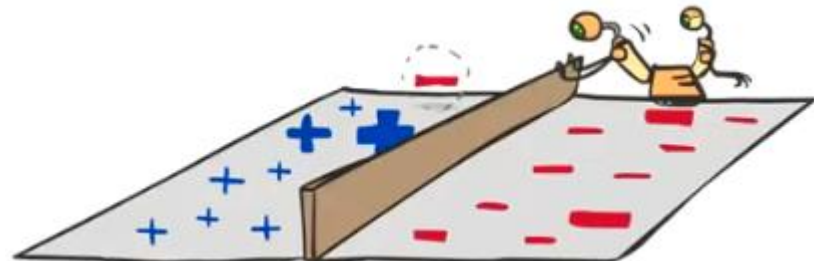
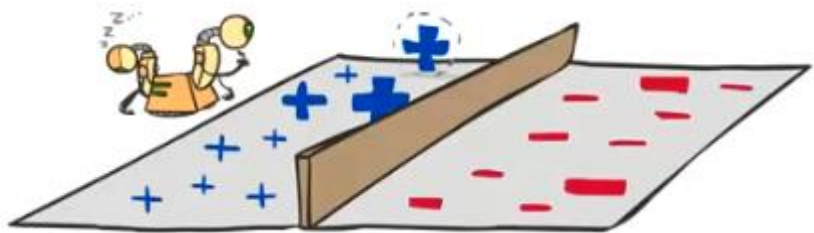
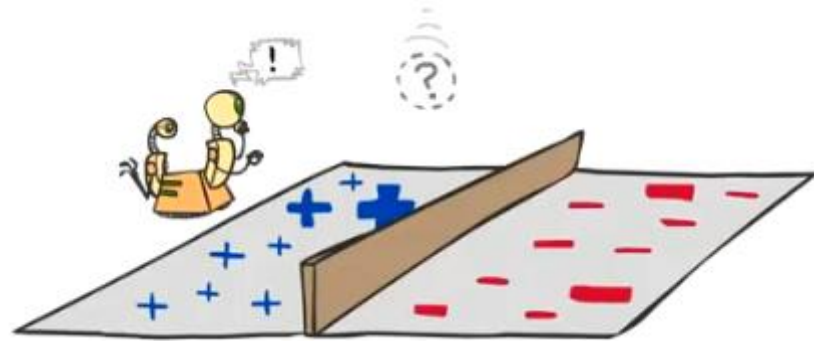
به روز رسانی وزن‌ها

۱۴



یادگیری: پرسپترون دودویی

۱۵



□ با بردار وزن صفر شروع کن.

□ به ازای هر نمونه‌ی آموزشی:

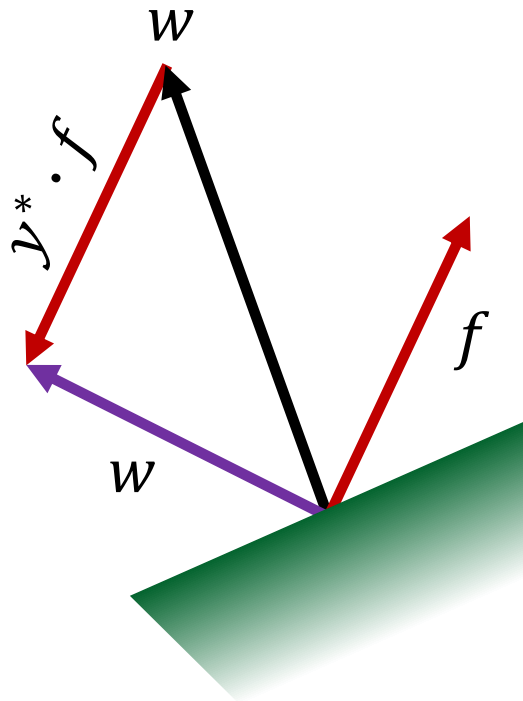
□ نمونه را با وزن‌های فعلی کلاس‌بندی کن.

□ اگر کلاس‌بندی درست، بدون تغییر!

□ اگر کلاس‌بندی نادرست، تغییر وزن‌ها!

یادگیری: پرسپترون دودویی

۱۶



□ با بردار وزن صفر شروع کن.

□ به ازای هر نمونه‌ی آموزشی:

□ نمونه را با وزن‌های فعلی کلاس‌بندی کن.

$$y = \begin{cases} +1, & w \cdot f(x) \geq 0 \\ -1, & w \cdot f(x) < 0 \end{cases}$$

□ اگر کلاس‌بندی درست، بدون تغییر!

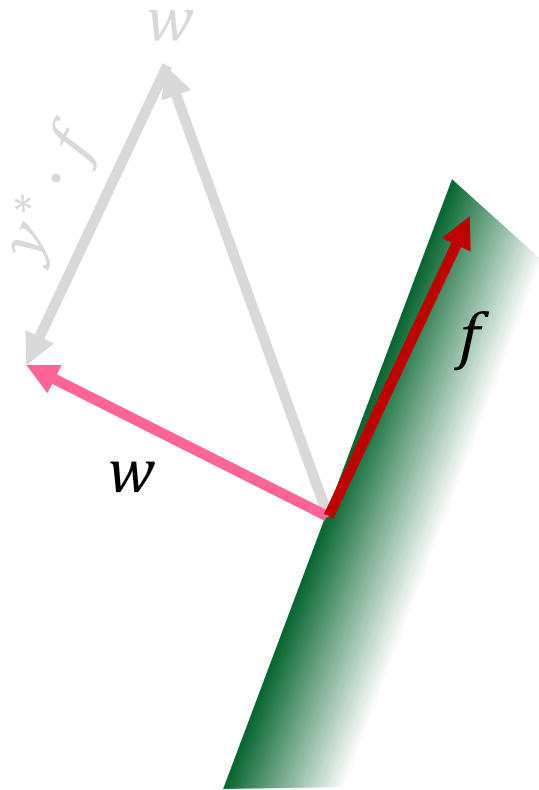
□ اگر کلاس‌بندی نادرست، تغییر وزن‌ها!

■ با افزودن یا کم کردن بردار ویژگی.

$$w = w + y^* \cdot f$$

یادگیری: پرسپترون دودویی

۱۷



□ با بردار وزن صفر شروع کن.

□ به ازای هر نمونه‌ی آموزشی:

□ نمونه را با وزن‌های فعلی کلاس‌بندی کن.

$$y = \begin{cases} +1, & w \cdot f(x) \geq 0 \\ -1, & w \cdot f(x) < 0 \end{cases}$$

□ اگر کلاس‌بندی درست، بدون تغییر!

□ اگر کلاس‌بندی نادرست، تغییر وزن‌ها!

