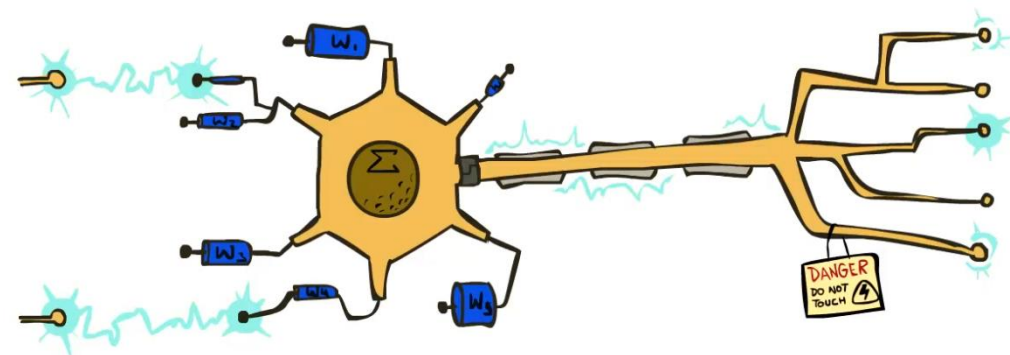
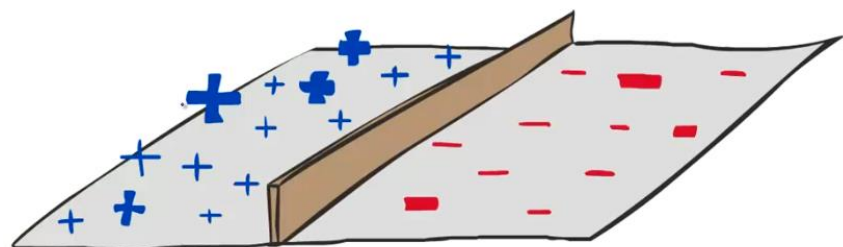


دسته‌بندی خطی

سید ناصر رضوی www.snrazavi.ir

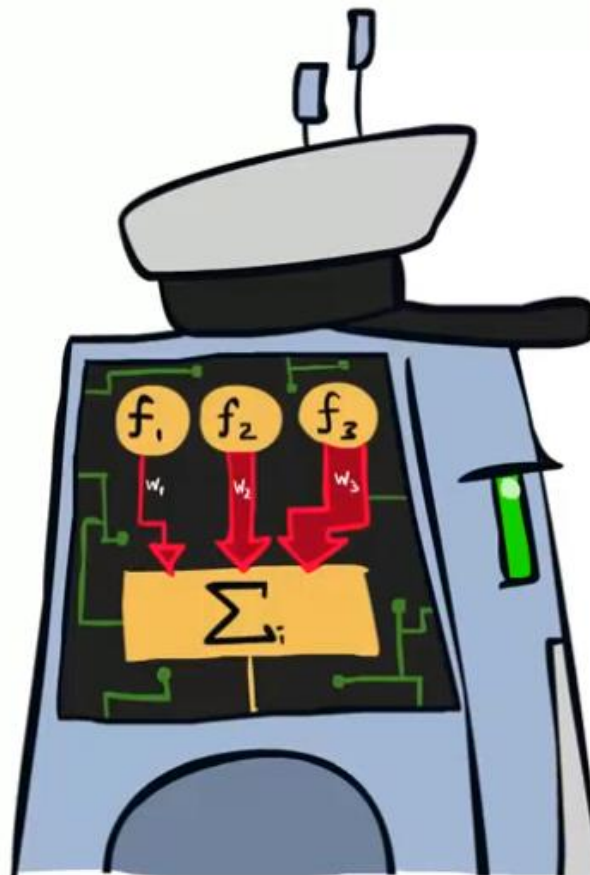
۱۳۹۶



- دسته‌بندی خطی
- الگوریتم یادگیری پرسپترون
- روش‌های بهبود پرسپترون
- الگوریتم یادگیری MIRA
- ماشین بردار پشتیبان (SVM)
- دسته‌بندی با بیشترین حاشیه

دسته‌بندی خطی

۳



بردار ویژگی

۴

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



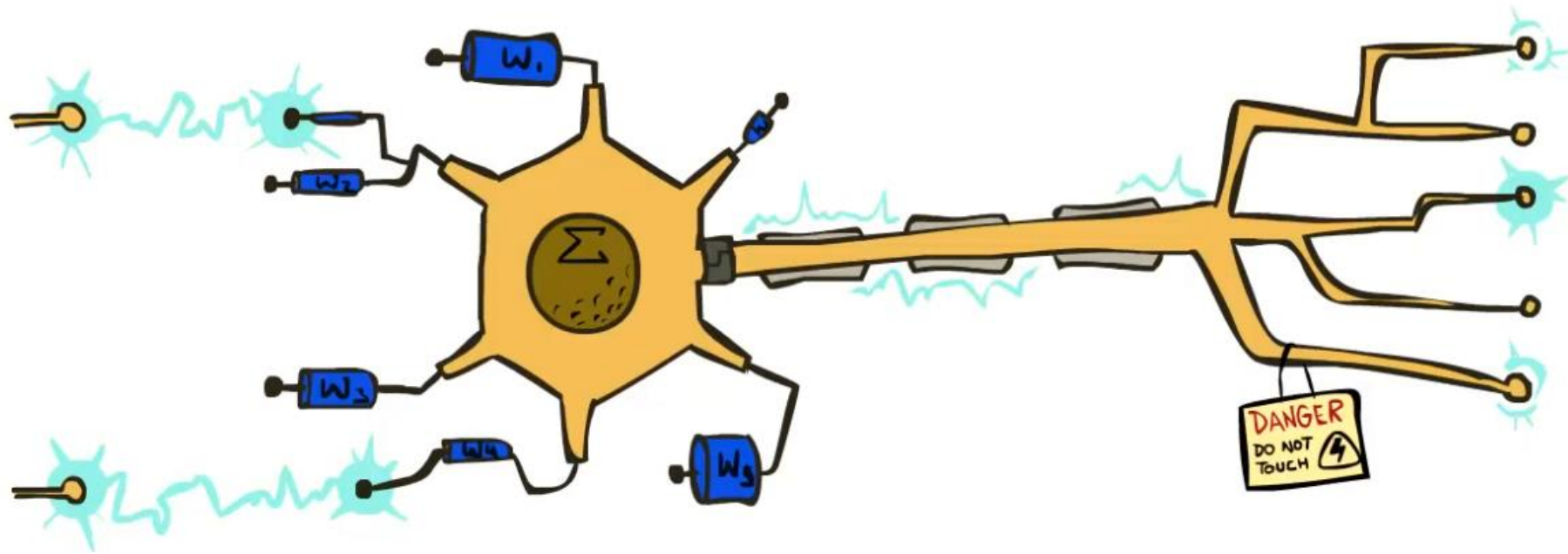
PIXEL (7 , 12)	: 1
PIXEL (7, 13)	: 0
...	
NUM_LOOPS	: 1
...	



"2"

پرسپترون‌ها

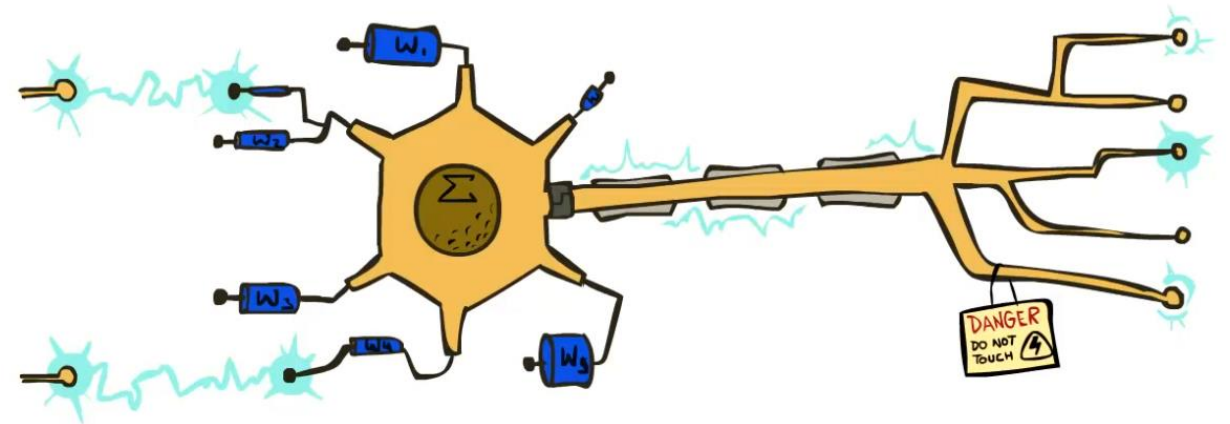
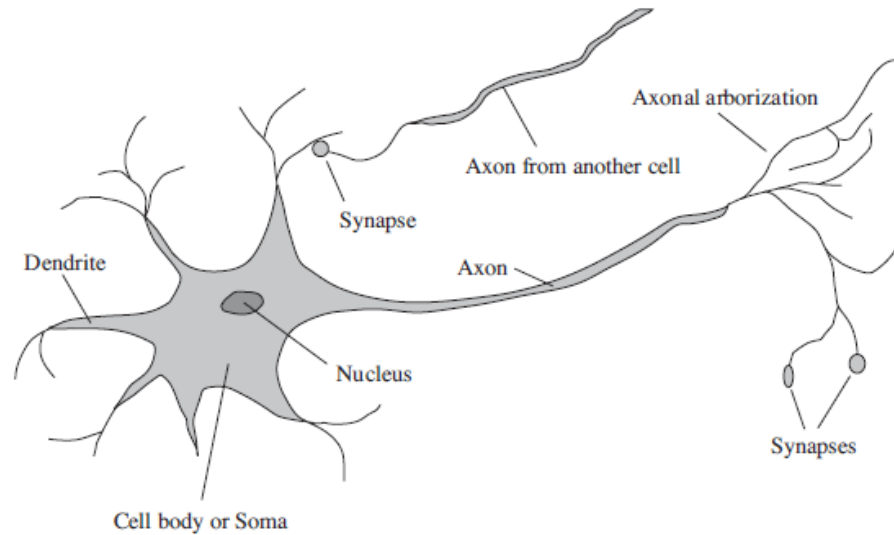
۵



