ماشین‌های بردار پشتیبان

سید ناصر رضوی www.snrazavi.ir

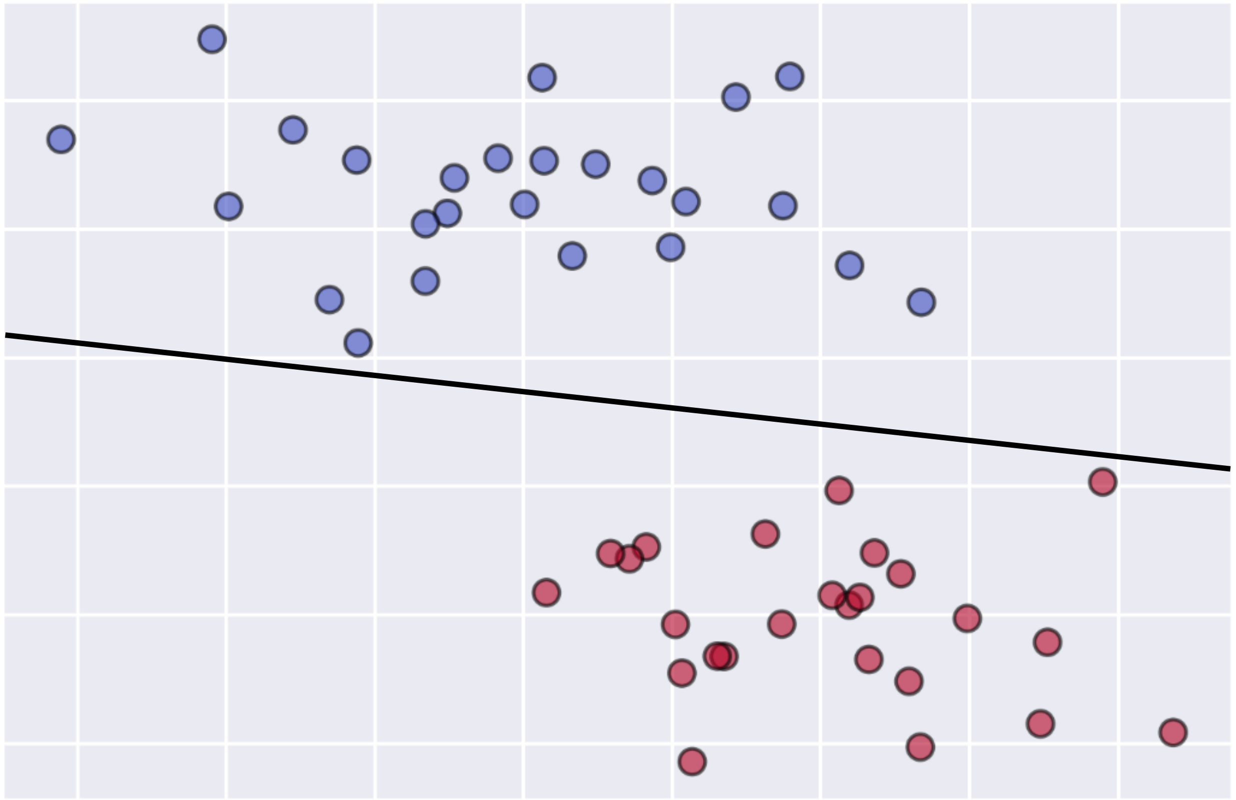
۱۳۹۷

فهرست مطالب

معرفی

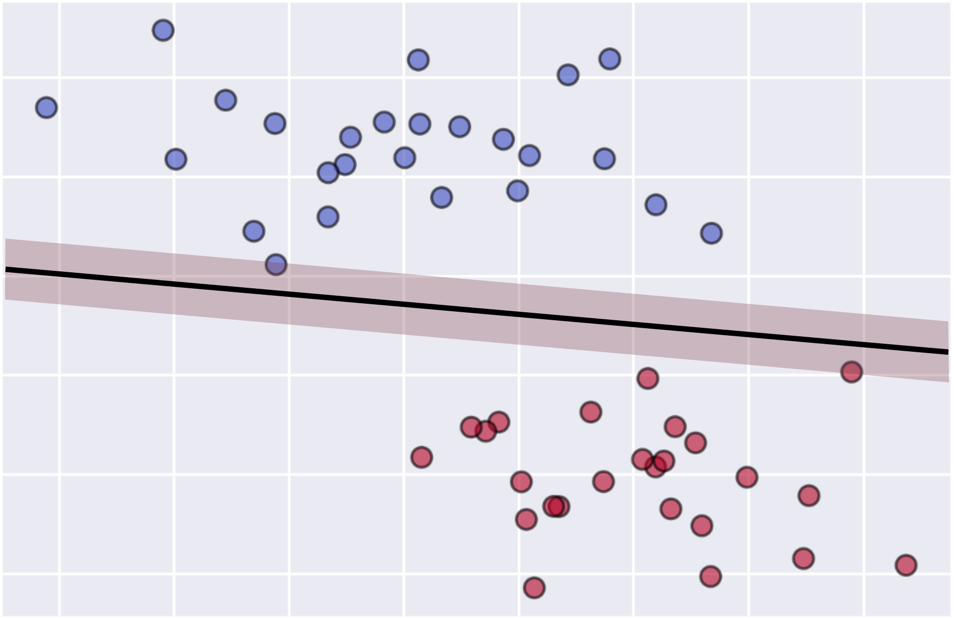
- ماشین‌های بردار پشتیبان [وپنیک، ۱۹۹۲]
  - یکی از پرطرفدارترین الگوریتم‌های یادگیری ماشین!
  - جداسازی بهتر داده‌ها نسبت به سایر روش‌های یادگیری ماشین (مسائل دسته‌بندی)!
  - استفاده از آن نسبتاً آسان است!
  - استفاده از ترفند کرنل:
    - دسته‌بندی، رگرسیون، تخمین توزیع، دسته‌بندی تک دسته‌ای و …