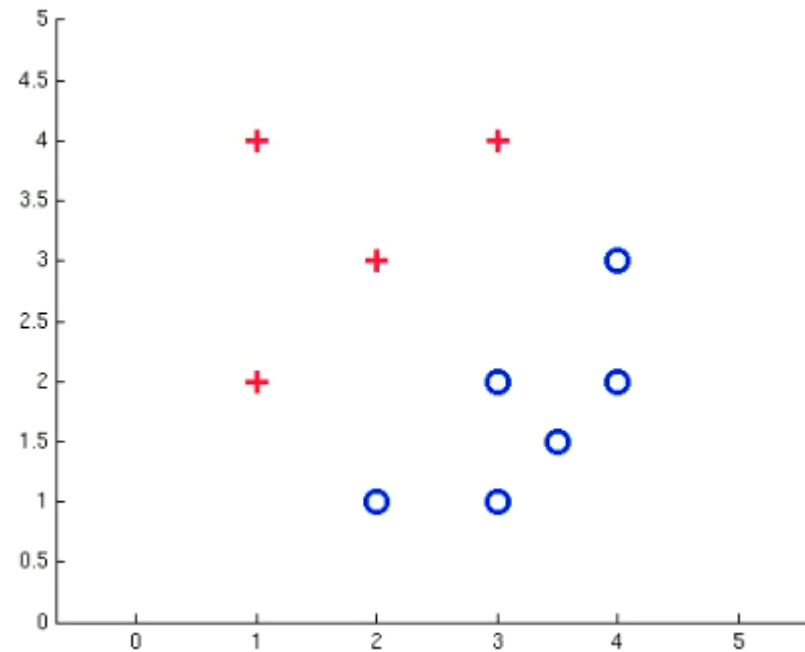
■ با افزودن یا کم کردن بردار ویژگی.

$$w = w + y^* \cdot f$$

پرسپترون: اجرای نمایشی

۱۸

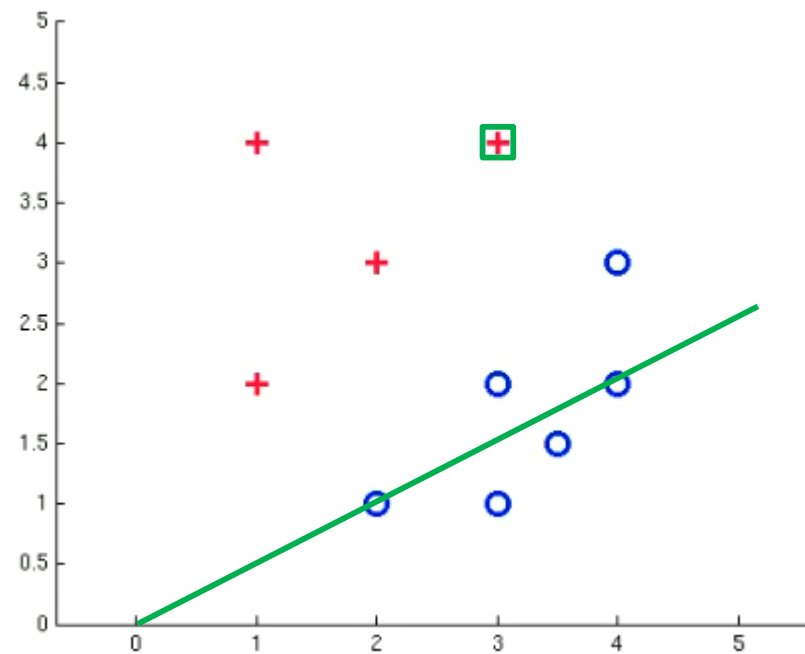
□ جداپذیر خطی.



پرسپترون: اجرای نمایشی

۱۹

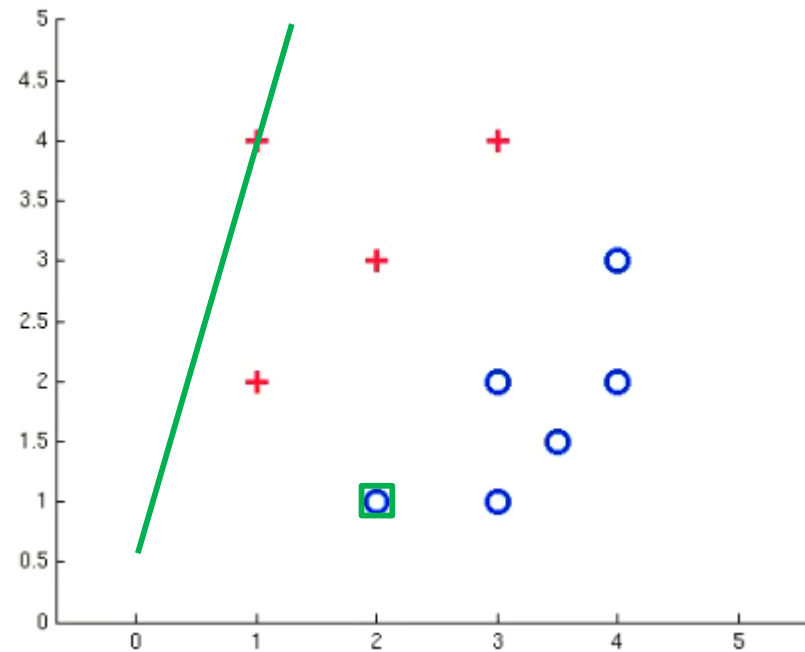
□ جداپذیر خطی.



پرسپترون: اجرای نمایشی

۲۰

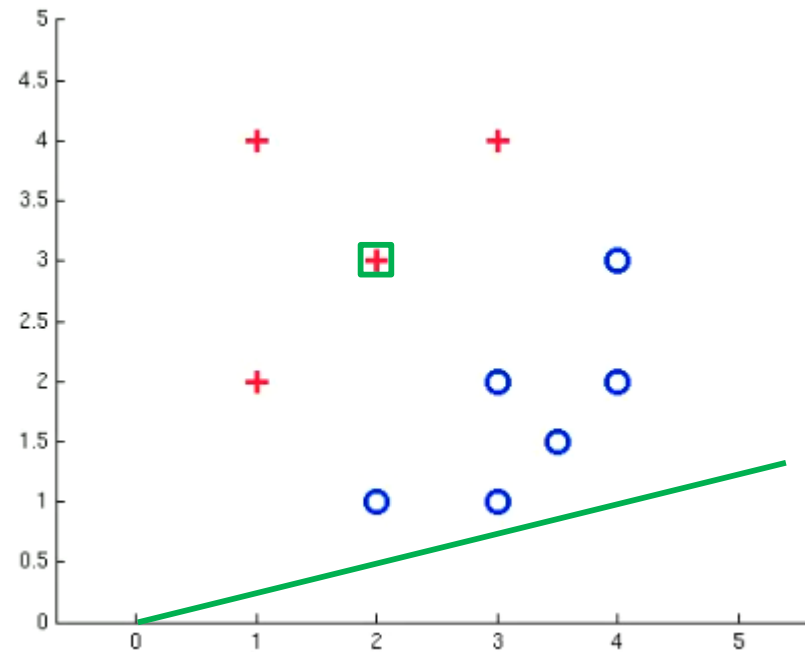
□ جداپذیر خطی.