کمی زیست‌شناسی (به زبان ساده)

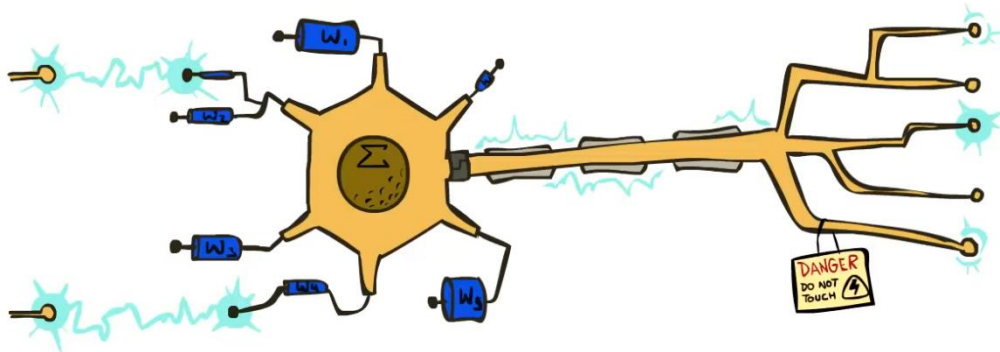
۶

□ یک الگو برداری بسیار ضعیف. نورون‌های انسان!



دسته‌بندی خطی

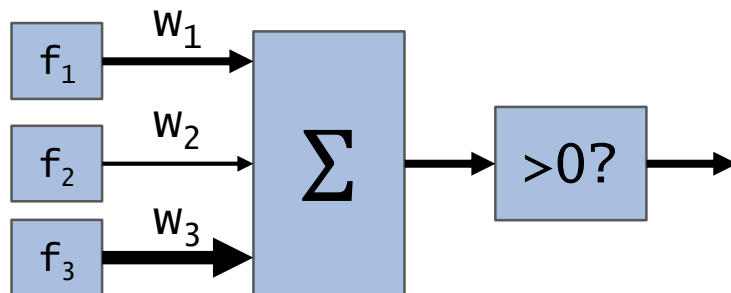
۷



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

ضرب داخلی

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



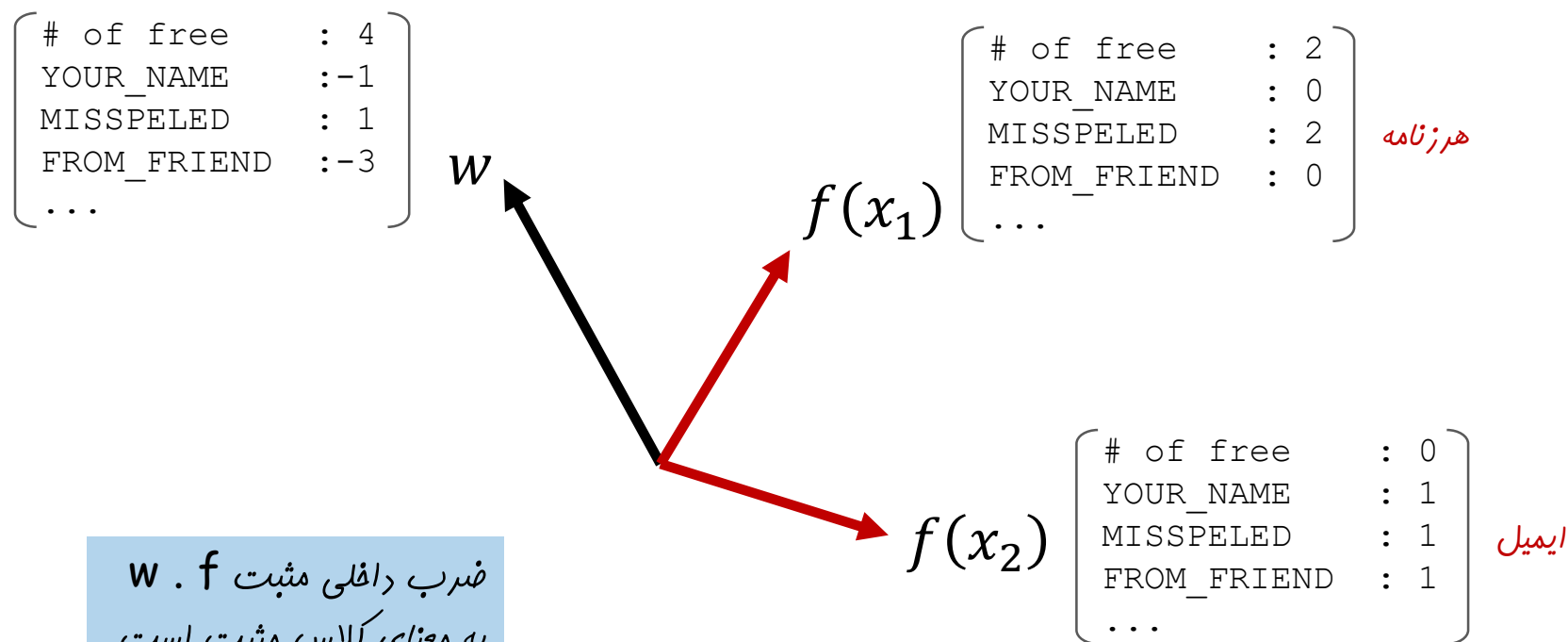
□ خروجی.

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

وزن‌ها

۸

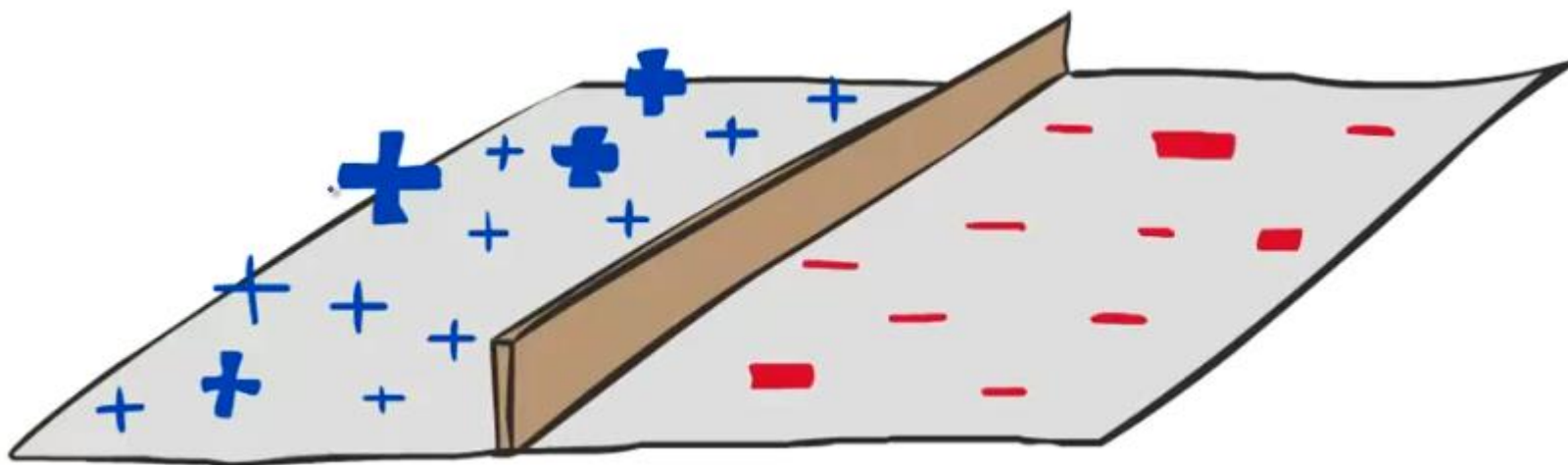
- حالت دودویی. مقایسه ویژگی‌ها با یک بردار وزن.
- یادگیری. تنظیم مقدار بردار وزن با توجه به نمونه‌های آموزشی.