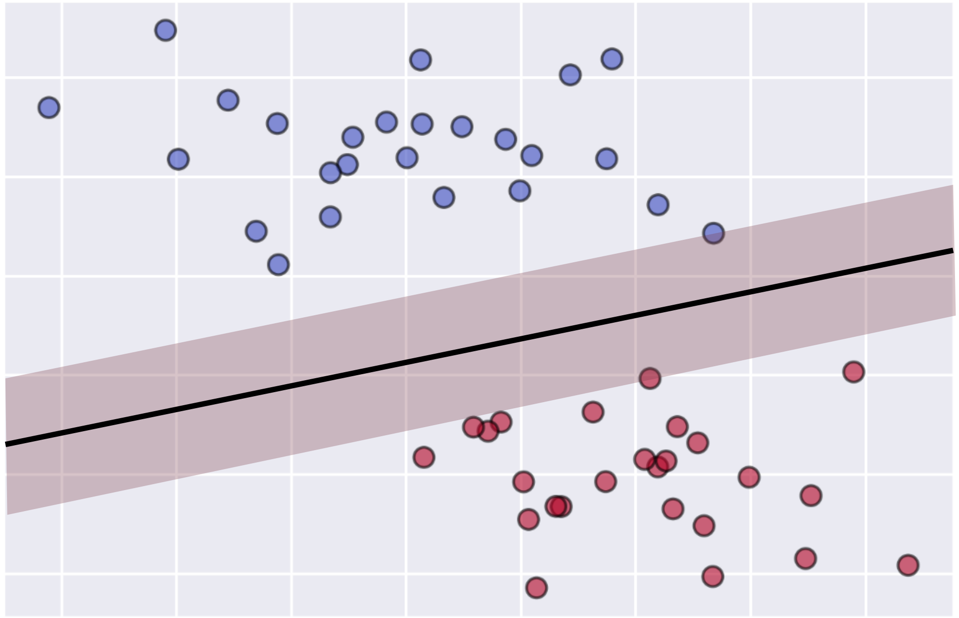
انگیزه: داده‌های تفکیک‌پذیر خطی



انگیزه: مرز تصمیم‌گیری بهینه

- پرسش. کدام مرز تصمیم‌گیری بهتر است؟

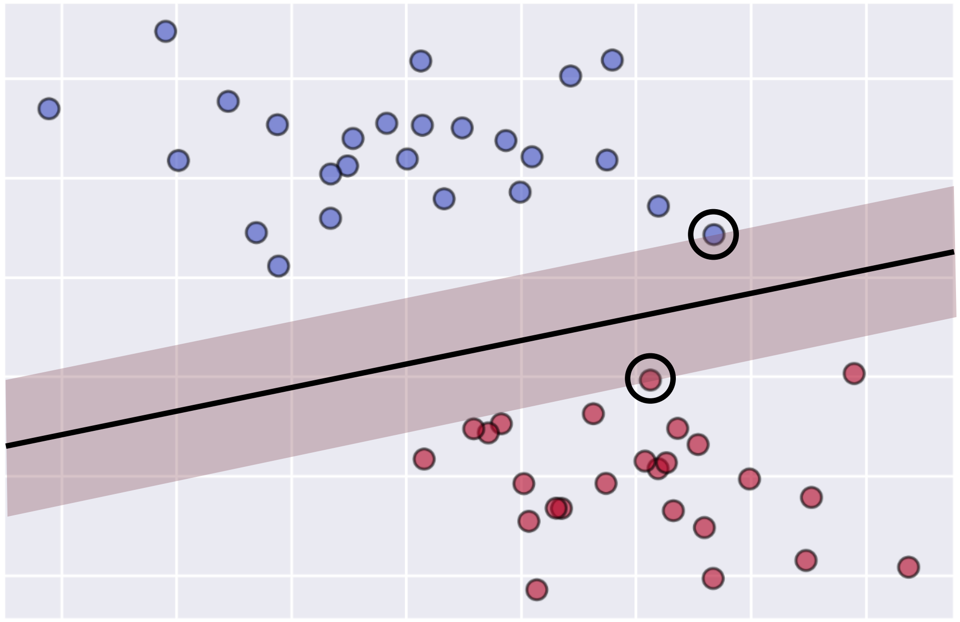




- راه‌حل بیشترین حاشیه. بیشترین پایداری در برابر تخریب داده‌ها. [افزایش قابلیت تعمیم]