پرسپترون: اجرای نمایشی

۲۱

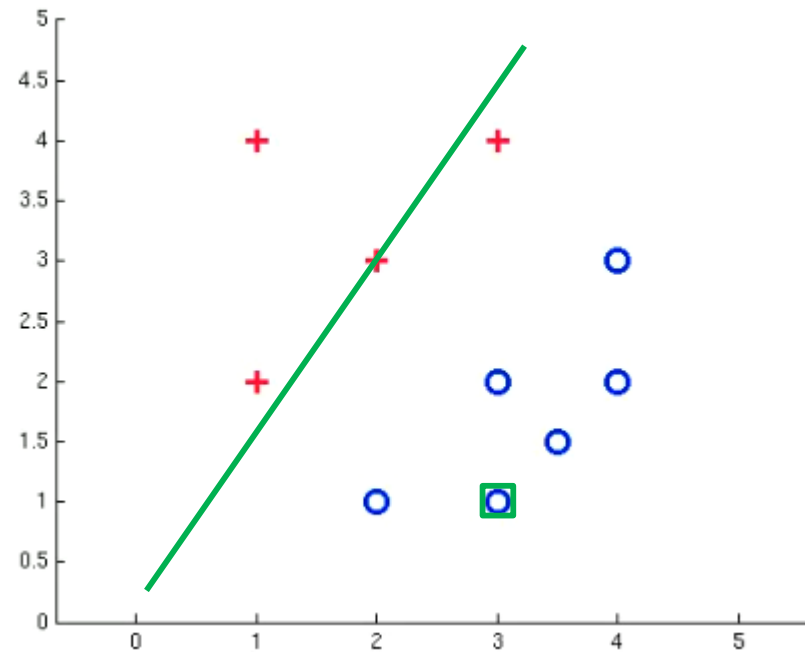
□ جداپذیر خطی.



پرسپترون: اجرای نمایشی

۲۲

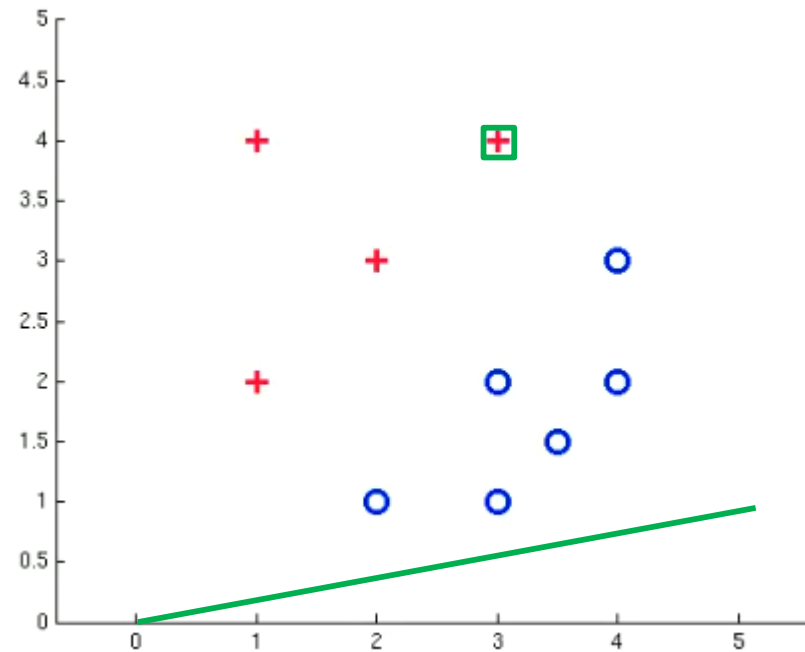
□ جداپذیر خطی.



پرسپترون: اجرای نمایشی

۲۳

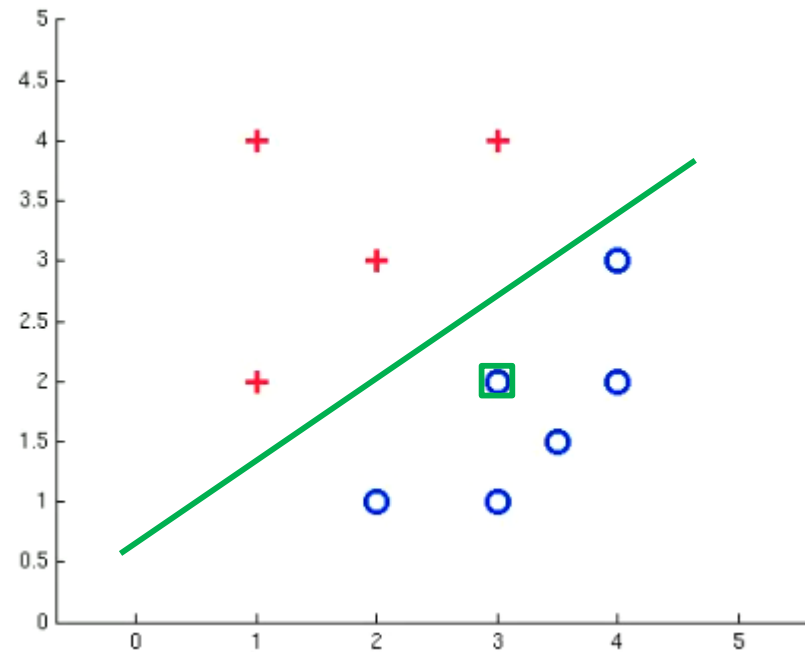
□ جداپذیر خطی.



پرسپترون: اجرای نمایشی

۲۴

□ جداپذیر خطی.



قاعده‌ی تصمیم‌گیری برای چندکلاس

۲۵

□ در صورت وجود چند کلاس.

□ به ازای هر کلاس y یک بردار وزن.