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

قواعد تصدیق‌گیری

۹



قاعده‌ی تصمیم‌گیری دودویی

۱۰

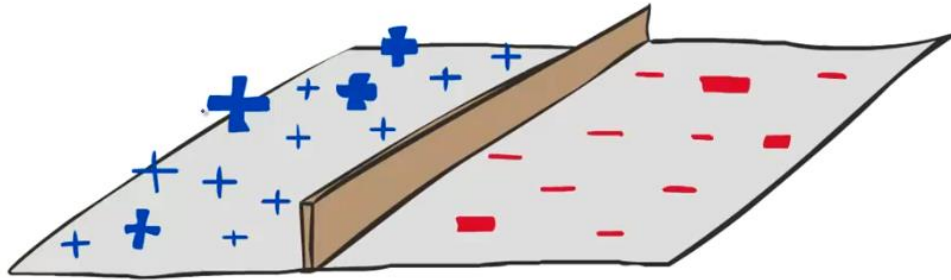
□ در فضای بردارهای ویژگی.

□ هر نمونه بیانگر یک نقطه است.

□ هر بردار وزن بیانگر یک ابرصفحه است.

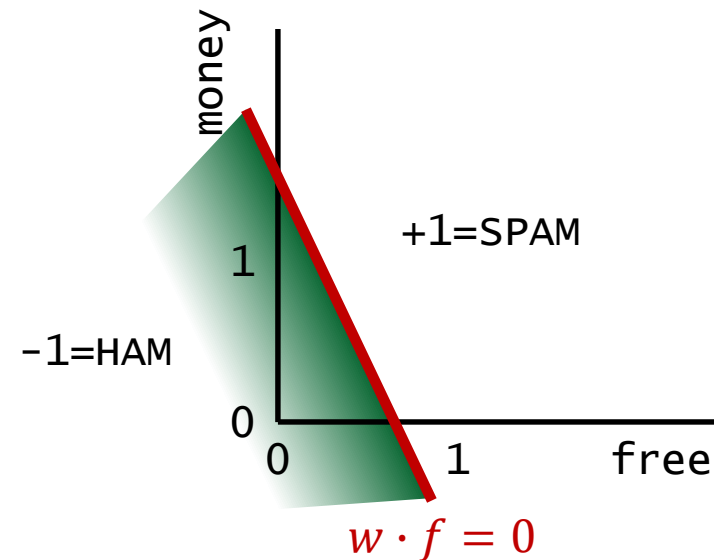
□ یک طرف ابرصفحه به کلاس مثبت اشاره دارد.

□ طرف دیگر آن به کلاس منفی اشاره دارد.



w

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



به روز رسانی وزن‌ها

۱۱



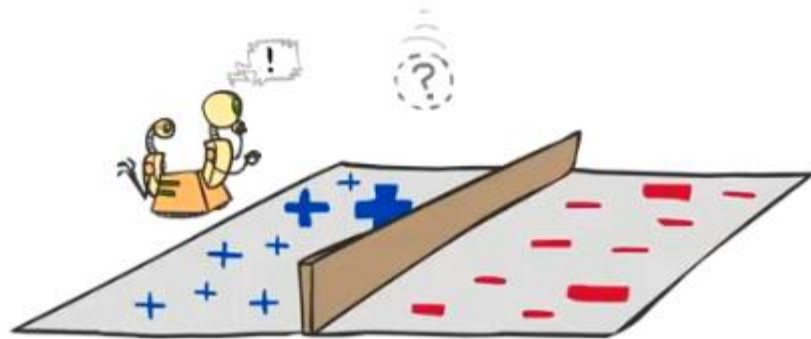
یادگیری: پرسپترون دودویی

۱۲

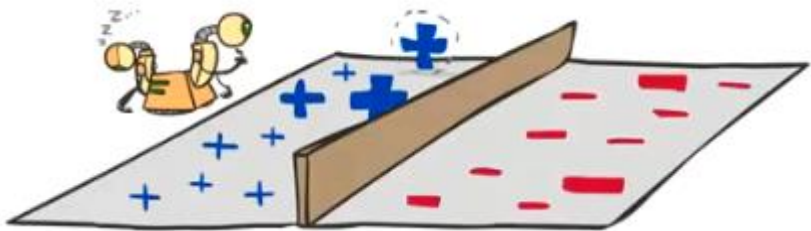
□ با بردار وزن صفر شروع کن.

□ به ازای هر نمونه آموزشی:

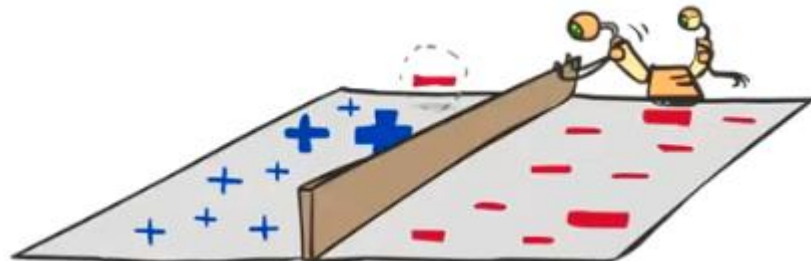
□ نمونه را با وزن‌های فعلی دسته‌بندی کن.



□ اگر دسته‌بندی درست، بدون تغییر!

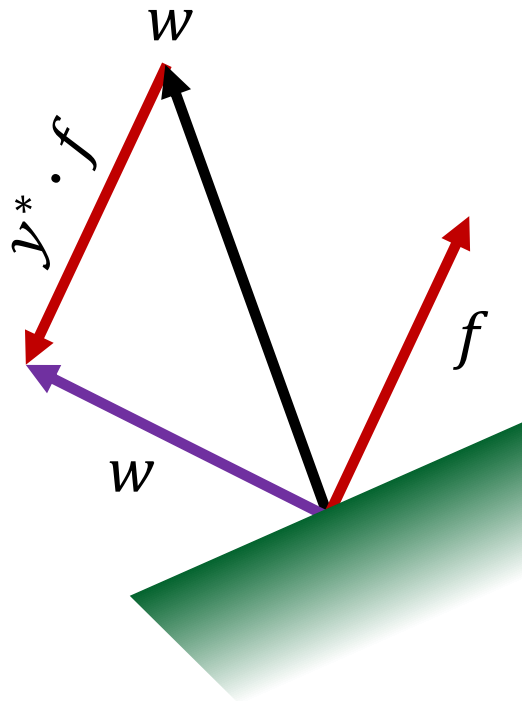


□ اگر دسته‌بندی نادرست، تغییر وزن‌ها!



یادگیری: پرسپترون دودویی

۱۳



□ با بردار وزن صفر شروع کن.

□ به ازای هر نمونه آموزشی:

□ نمونه را با وزن‌های فعلی دسته‌بندی کن.

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

□ اگر دسته‌بندی درست، بدون تغییر!

□ اگر دسته‌بندی نادرست، تغییر وزن‌ها!

■ با افزودن یا کم کردن بردار ویژگی.

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

یادگیری: پرسپترون دودویی

۱۴

□ با بردار وزن صفر شروع کن.

□ به ازای هر نمونه آموزشی:

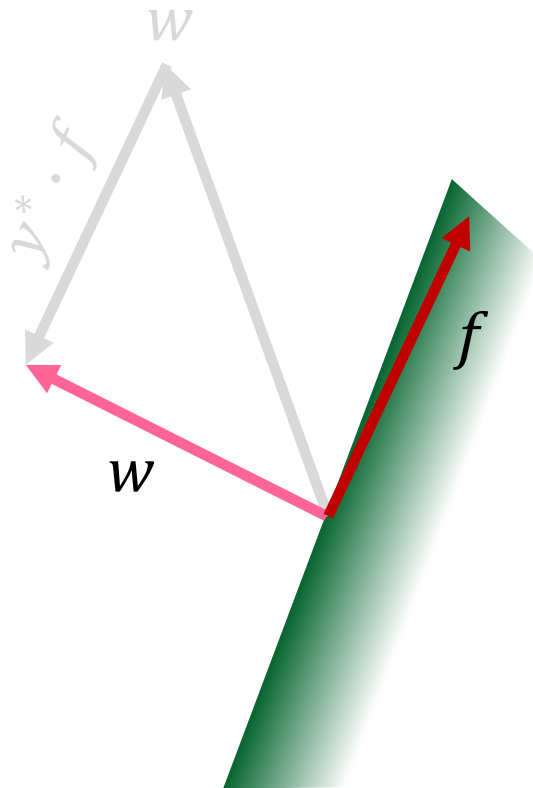
□ نمونه را با وزن‌های فعلی دسته‌بندی کن.

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

□ اگر دسته‌بندی درست، بدون تغییر!

□ اگر دسته‌بندی نادرست، تغییر وزن‌ها!

■ با افزودن یا کم کردن بردار ویژگی.