انگیزه: بردارهای پشتیبان

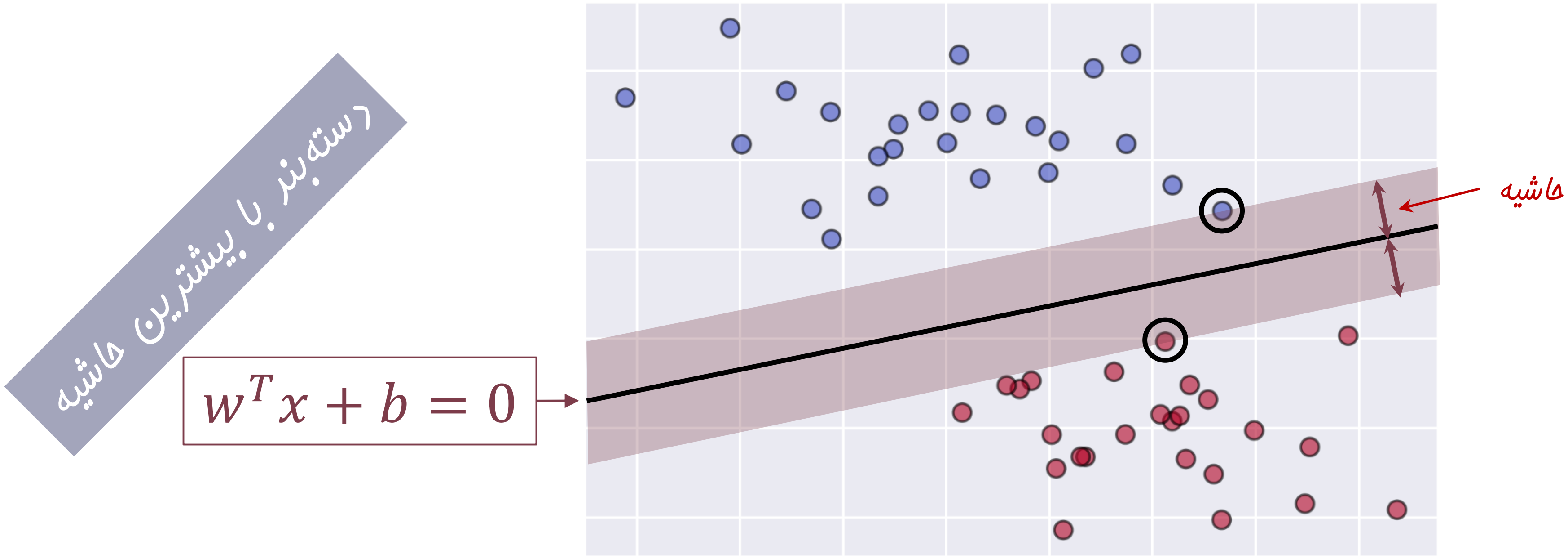
- بردار پشتیان. نزدیک‌ترین داده‌ها به مرز تصمیم‌گیری.



- هدف. بیشینه کردن فاصله بردارهای پشتیبان از مرز تصمیم‌گیری.

انگیزه: بیشینه‌سازی حاشیه

- حاشیه. فاصله بردارهای پشتیبان تا مرز تصمیم‌گیری.



- هدف. بیشینه کردن فاصله بردارهای پشتیبان از مرز تصمیم‌گیری.

مرز تصمیه‌گیری بهینه: نمادها

- نمونه‌های آموزشی.

$$X = (\boldsymbol{x}^t, y^t), \qquad y^t = \begin{cases} +1 & if\ x^t \in C_1 \\ -1 & if\ x^t \in C_2 \end{cases}$$

- هدف. یافتن بردار w و مقدار b به گونه‌ای که:

$$\boldsymbol{w}^T \boldsymbol{x}^t + b \geq +1 \quad \text{for} \quad y^t = +1$$

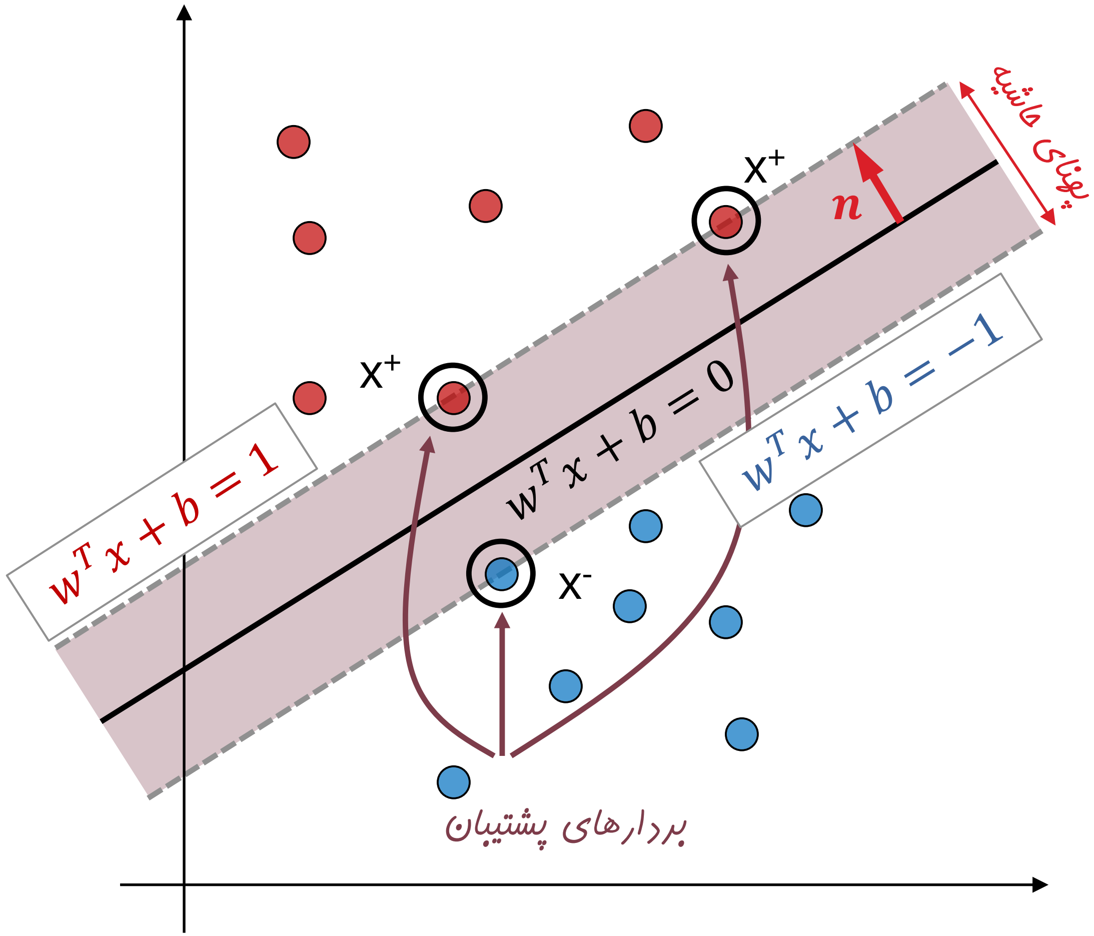
$$\boldsymbol{w}^T \boldsymbol{x}^t + b \leq -1 \quad \text{for} \quad y^t = -1$$