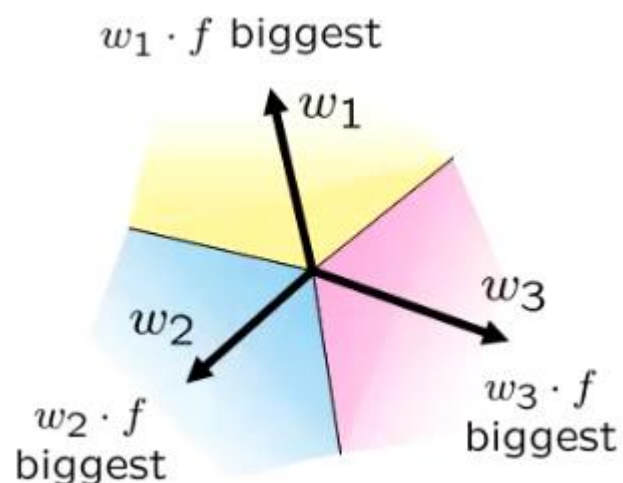
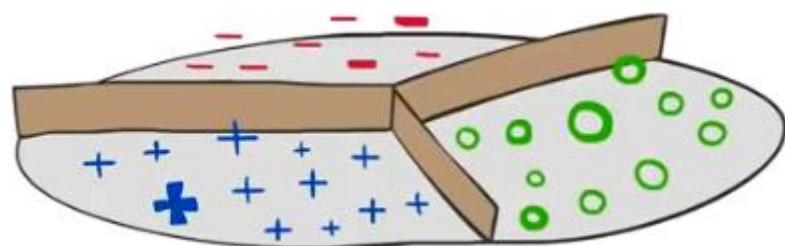
w_y

□ محاسبه‌ی امتیاز مربوط به کلاس y .

$w_y \cdot f(x)$

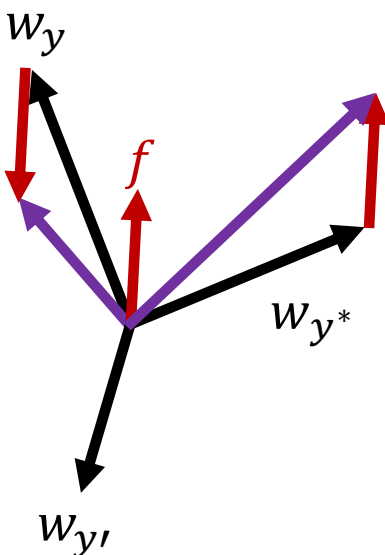
□ پیش‌بینی. برنده کلاس با بیشترین امتیاز.

$$y = \arg \max_y w_y \cdot f(x)$$



یادگیری: پرسپترون چند کلاسی

۲۶



□ با بردار وزن صفر شروع کن.

□ به ازای هر نمونه‌ی آموزشی:

□ نمونه را با وزن‌های فعلی کلاس‌بندی کن.

$$y = \arg \max_y w_y \cdot f(x)$$

□ اگر کلاس‌بندی درست، بدون تغییر!

□ اگر کلاس‌بندی نادرست، تغییر وزن‌ها!

■ کاهش امتیاز کلاس نادرست، افزایش امتیاز کلاس درست

$$\begin{aligned} w_y &= w_y - f \\ w_{y^*} &= w_{y^*} + f \end{aligned}$$

یادگیری: پرسپترون چند کلاسی

۲۷

□ با بردار وزن صفر شروع کن.

□ به ازای هر نمونه آموزشی:

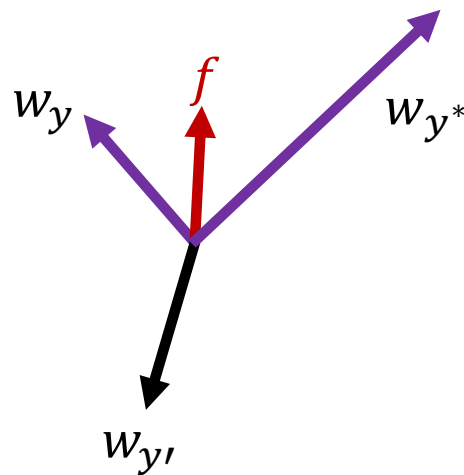
□ نمونه را با وزن‌های فعلی کلاس‌بندی کن.

$$y = \arg \max_y w_y \cdot f(x)$$

□ اگر کلاس‌بندی درست، بدون تغییر!

□ اگر کلاس‌بندی نادرست، تغییر وزن‌ها!

■ کاهش امتیاز کلاس نادرست، افزایش امتیاز کلاس درست



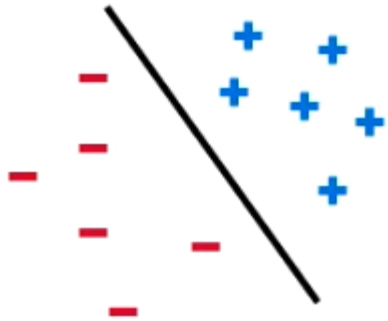
$$\begin{aligned} w_y &= w_y - f \\ w_{y*} &= w_{y*} + f \end{aligned}$$

پرسپترون: اجرای نمایشی

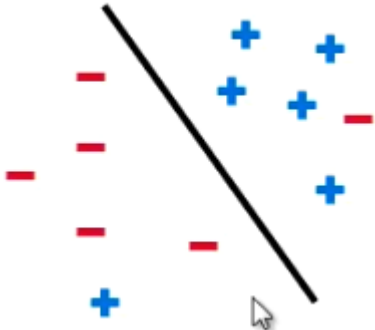
پرسپترون: ویژگی‌ها

۲۹

تفکیک پذیر



تفکیک ناپذیر



□ تفکیک پذیری. اگر مجموعه‌ای از پارامترها وجود داشته باشند که داده‌های آموزشی را به درستی تفکیک کنند.

□ همگرایی. اگر داده‌های آموزشی تفکیک پذیر باشند، پرسپترون در نهایت همگرا خواهد شد (حالت دودویی).