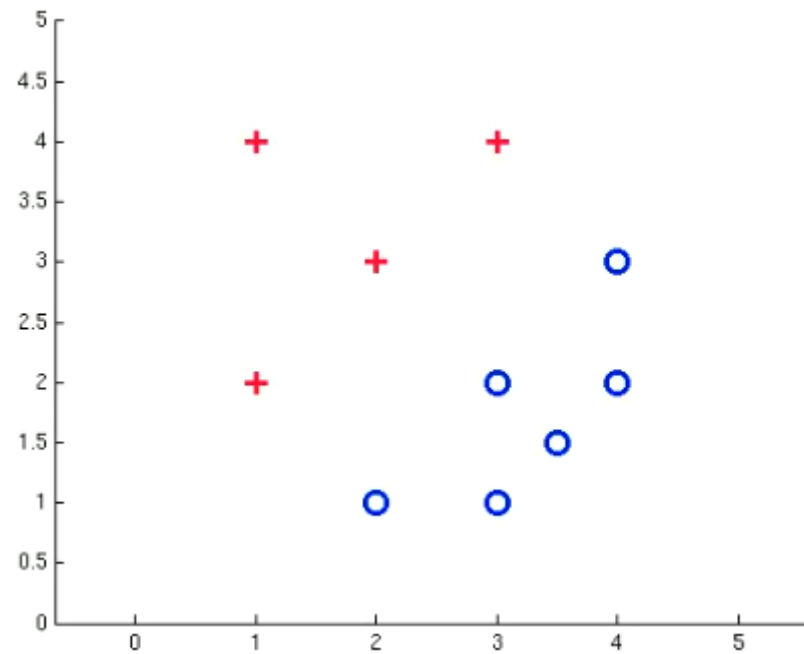


$$w = w + y^* \cdot f$$

پرسپترون: اجرای نمایشی

۱۵

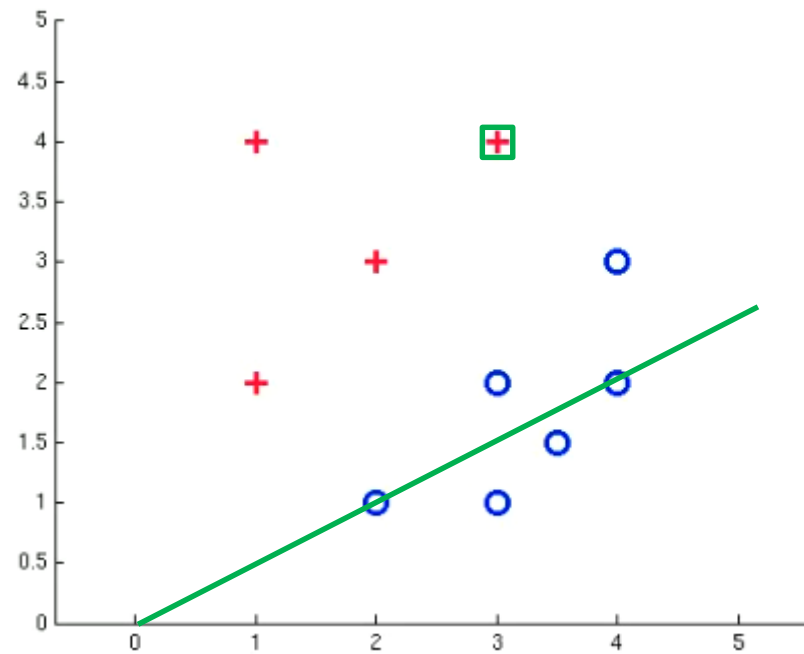
□ جداپذیر خطی.



پرسپترون: اجرای نمایی

۱۶

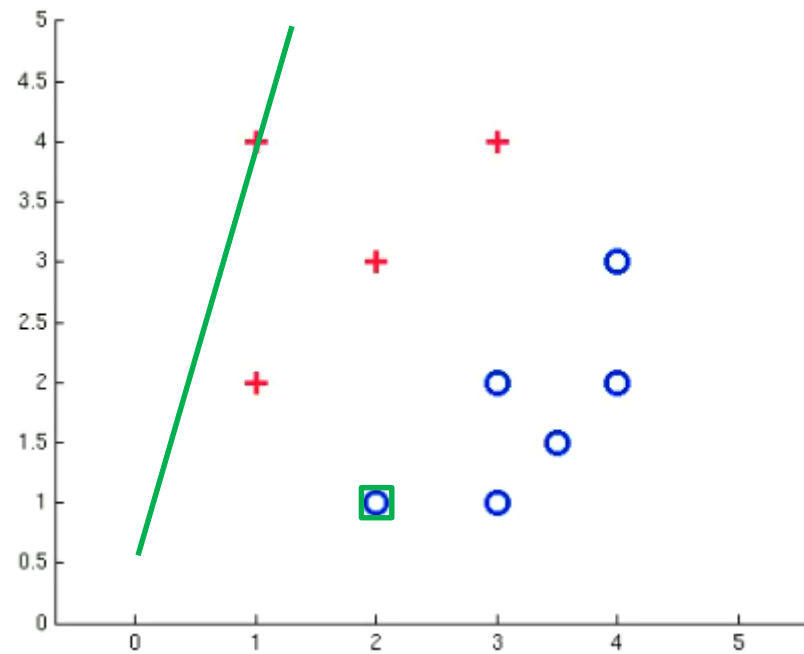
□ جداپذیر خطی.



پرسپترون: اجرای نمایشی

۱۷

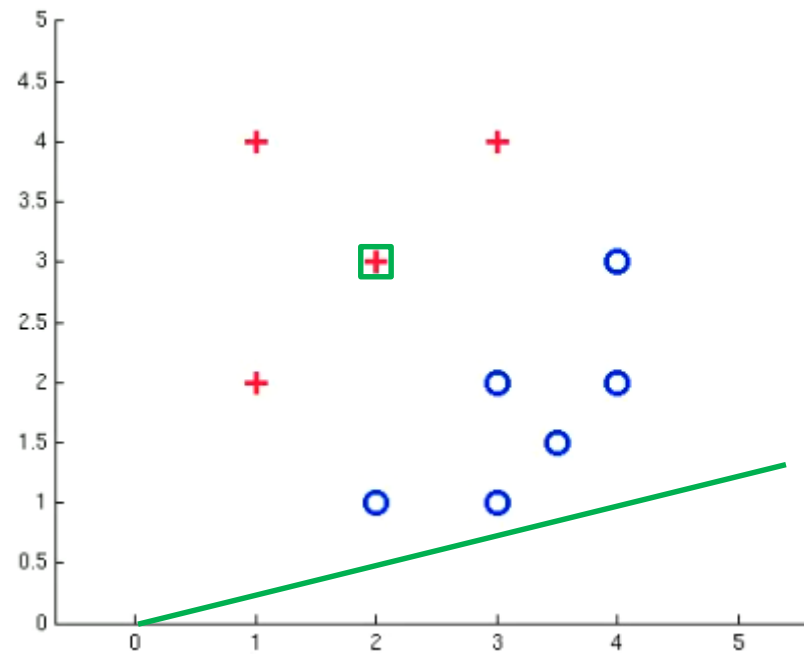
□ جداپذیر خطی.



پرسپترون: اجرای نمایشی

۱۸

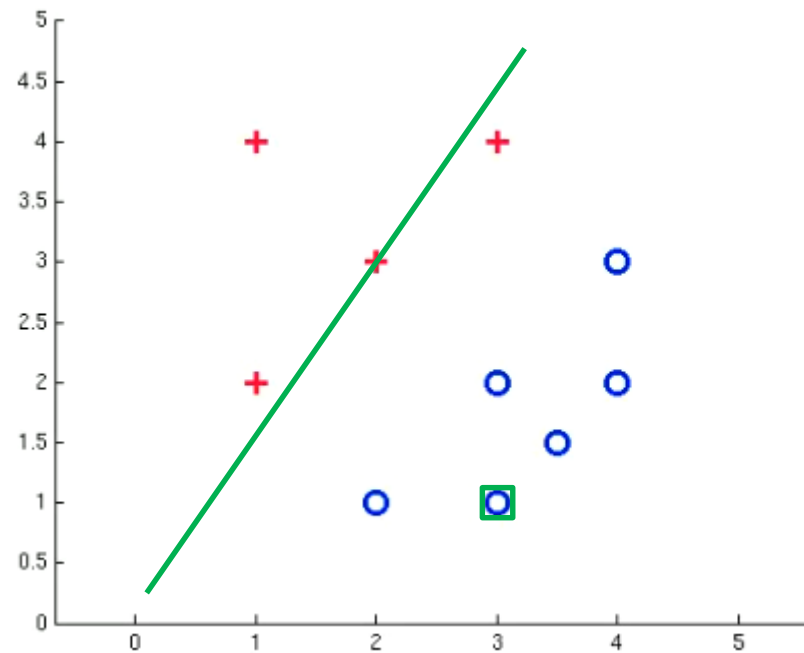
□ جداپذیر خطی.



پرسپترون: اجرای نمایشی

۱۹

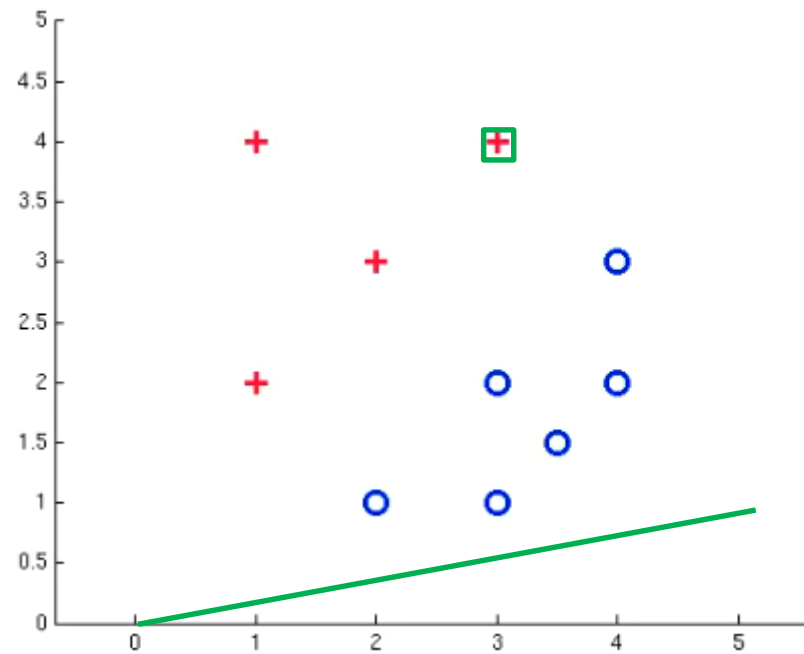
□ جداپذیر خطی.