$$\Rightarrow \quad \boxed{y^t(\boldsymbol{w}^T \boldsymbol{x}^t + b) \geq +1}$$

$$\max(0{,}1 - y^t(\boldsymbol{w}^T \boldsymbol{x}^t + b))$$

تابع هدف: محاسبه حاشیه

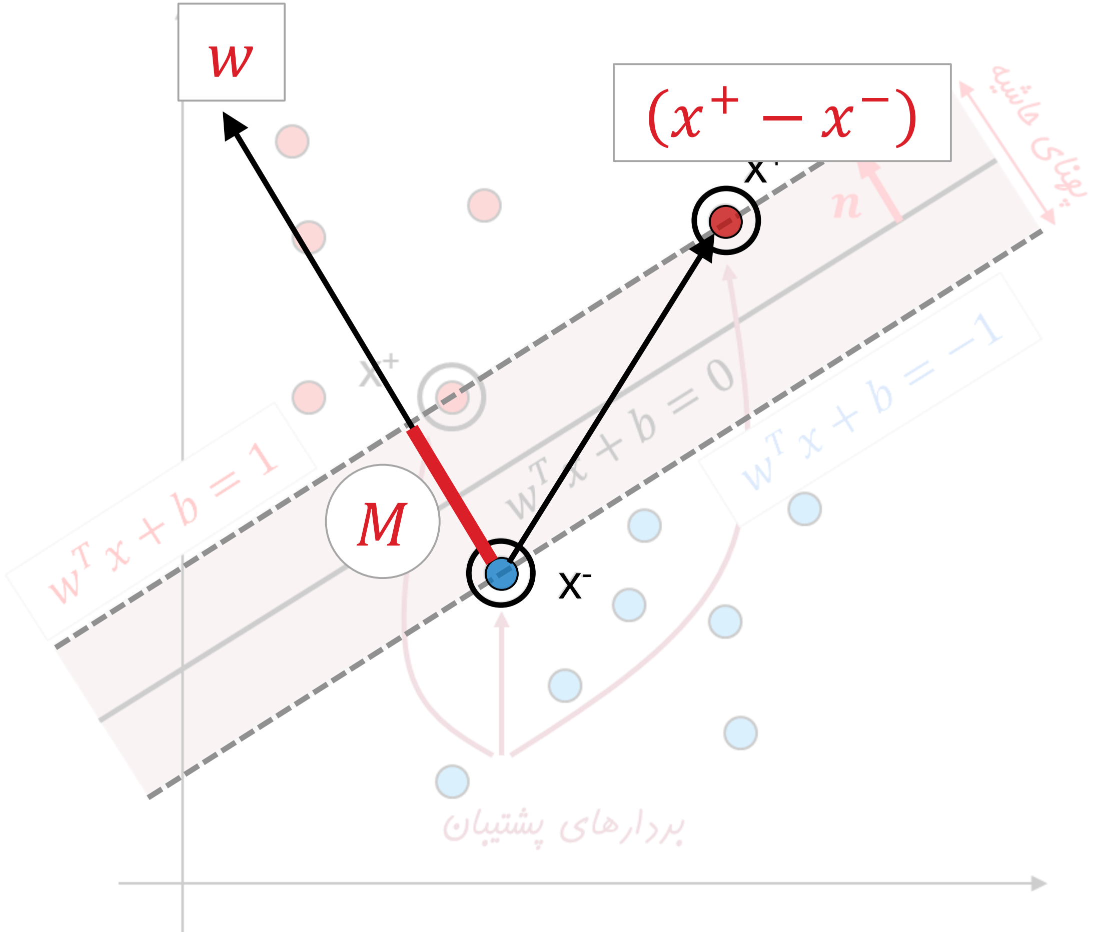
- می‌دانیم:



$$w^T x^+ + b = +1$$

$$w^T x^- + b = -1$$

تابع هدف: محاسبه حاشیه



- می‌دانیم:

$$w^T x^+ + b = +1$$

$$w^T x^- + b = -1$$

- بنابراین:

$$w^T (x^+ - x^-) = 2$$

$$\Rightarrow \|w\| \cdot \|x^+ - x^-\| \cos \alpha = 2$$

$$\Rightarrow \|w\| \cdot M = 2 \quad \Rightarrow \quad \boxed{M = 2/\|w\|}$$

تابع هدف

- هدف. بیشینه کردن اندازه حاشیه [فاصله بردارهای پشتیبان از مرز تصمیم‌گیری].

$$\boxed{M = \frac{2}{\|w\|}}$$

- توجه. برای بیشینه کردن حاشیه، می‌توان اندازه بردار w را کمینه نمود.