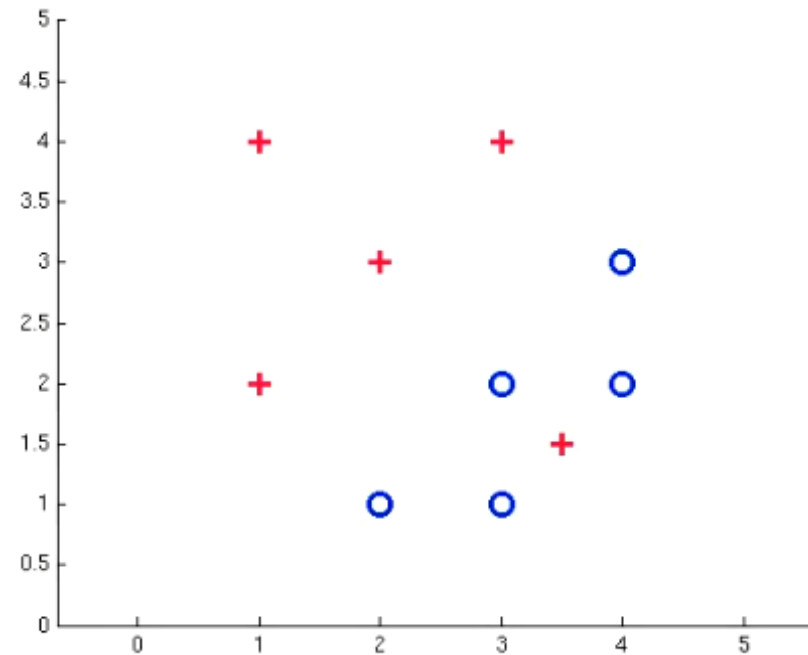
□ کران بالای خطا. حداکثر تعداد خطاها (در حالت دودویی) با حاشیه و یا درجه‌ی تفکیک پذیری مرتبط است.

$$mistakes < \frac{k}{\delta^2}$$

پرسپترون: مثال

۳۰

□ داده‌های تفکیک‌ناپذیر خطی.

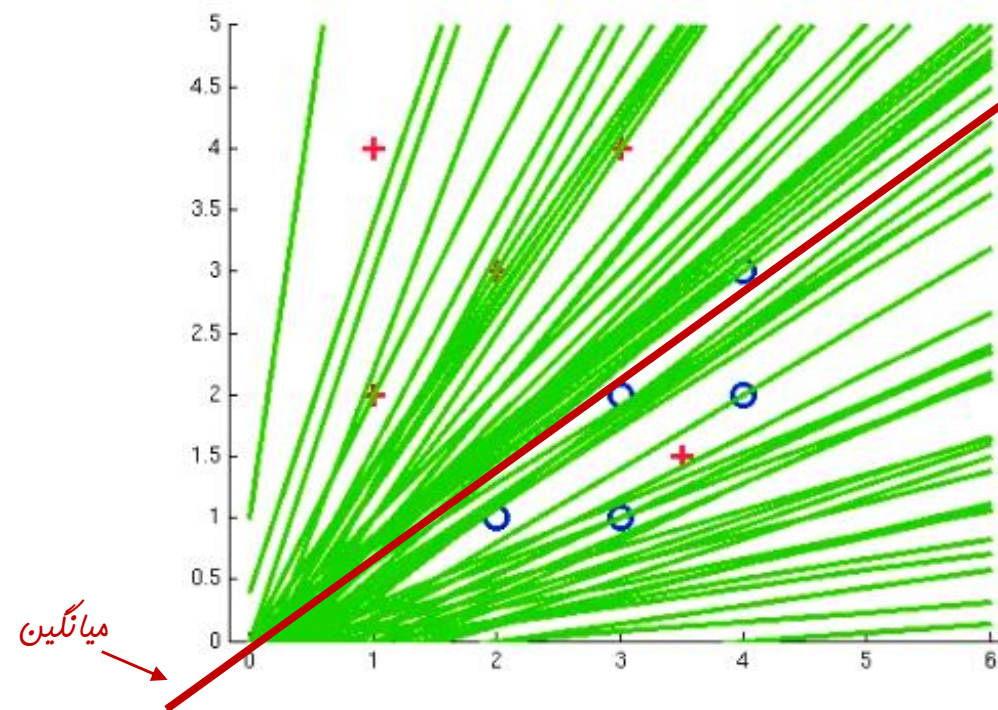


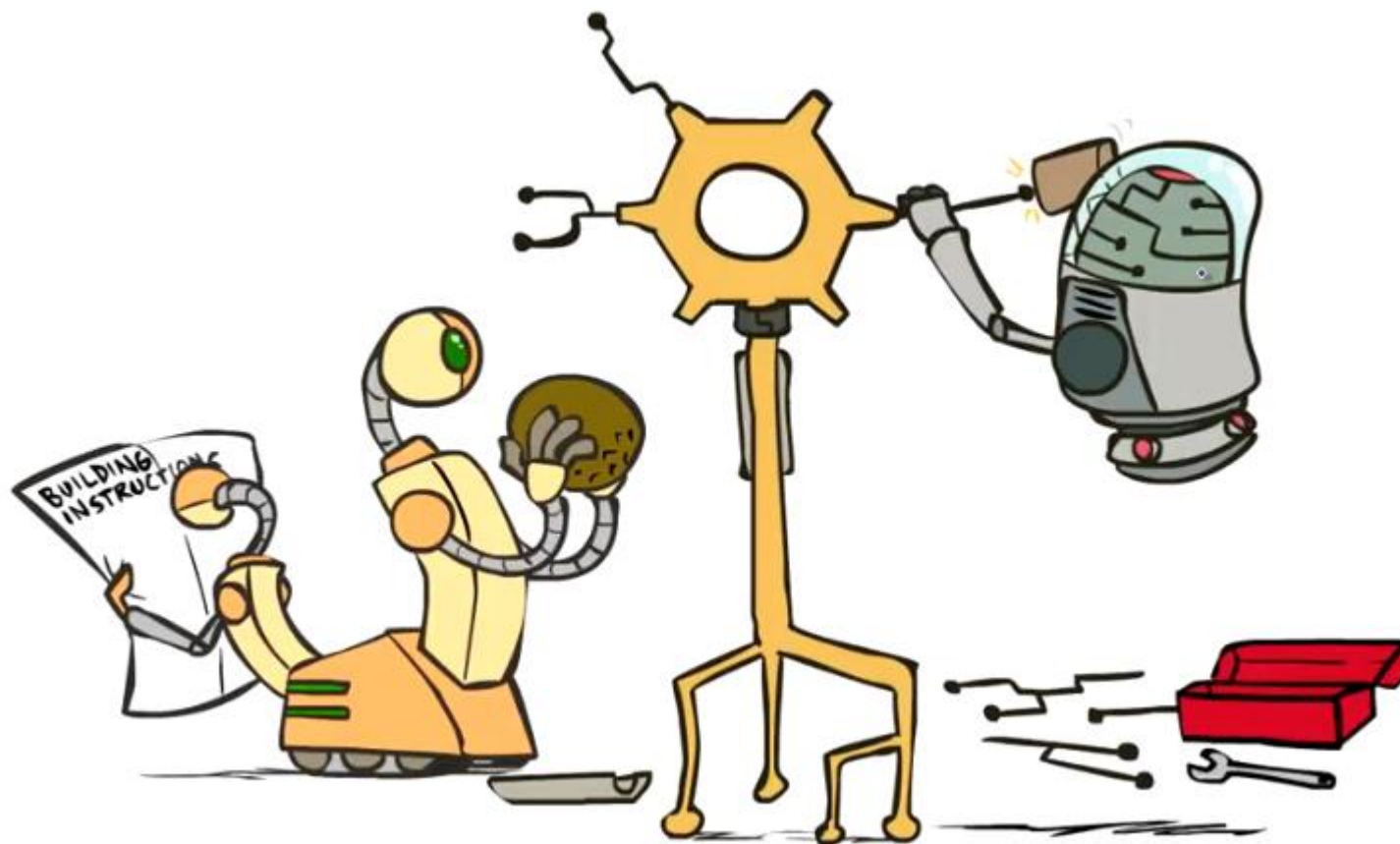
پرسپترون: مثال

۳۱

□ داده‌های تفکیک‌ناپذیر خطی.