پرسپترون: اجرای نمایشی

۲۰

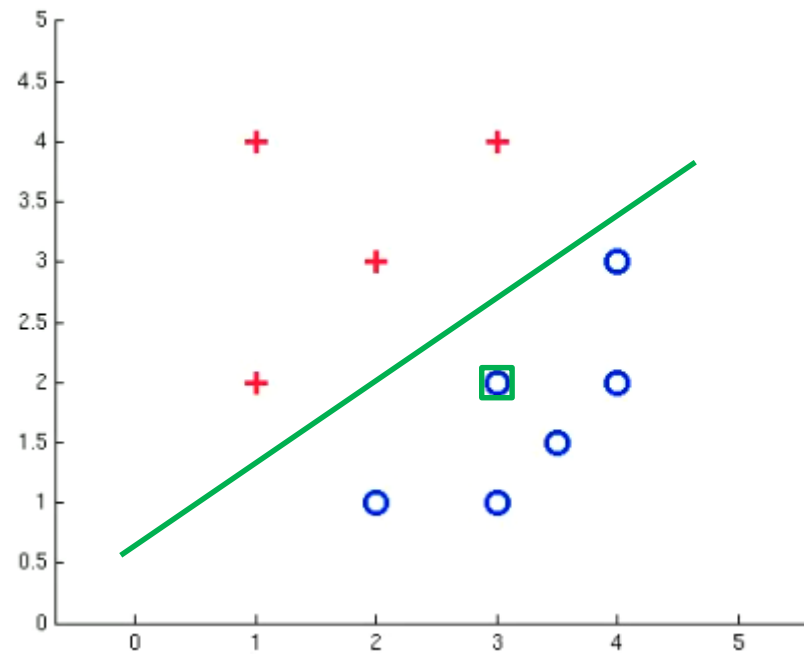
□ جداپذیر خطی.



پرسپترون: اجرای نمایشی

۲۱

□ جداپذیر خطی.



قاعده تصمیم‌گیری برای چندکلاس

۲۲

□ در صورت وجود چند کلاس.

□ به ازای هر کلاس y یک بردار وزن.

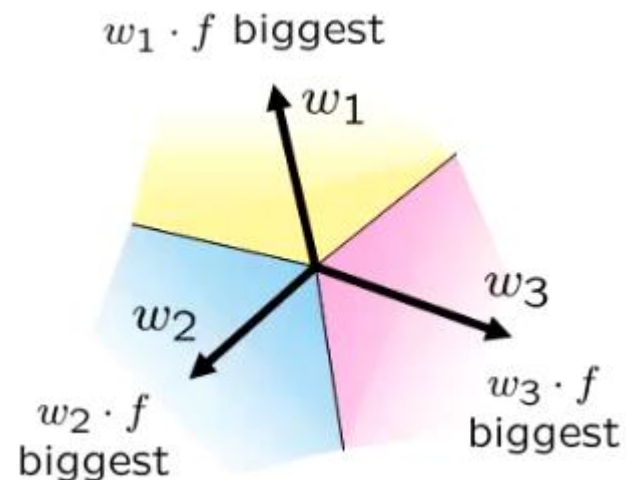
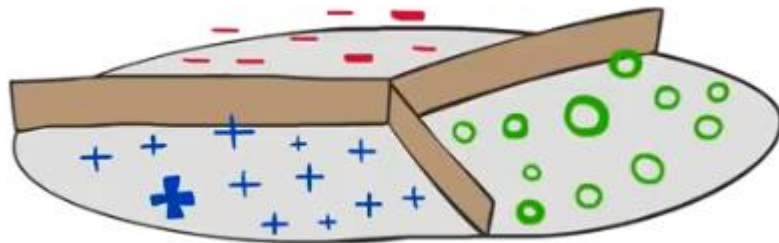
w_y

□ محاسبه امتیاز مربوط به کلاس y .

$w_y \cdot f(x)$

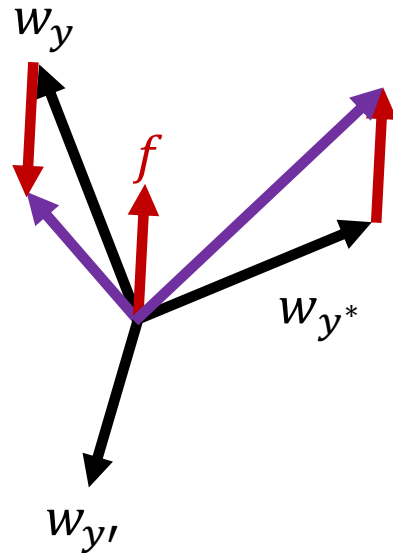
□ پیش‌بینی. برنده کلاس با بیشترین امتیاز.

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



یادگیری: پرسپترون چند کلاسی

۲۳



□ با بردار وزن صفر شروع کن.

□ به ازای هر نمونه آموزشی:

□ نمونه را با وزن‌های فعلی دسته‌بندی کن.

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

□ اگر دسته‌بندی درست، بدون تغییر!

□ اگر دسته‌بندی نادرست، تغییر وزن‌ها!

■ کاهش امتیاز کلاس نادرست، افزایش امتیاز کلاس درست

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

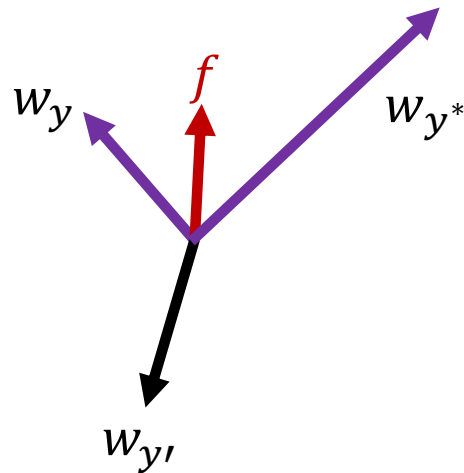
یادگیری: پرسپترون چند کلاسی

۲۴

□ با بردار وزن صفر شروع کن.

□ به ازای هر نمونه آموزشی:

□ نمونه را با وزنهای فعلی دسته‌بندی کن.



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

□ اگر دسته‌بندی درست، بدون تغییر!

□ اگر دسته‌بندی نادرست، تغییر وزن‌ها!

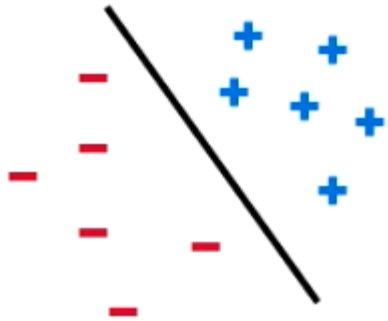
■ کاهش امتیاز کلاس نادرست، افزایش امتیاز کلاس درست

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

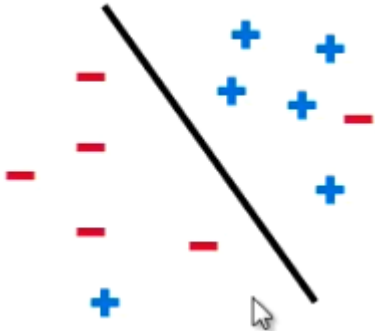
پرسپترون: ویژگی‌ها

۲۵

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).

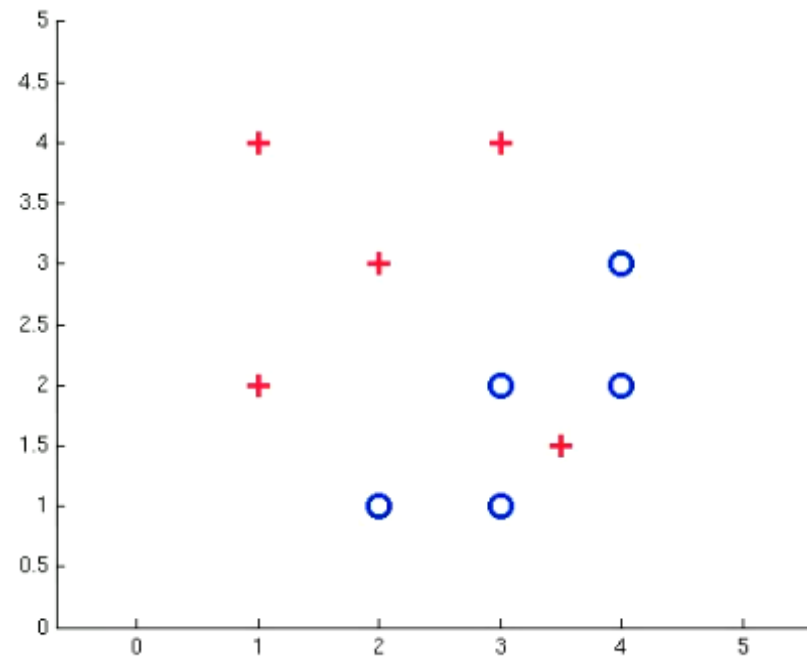
□ کران بالای خطا. حداکثر تعداد خطاها (در حالت دودویی) با حاشیه و یا درجه تفکیک پذیری مرتبط است.