- محدودیت‌ها. مرز تصمیم‌گیری باید داده‌های هر دو دسته را به درستی از یکدیگر تفکیک کند.

تابع هدف: بیان رسمی

- تابع هدف.

$$
\begin{aligned}
& \min \frac{1}{2}\|\boldsymbol{w}\|^2 \\
\text{s.t.} \quad & (\boldsymbol{w}^T\boldsymbol{x}^t + b) \geq +1 \quad if\ y^t = +1 \\
& (\boldsymbol{w}^T\boldsymbol{x}^t + b) \leq -1 \quad if\ y^t = -1
\end{aligned}
$$

- ساده‌سازی.

$$
\begin{aligned}
& \min \frac{1}{2}\|\boldsymbol{w}\|^2 \\
\text{s.t.} \quad & y^t(\boldsymbol{w}^T\boldsymbol{x}^t + b) \geq +1
\end{aligned}
$$

تابع هدف: بیان رسمی

- تابع هدف.

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2$$
$$\text{s.t.} \quad y^t(\boldsymbol{w}^T\boldsymbol{x}^t + b) \geq +1$$

بهینه‌سازی محدب

- حل مسئله با استفاده از ضرایب لاگرانژ.

$$L_p = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{t=1}^{m} \alpha^t[y^t(\boldsymbol{w}^T\boldsymbol{x}^t + b) - 1]$$
$$= \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{t=1}^{m} \alpha^t y^t(\boldsymbol{w}^T\boldsymbol{x}^t + b) + \sum_{t=1}^{m} \alpha^t$$

جوزف لویی لاگرانژ

تابع هدف: بیان رسمی

- تابع هدف.

$$L_p = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{t=1}^{m} \alpha^t [y^t(\boldsymbol{w}^T \boldsymbol{x}^t + b) - 1]$$

$$= \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{t=1}^{m} \alpha^t y^t (\boldsymbol{w}^T \boldsymbol{x}^t + b) + \sum_{t=1}^{m} \alpha^t$$

$$\frac{\partial L_p}{\partial \boldsymbol{w}} = 0 \Rightarrow \boldsymbol{w} = \sum_{t=1}^{m} \alpha^t y^t \boldsymbol{x}^t$$

مرز تصمیم‌گیری یک ترکیب خطی از داده‌های آموزشی

$$\frac{\partial L_p}{\partial b} = 0 \Rightarrow \sum_{t=1}^{m} \alpha^t y^t = 0$$

تابع هدف: شکل دوگان

- تابع هدف.

$$L_d = \frac{1}{2}(\boldsymbol{w}^T\boldsymbol{w}) - \boldsymbol{w}^T\sum_{t=1}^{m}\alpha^t y^t \boldsymbol{x}^t - b\sum_{t=1}^{m}\alpha^t y^t + \sum_{t=1}^{m}\alpha^t$$

$$= -\frac{1}{2}(\boldsymbol{w}^T\boldsymbol{w}) + \sum_{t=1}^{m}\alpha^t$$

$$= -\frac{1}{2}\sum_{t=1}^{m}\sum_{s=1}^{m}\alpha^t\alpha^s y^t y^s (\boldsymbol{x}^t)^T\boldsymbol{x}^s + \sum_{t=1}^{m}\alpha^t$$

الگوریتم
بهینه‌سازی ترتیبی مینیمال
پِلَت (۱۹۹۹)

subject to $\sum_{t=1}^{m}\alpha^t y^t = 0$ and $\alpha^t \geq 0\ \forall t$

- مقدار بسیاری از ضرایب آلفا برابر با صفر است و تنها تعداد اندکی دارای مقدار بزرگ‌تر از صفر هستند؛
- داده‌هایی که به ازای آنها مقدار آلفا بزرگ‌تر از صفر است، همان بردارهای پشتیبان هستند.

تابع هدف: شکل برداری

- تابع هدف.

$$L_d = -\frac{1}{2}\sum_{t=1}^{m}\sum_{s=1}^{m}\alpha^t\alpha^s y^t y^s (\boldsymbol{x}^t)^T \boldsymbol{x}^s + \sum_{t=1}^{m}\alpha^t$$

$$= -\frac{1}{2}\alpha^T Q\alpha + e^T\alpha$$

الگوریتم
بهینه‌سازی ترتیبی مینیمال
پِلَت (۱۹۹۹)

$Q_{ts} = y^t y^s (x^t)^T x^s,$ $\quad e = [1 \quad 1 \quad \dots \quad 1]^T \in \mathbb{R}^m$

subject to $\sum_{t=1}^{m}\alpha^t y^t = 0$ and $\alpha^t \geq 0\ \forall t$

- مقدار بسیاری از ضرایب آلفا برابر با صفر است و تنها تعداد اندکی دارای مقدار بزرگ‌تر از صفر هستند؛
- داده‌هایی که به ازای آنها مقدار آلفا بزرگ‌تر از صفر است، همان بردارهای پشتیبان هستند.