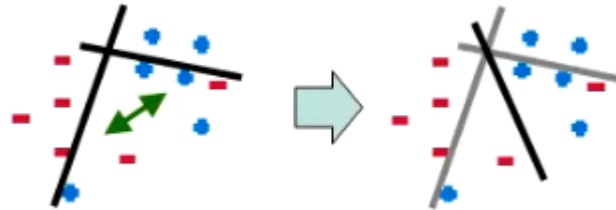
فرضیه‌های به دست آمده از ۱۰۰ اجرا



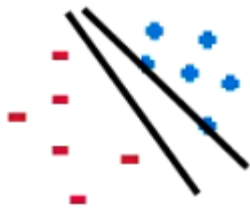


مشکلات پرسپترون

۳۳

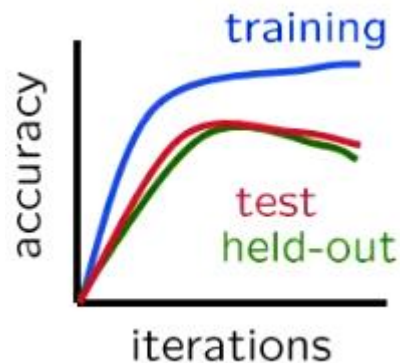


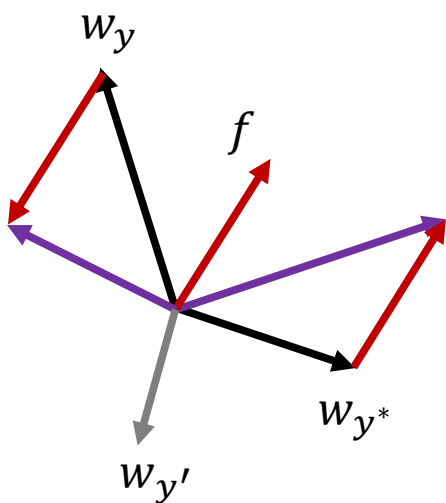
□ **نویز.** اگر داده‌ها تفکیک‌پذیر نباشند، وزن‌ها ممکن است همگرا نشوند!
□ میانگین‌گیری از بردارهای وزن در طول زمان می‌تواند کمک کند. [پرسپترون میانگین]



□ **تعمیم حد وسط.** یافتن راه‌حلی که داده‌ها را به خوبی تفکیک می‌کند، حتی اگر داده‌ها تفکیک‌پذیر باشند.

□ **آموزش بیش از حد.** دقت بر روی داده‌های آزمایشی و اعتبارسنجی ابتدا افزایش و سپس کاهش می‌یابد.
□ نوعی از بیش‌برازش.





حدس زدن y به جای y^* برای
نمونه x با ویژگی‌های $f(x)$

$$w_y = w_y - \tau f$$

$$w_{y^*} = w_{y^*} + \tau f$$

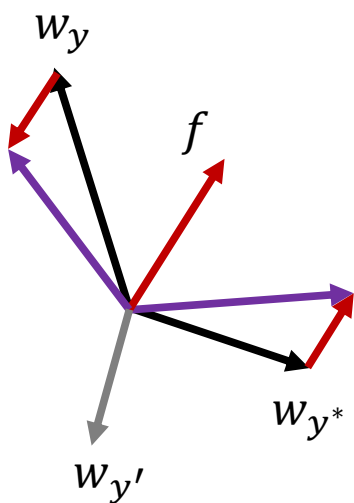
□ ایده. به روز رسانی وزن‌ها برای کاهش اثر این مشکلات.

□ $MIRA^*$. انتخاب یک اندازه برای مقدار به روز رسانی به گونه‌ای که اشتباه فعلی را تصحیح کند ...
□ ... اما مقدار تغییر w را به حداقل برساند.

$$\min_w \frac{1}{2} \sum_y \|w_y - w_{y'}\|^2$$

$$w_{y^*} \cdot f(x) \geq w_y \cdot f(x) + 1$$

افزایش قابلیت تعمیم



حدس زدن y به جای y^* برای
نمونه x با ویژگی‌های $f(x)$

$$w_y = w_y - \tau f$$

$$w_{y^*} = w_{y^*} + \tau f$$

□ ایده. به روز رسانی وزن‌ها برای کاهش اثر این مشکلات.

□ **MIRA***. انتخاب یک اندازه برای مقدار به روز رسانی به گونه‌ای که اشتباه فعلی را تصحیح کند ...
□ ... اما مقدار تغییر w را به حداقل برساند.

$$\min_w \frac{1}{2} \sum_y \|w_y - w_{y'}\|^2$$

$$w_{y^*} \cdot f(x) \geq w_y \cdot f(x) + 1$$

افزایش قابلیت تعمیم

به روز رسانی تصحیح کننده حداقلی

۳۶

$$\min_w \frac{1}{2} \sum_y \|w_y - w'_y\|^2$$

$$w_{y^*} \cdot f \geq w_y \cdot f + 1$$



$$\min_{\tau} \|\tau f\|^2$$

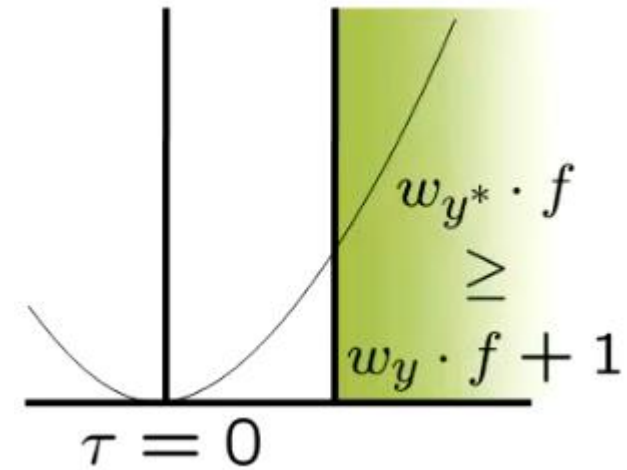
$$w_{y^*} \cdot f \geq w_y \cdot f + 1$$



$$(w'_{y^*} + \tau f) \cdot f \geq (w'_y - \tau f) \cdot f + 1$$

$$\tau = \frac{(w'_y - w'_{y^*}) \cdot f + 1}{2f \cdot f}$$

$$\begin{aligned} w_y &= w'_y - \tau f(x) \\ w_{y^*} &= w'_{y^*} + \tau f(x) \end{aligned}$$



حداکثر اندازه گام

۳۷

□ در عمل به روز رسانی‌های بسیار بزرگ نیز خوب نیستند!