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

پرسپترون: مثال

۲۶

□ داده‌های تفکیک‌ناپذیر خطی.

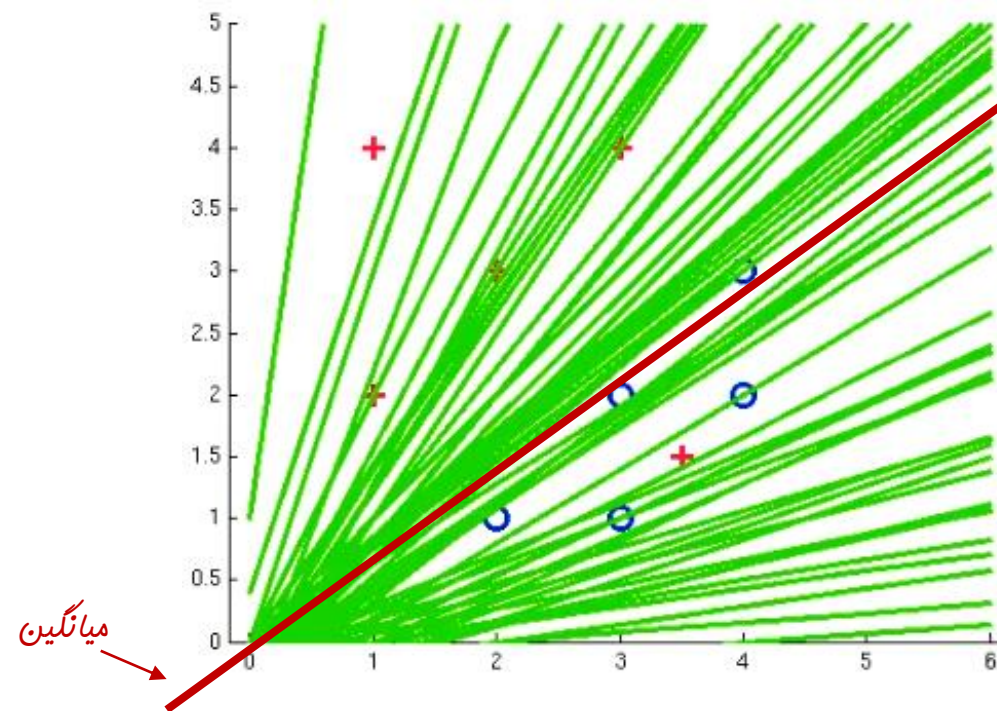


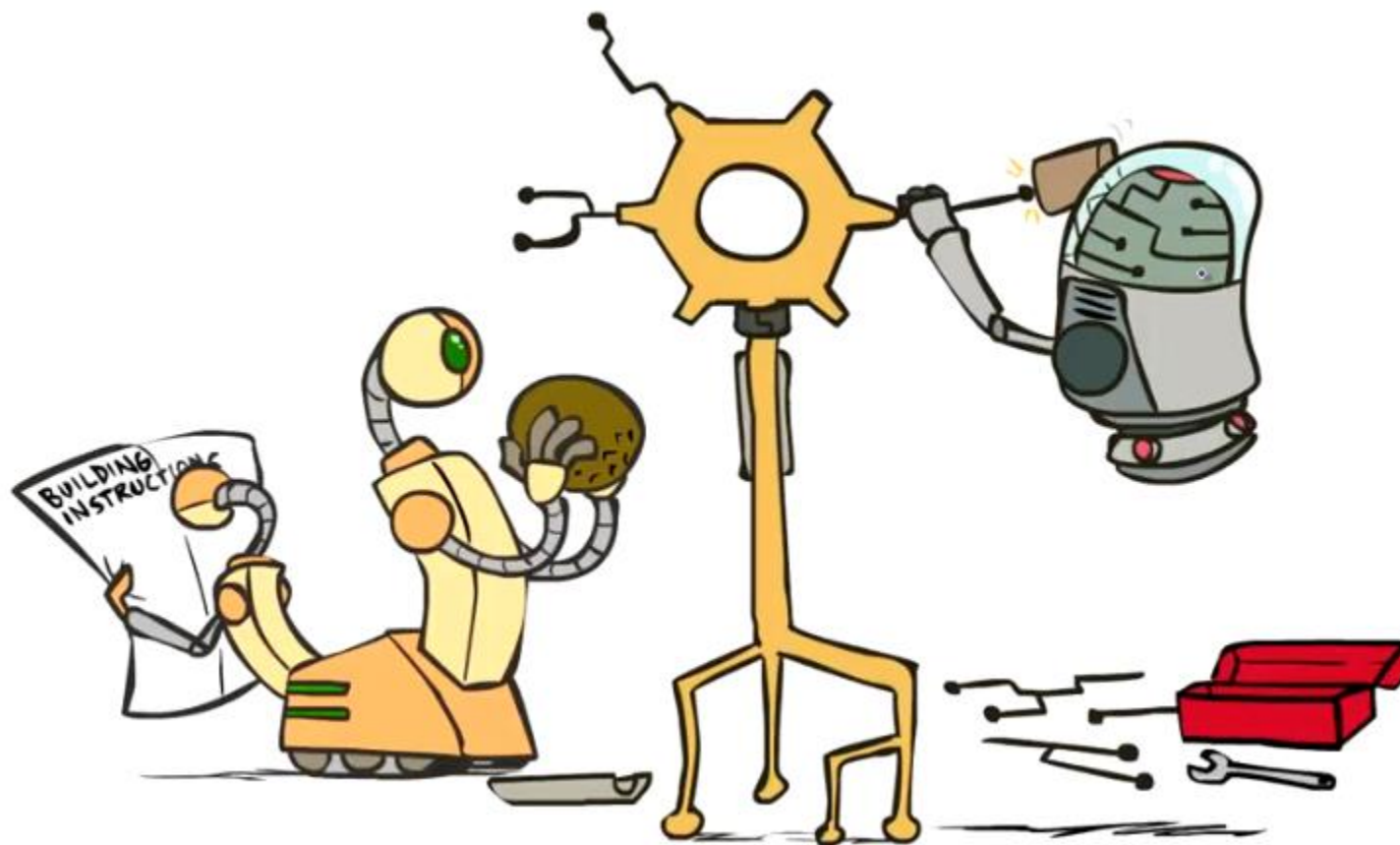
پرسپترون: مثال

۲۷

□ داده‌های تفکیک‌ناپذیر خطی.

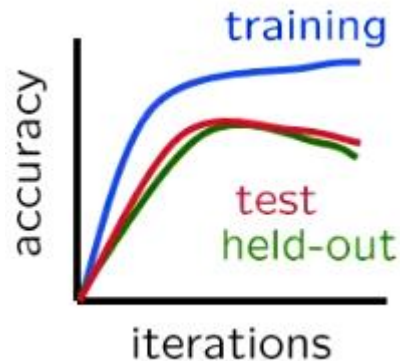
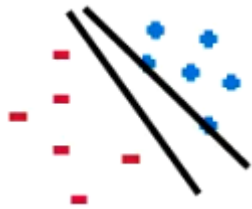
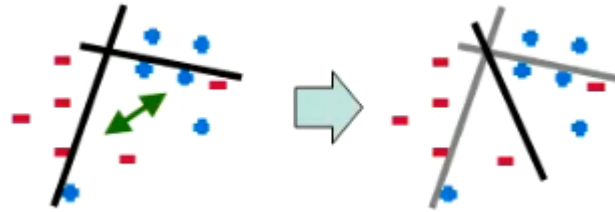
فرضیه‌های به دست آمده از ۱۰۰ اجرا





مشکلات پرسپترون

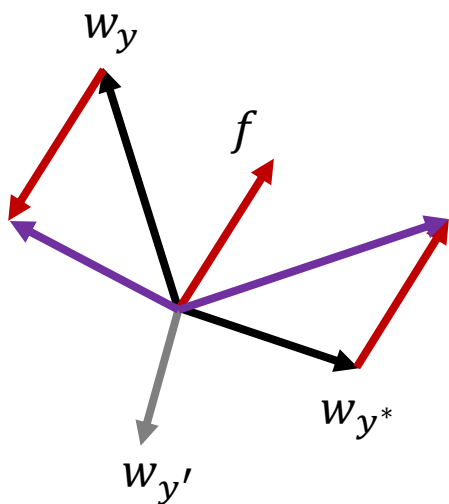
۲۹



- **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.

بهبود پرسپترون

۳۰



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

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

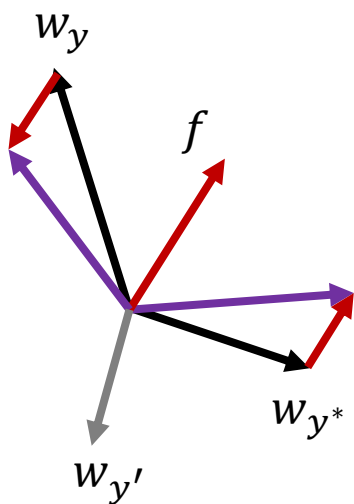
□ ایده. به روز رسانی وزن‌ها برای کاهش اثر این مشکلات.