داده‌های تفکیک‌ناپذیر خطی: حاشیه نرم

- اگر داده‌ها به صورت خطی تفکیک‌پذیر نباشند چه می‌شود؟
- حاشیه نرم. اجازه دادن اندکی خطا در جداسازی!
- خطای نرم.
- تابع هدف جدید.

$$y^t(\boldsymbol{w}^T\boldsymbol{x}^t + b) \geq 1 - \varepsilon^t$$

$$soft\ error = \sum_{t=1}^{m} \varepsilon^t$$

ضریب جریمه

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{t=1}^{m} \varepsilon^t$$

$$\text{s.t.} \quad y^t(\boldsymbol{w}^T\boldsymbol{x}^t + b) \geq 1 - \varepsilon^t$$

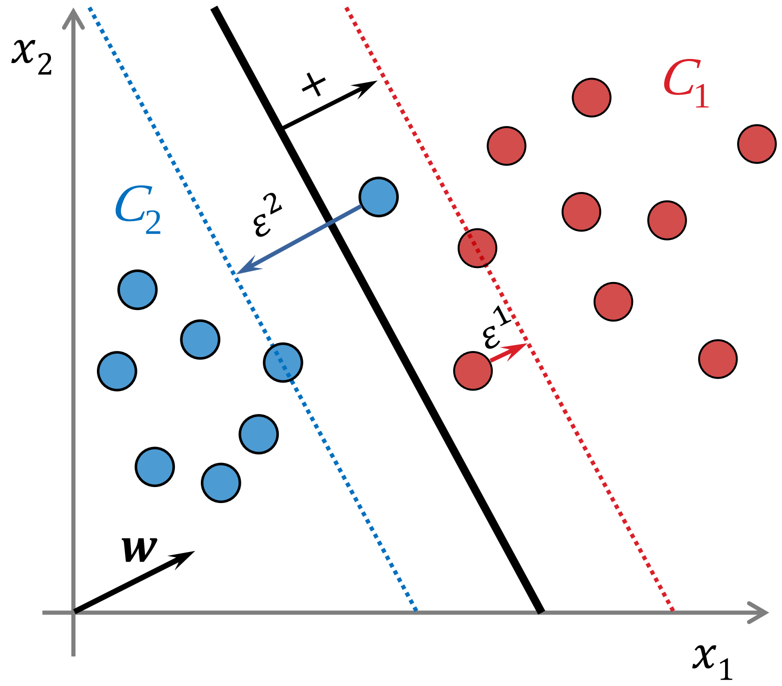
$$\varepsilon^t \geq 0$$

داده‌های تفکیک‌ناپذیر خطی: حاشیه نرم

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{t=1}^{m}\varepsilon^t$$

$$\text{s.t.} \quad y^t(\boldsymbol{w}^T\boldsymbol{x}^t + b) \geq 1 - \varepsilon^t$$

$$\varepsilon^t \geq 0$$



ضرایب لاگرانژ

$$L_p = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{t=1}^{m}\varepsilon^t - \sum_{t=1}^{m}\alpha^t[y^t(\boldsymbol{w}^T\boldsymbol{x}^t + b) - 1 + \varepsilon^t] - \sum_{t=1}^{m}\mu^t\varepsilon^t$$

داده‌های تفکیک‌ناپذیر خطی: حاشیه نرم

$$L_p = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{t=1}^{m}\varepsilon^t - \sum_{t=1}^{m}\alpha^t[y^t(\boldsymbol{w}^T\boldsymbol{x}^t + b) - 1 + \varepsilon^t] - \sum_{t=1}^{m}\mu^t\varepsilon^t$$

$$\frac{\partial L_p}{\partial \boldsymbol{w}} = 0 \Rightarrow \boldsymbol{w} = \sum_{t=1}^{m}\alpha^t y^t \boldsymbol{x}^t$$

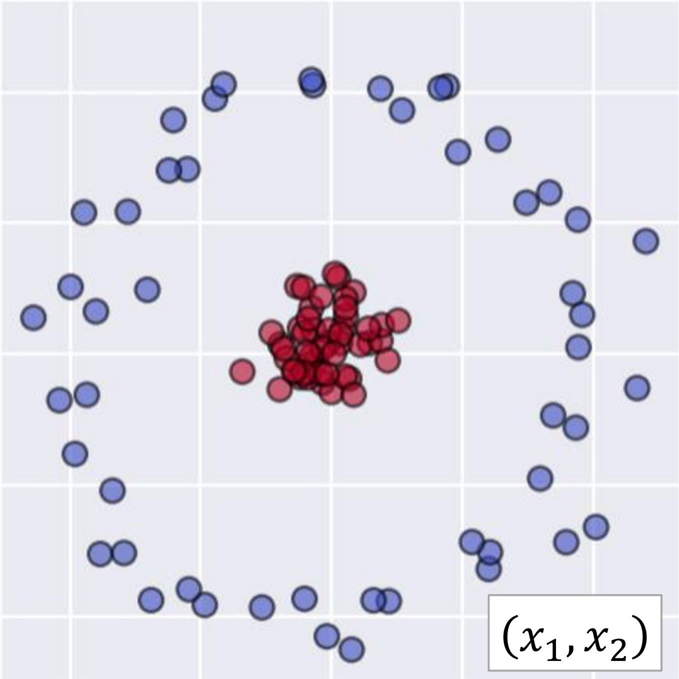
$$\frac{\partial L_p}{\partial b} = 0 \Rightarrow \sum_{t=1}^{m}\alpha^t y^t = 0$$

$$\frac{\partial L_p}{\partial \varepsilon^t} = 0 \Rightarrow C - \alpha^t - \mu^t = 0 \Rightarrow 0 \le \alpha^t \le C$$