□ نمونه‌ها ممکن است به غلط برچسب گذاری شوند.

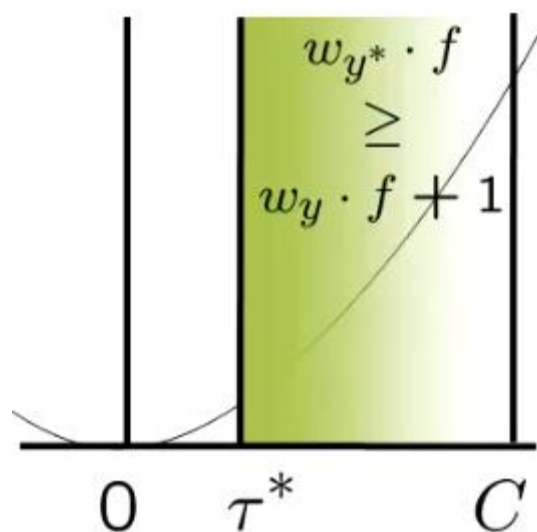
□ ممکن است به اندازه‌ی کافی ویژگی وجود نداشته باشد.

□ راه‌حل: محدود کردن حداکثر اندازه‌ی τ با مقدار ثابت C .

$$\tau^* = \min\left(\frac{(w'_y - w'_{y^*}) \cdot f + 1}{2f \cdot f}, C\right)$$

□ معمولاً سریع‌تر از پرسپترون همگرا می‌شود.

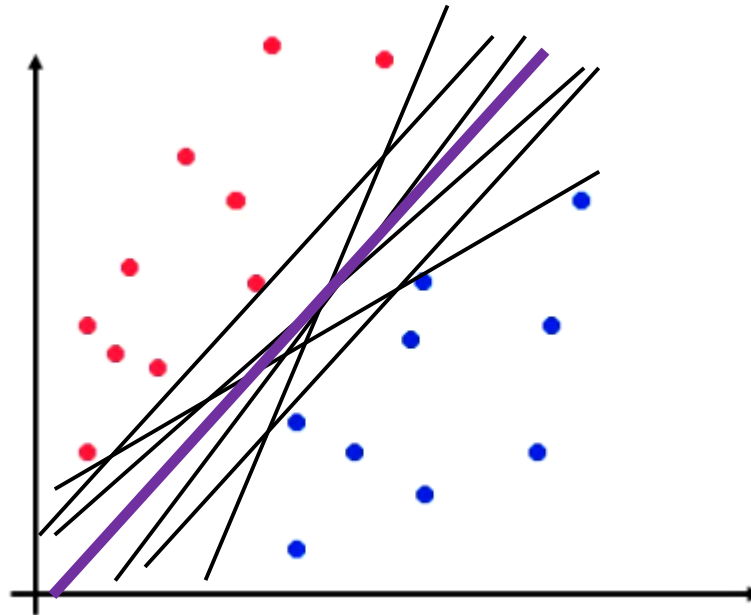
□ معمولاً بهتر است، به ویژه برای داده‌های دارای نویز.



تفکیک کننده‌های خطی

۳۸

□ س. کدام تفکیک کننده بهینه است؟



ماشین‌های بردار پشتیبان

۳۹

- بیشینه‌سازی حاشیه. مناسب از لحاظ نظری و در عمل.
- تنها بردارهای پشتیبان اهمیت دارند؛ نمونه‌های دیگر قابل صرف نظر هستند.
- ماشین‌های بردار پشتیبان، تفکیک کننده با بیشترین حاشیه را پیدا می‌کنند.