□ $MIRA^*$. 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 مداخلی

۳۲

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

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



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

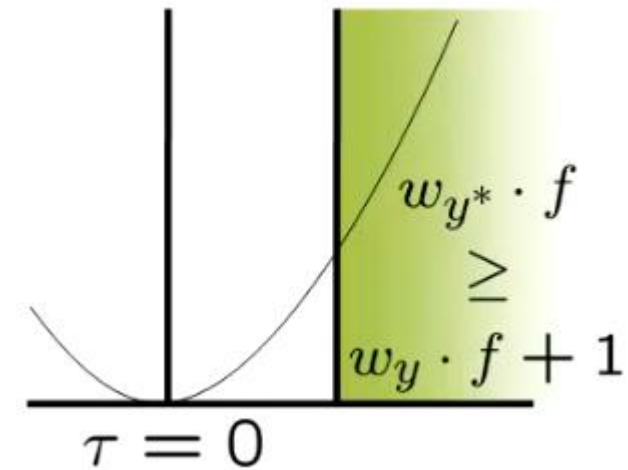
$$w_{y^*} \cdot f \geq w_y \cdot f + 1$$



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

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

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



حداکثر اندازه گام

۳۳

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

□ نمونه‌ها ممکن است به غلط برچسب گذاری شوند.

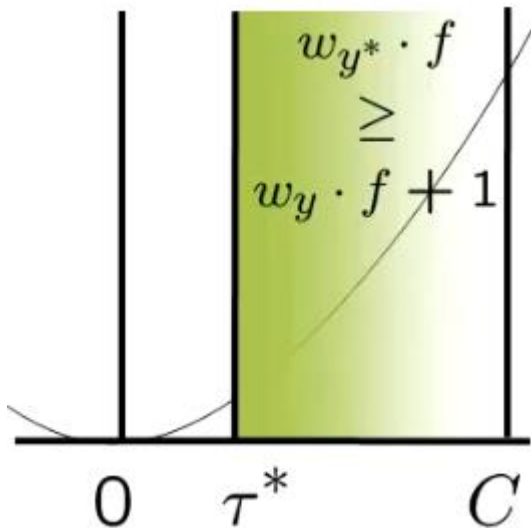
□ ممکن است به اندازه کافی ویژگی وجود نداشته باشد.

□ 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.

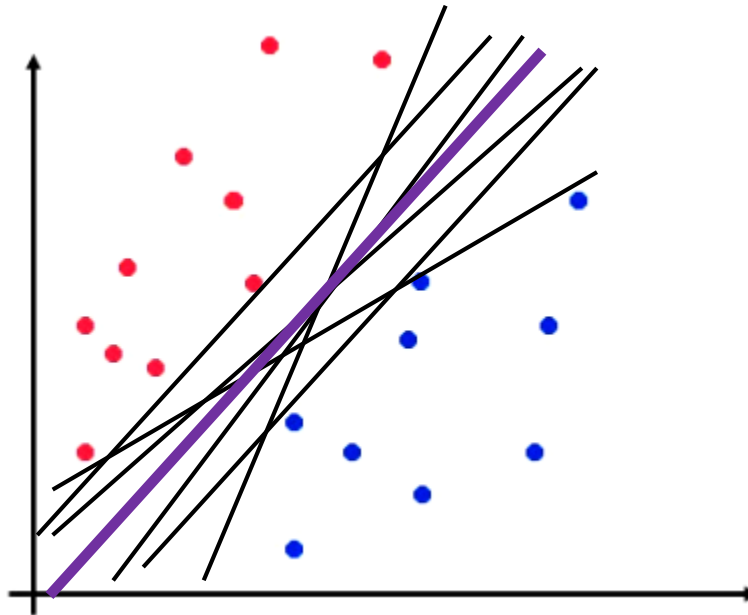
□ معمولاً بهتر است، به ویژه برای داده‌های دارای نویز.



تفکیک کننده‌های خطی

۳۴

□ س. کدام تفکیک کننده بهینه است؟



ماشین‌های بردار پشتیبان

۳۵

- 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

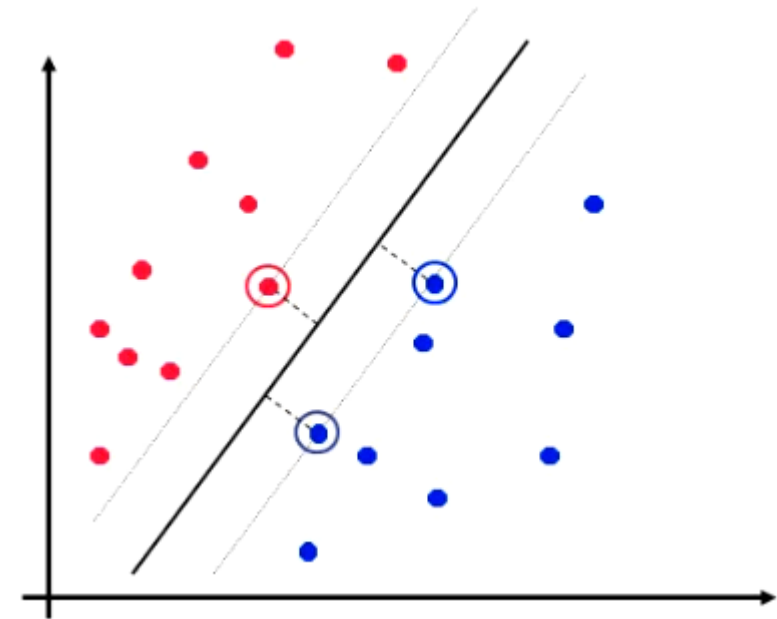
$$\min_w \frac{1}{2} \|w - w'\|^2$$

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

SVM

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

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



فهرست مطالب

- دسته‌بندی خطی
- روش‌های پارامتری و غیرپارامتری
- تابع هزینه (تابع زیان)

رویکرد پارامتری: دسته‌بندی خطی

۳۷

□ دسته‌بندی خطی.



$[32 \times 32 \times 3]$

برداری از اعداد در بازه $[0, 1]$

تصویر
پارامترها
 $f(\mathbf{x}, \mathbf{w})$



۱۰ عدد

هر عدد بیانگر امتیاز
مربوط به یک کلاس

رویکرد پارامتری: دسته‌بندی خطی

۳۸

□ دسته‌بندی خطی.

پارامترها، یا «وزن‌ها»

10×1 10×30.72 30.72×1 10×1



$[32 \times 32 \times 3]$

برداری از اعداد در بازه $[0, 1]$

$$f(x, W) = Wx (+b)$$

۱۰ عدد

هر عدد بیانگر امتیاز
مربوط به یک کلاس

دسته‌بندی خطی: مثال

۳۹

□ چهار ویژگی و سه کلاس [سگ، گربه، کشتی]

تبدیل تصویر به یک بردار ستونی



تصویر ورودی

0.2	-0.5	0.1	2.0
1.5	1.3	2.1	0.0
0.0	0.25	0.2	-0.3

W

56
231
24
2

x_i

+

1.1
3.2
-1.2

b

→

-96.8
437.9
61.95

$f(x_i; W, b)$

امتیاز گربه

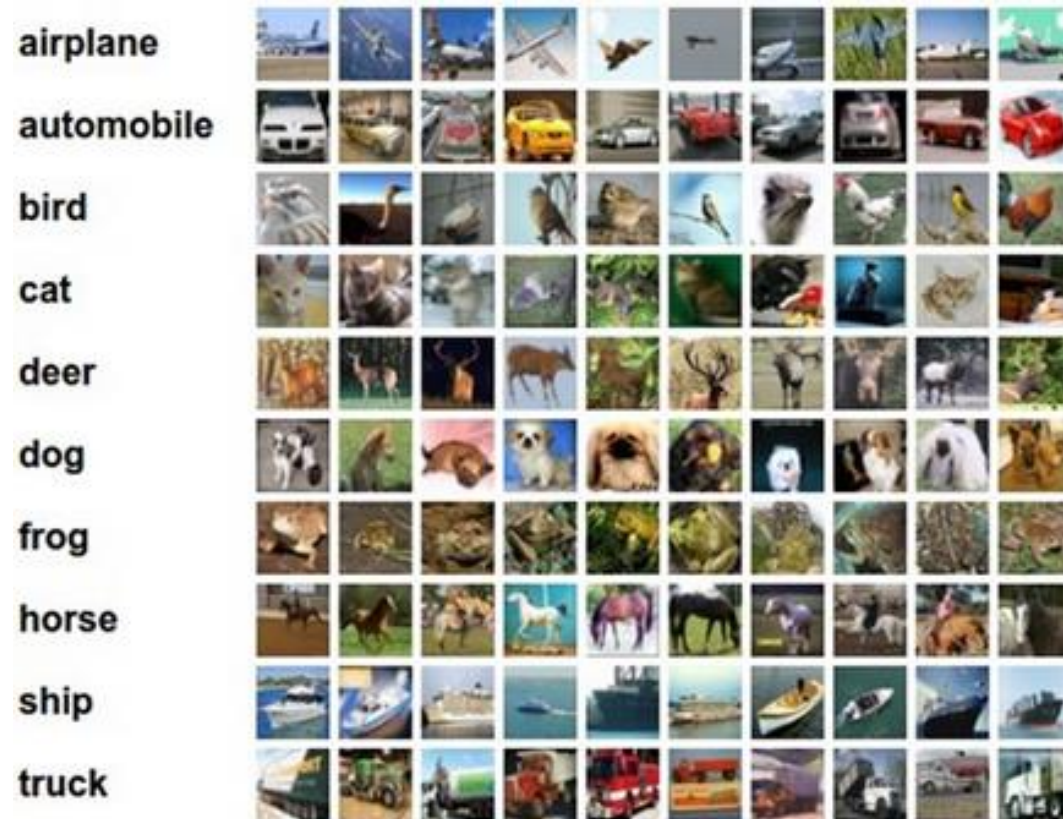
امتیاز سگ

امتیاز کشتی

تفسیر دسته‌بندی خطی

□ یک دسته‌بند خطی چه عملی انجام می‌دهد؟

$$f(x_i; W, b) = Wx_i + b$$



تفسیر دسته‌بندی خطی

□ یک دسته‌بند خطی چه عملی انجام می‌دهد؟

$$f(x_i; W, b) = Wx_i + b$$

نمونه‌ای از وزن‌های آموزش داده شده با استفاده از
یک دسته‌بند خطی بر روی مجموعه CIFAR-10