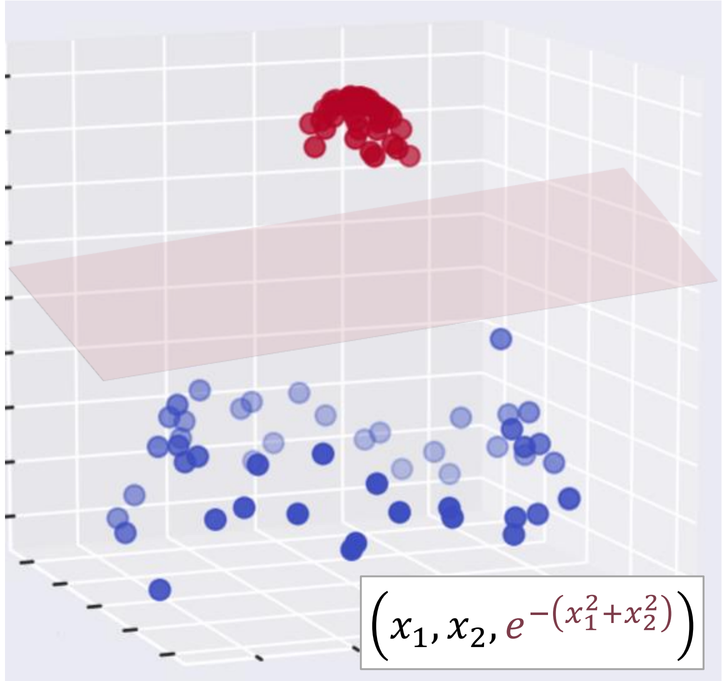
ترفند کرنل و دسته‌بندی غیرخطی

توابع کرنل

- ایده. نگاشت مسئله به یک فضای ویژگی جدید با استفاده از تبدیلات غیرخطی.
  - استفاده از یک مدل خطی در فضای جدید به منظور دسته‌بندی داده‌ها.