MIRA

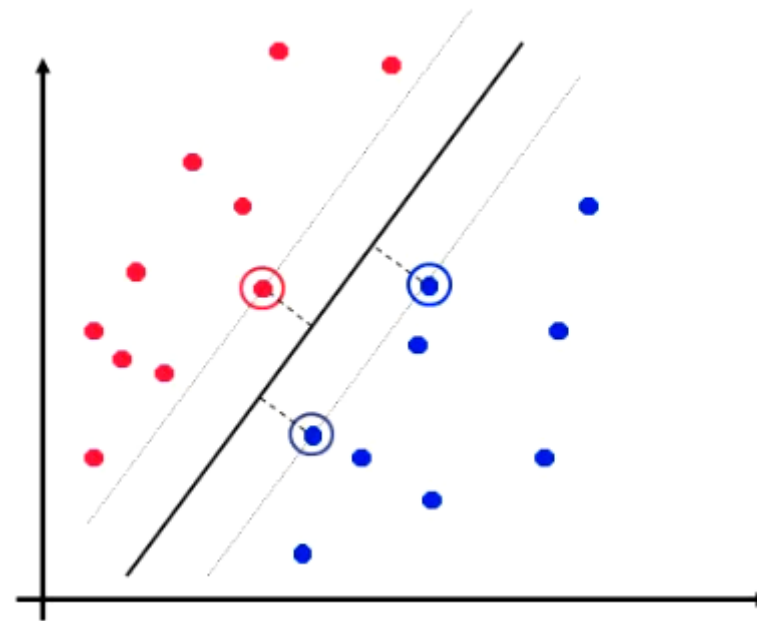
$$\min_w \frac{1}{2} \|w - w'\|^2$$

$$w_{y^*} \cdot f(x_i) \geq w_y \cdot f(x_i) + 1$$

SVM

$$\min_w \frac{1}{2} \|w\|^2$$

$$\forall i, y \quad w_{y^*} \cdot f(x_i) \geq w_y \cdot f(x_i) + 1$$



کلاس بندی: مقایسه

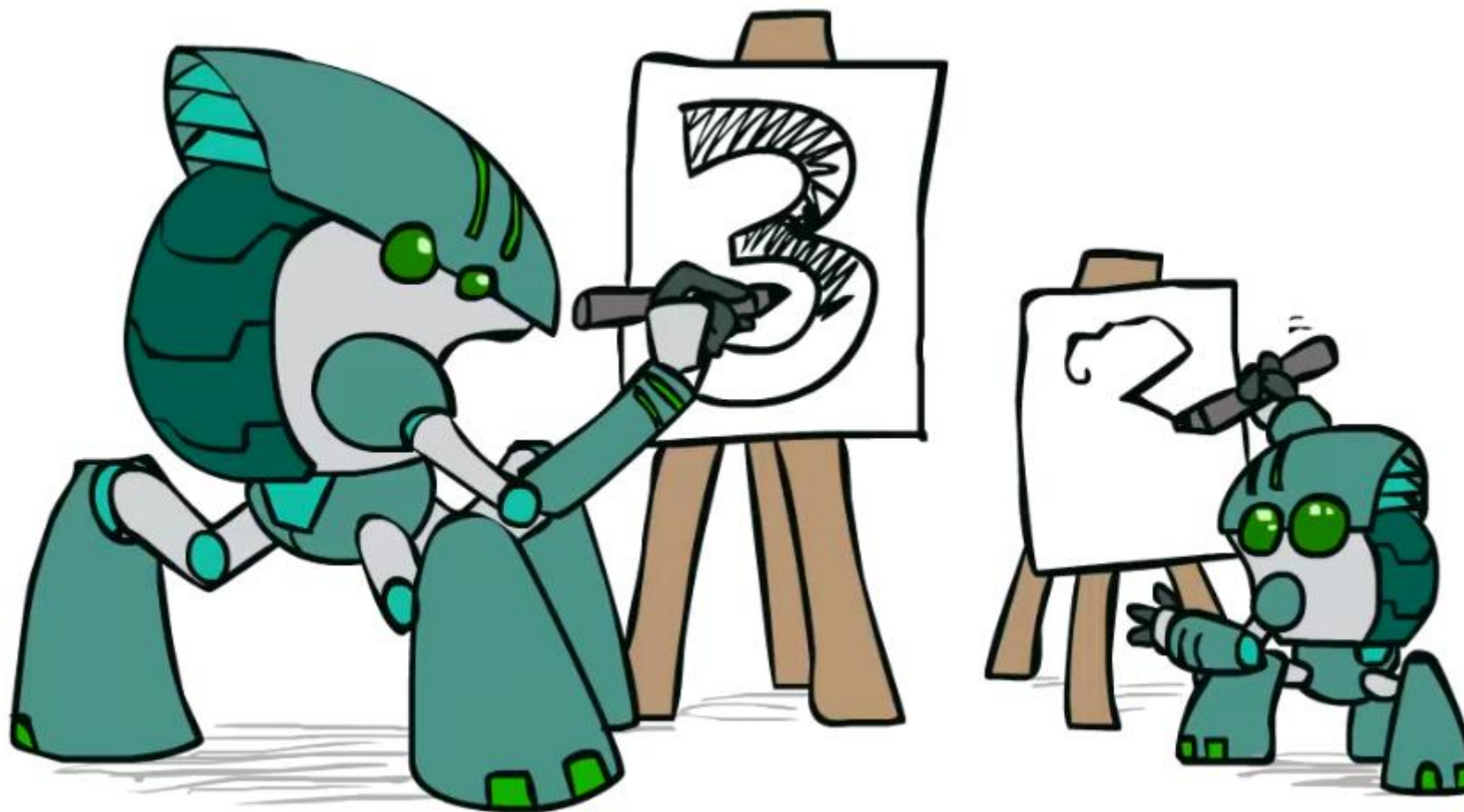
۴۰

□ بیز ساده.

- یک مدل بر روی مجموعه‌ی آموزشی ایجاد می‌کند
- احتمال تعلق یک نمونه به هر کلاس را پیش‌بینی می‌کند
- فرض قوی درباره‌ی استقلال ویژگی‌ها
- یک گذر بر روی داده‌ها (شمارش)

□ پرسپترون‌ها / میرا.

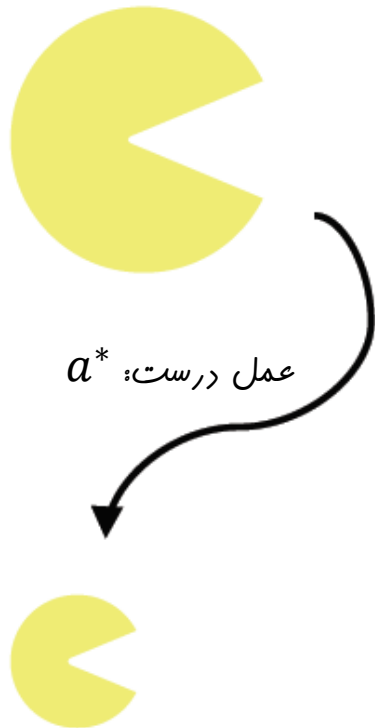
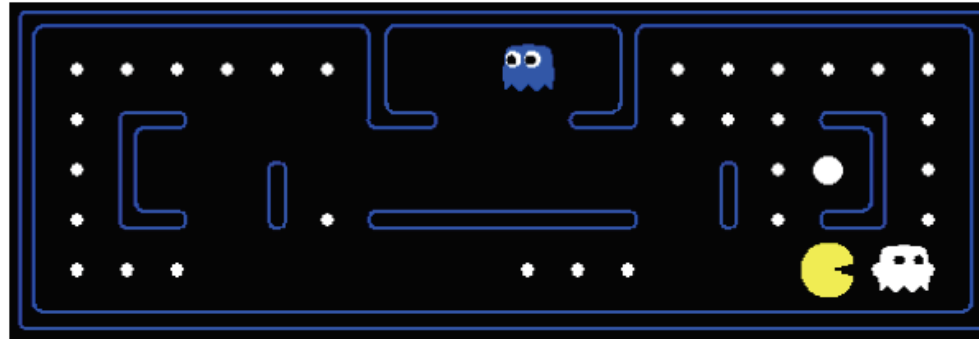
- فرضیات کمتر درباره‌ی داده‌ها
- یادگیری مبتنی بر خطا
- چندین گذر بر روی داده‌ها (پیش‌بینی)
- اغلب دقیق‌تر



کارآموزی برای پکمن!

۴۲

□ نمونه‌ها همان حالت‌های محیط هستند.



عمل درست: a^*

$$\forall a \neq a^* \\ w \cdot f(a^*) > w \cdot f(a)$$

□ نامزدها زوج مرتب‌های (s, a) هستند.

□ عمل‌های «درست»: عمل‌های انجام شده به وسیله فرد خبره

□ ویژگی‌ها بر روی زوج‌های مرتب (s, a) تعریف می‌شوند: $f(s, a)$

□ امتیاز یک حالت q مانند (s, a) به صورت زیر محاسبه می‌شود:

$$w \cdot f(s, a)$$

کارآموزی: اجرای نمایش