airplane

automobile

bird

cat

deer

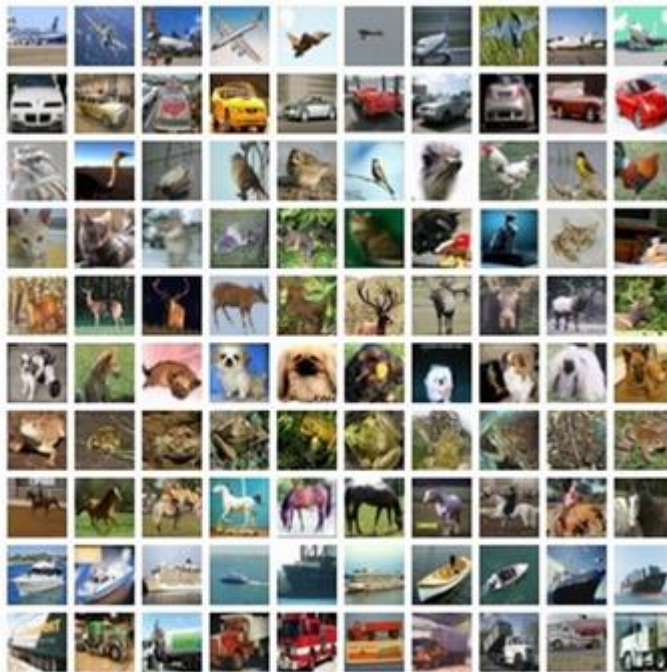
dog

frog

horse

ship

truck



plane

car

bird

cat

deer

dog

frog

horse

ship

truck



تفسیر دسته‌بندی خطی

۴۲

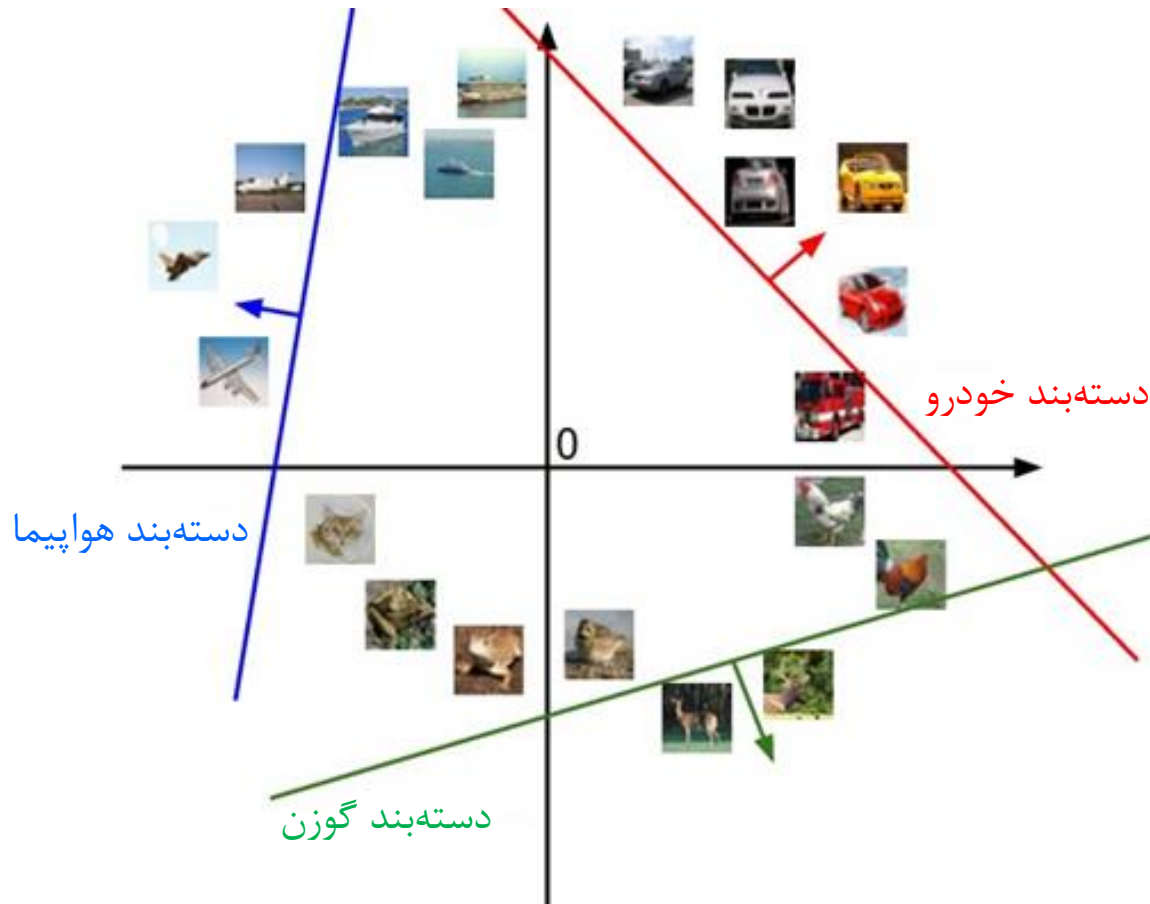
□ یک دسته‌بند خطی چه عملی انجام می‌دهد؟

$$f(x_i; W, b) = Wx_i + b$$



$[32 \times 32 \times 3]$

برداری از اعداد در بازه $[0, 1]$



دسته‌بندی خطی: تابع امتیاز

۴۳

□ امتیاز کلاس‌های مختلف برای سه تصویر ورودی با استفاده از وزن‌های تصادفی.



-3.45	-0.51	3.42	هواپیما
-8.87	6.04	4.64	خودرو
0.09	5.31	2.65	پرند
2.90	-4.22	5.10	گربه
4.48	-4.19	2.64	گوزن
8.02	3.58	5.55	سگ
3.78	4.49	-4.34	قورباغه
1.06	-4.37	-1.50	اسب
-0.36	-2.09	-4.79	کشتی
-0.72	-2.93	6.14	کامیون

$$f(x_i; W, b) = Wx_i + b$$

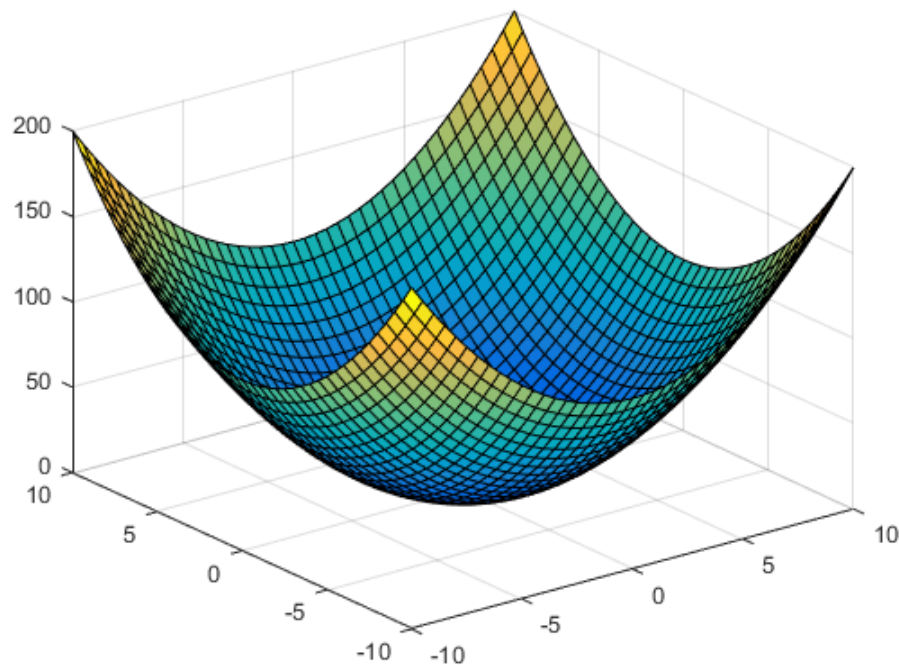


$[32 \times 32 \times 3]$

برداری از اعداد در بازه $[0, 1]$

در ادامه ...

۴۴



□ تابع هزینه [تابع زیان].

□ بیانگر میزان **بد بودن** ماتریس وزن به صورت یک مقدار عددی!

□ بهینه‌سازی.

□ یافتن یک ماتریس وزن به گونه‌ای که تابع هزینه را **کمینه** نماید.

□ دسته‌بندی غیرخطی.