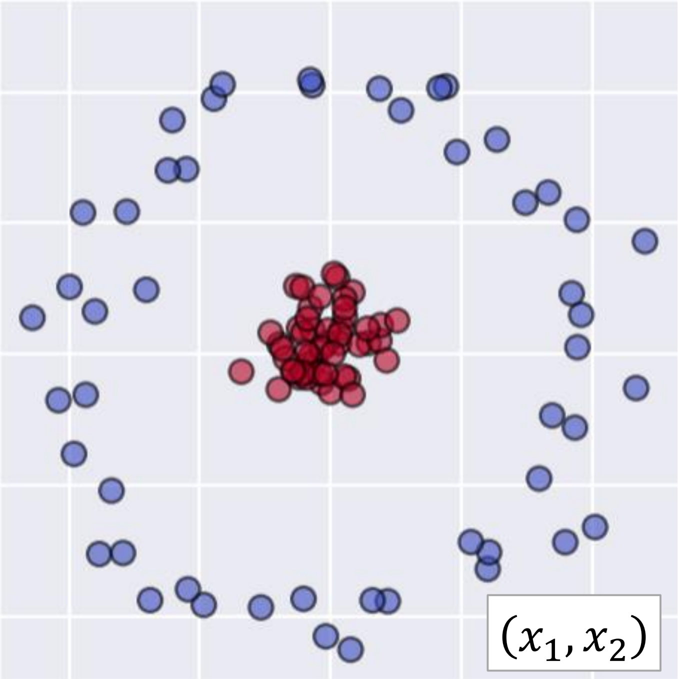


$\Phi:\ \boldsymbol{x} \rightarrow \varphi(\boldsymbol{x})$

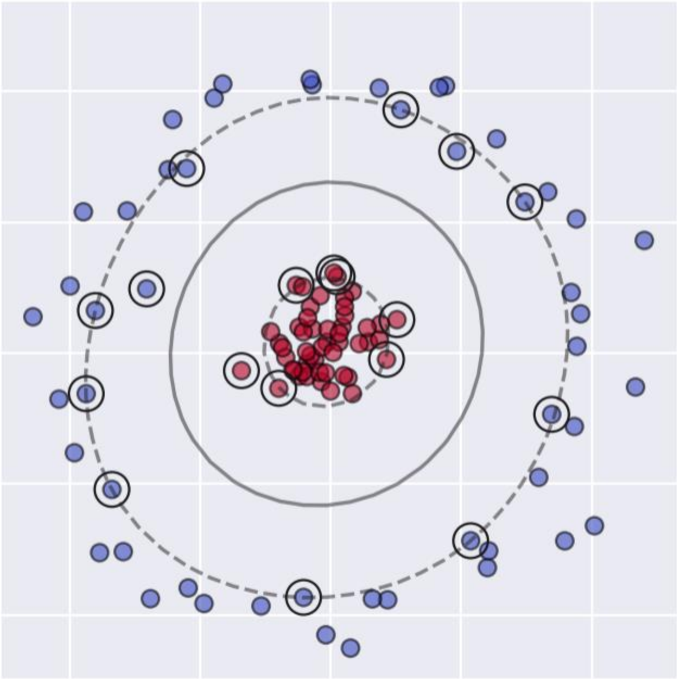


توابع کرنل

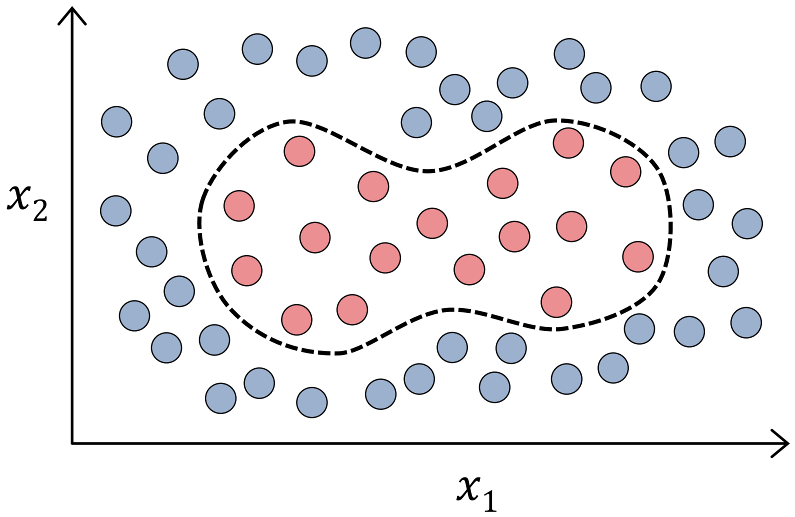
- ایده. نگاشت مسئله به یک فضای ویژگی جدید با استفاده از تبدیلات غیرخطی.
  - استفاده از یک مدل خطی در فضای جدید به منظور دسته‌بندی داده‌ها.
  - مدل خطی در فضای جدید متناظر با یک مدل غیرخطی در فضای اصلی است.



$$\Phi:\ \boldsymbol{x} \rightarrow \varphi(\boldsymbol{x})$$



مرز تصمیم‌گیری غیرخطی



- پیش‌بینی. $y = 1$ اگر:

$$h(x) = b + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2 + w_5x_1x_2 + \cdots \geq 0$$

- ویژگی‌ها.

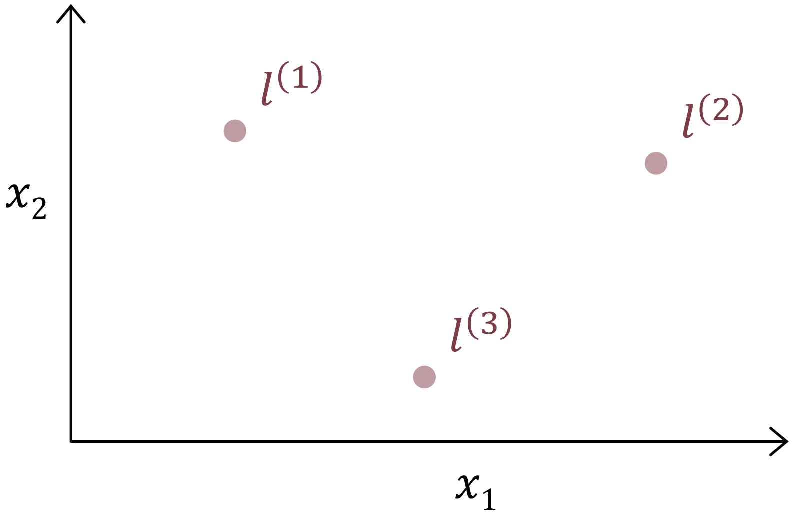
$$f_1 = x_1, \quad f_2 = x_2, \quad f_3 = x_1^2, \quad f_4 = x_2^2, \quad f_5 = x_1x_2, \quad \cdots$$

$$h(f) = b + w_1f_1 + w_2f_2 + w_3f_3 + w_4f_4 + w_5f_5 + \cdots$$ ⟵ مرز تصمیم‌گیری خطی

- پرسش. آیا روش بهتری برای انتخاب ویژگی‌ها وجود دارد؟

کرنل‌ها به عنوان معیار شباهت

- ایده. با داشتن x، مجموعه جدید ویژگی‌ها را بر اساس شباهت آن با نقاط راهنمای $l^{(1)}$، $l^{(2)}$ و $l^{(3)}$ انتخاب کن.



$$f_1 = sim(x, l^{(1)}) = \exp\left(-\frac{\left\|x - l^{(1)}\right\|^2}{2\sigma^2}\right)$$

$$f_2 = sim(x, l^{(2)}) = \exp\left(-\frac{\left\|x - l^{(2)}\right\|^2}{2\sigma^2}\right)$$

$$f_3 = sim(x, l^{(3)}) = \exp\left(-\frac{\left\|x - l^{(3)}\right\|^2}{2\sigma^2}\right)$$

کرنل (کرنل گوسی)

- تابع کرنل. معیاری به منظور محاسبه شباهت میان داده‌های x و y

کرنل‌ها به عنوان معیار شباهت

- تابع کرنل.

$$\boxed{f_i = sim\left(x, l^{(i)}\right) = \exp\left(-\frac{\left\|x - l^{(i)}\right\|^2}{2\sigma^2}\right)}$$

- حالت اول. $x \approx l^{(i)}$

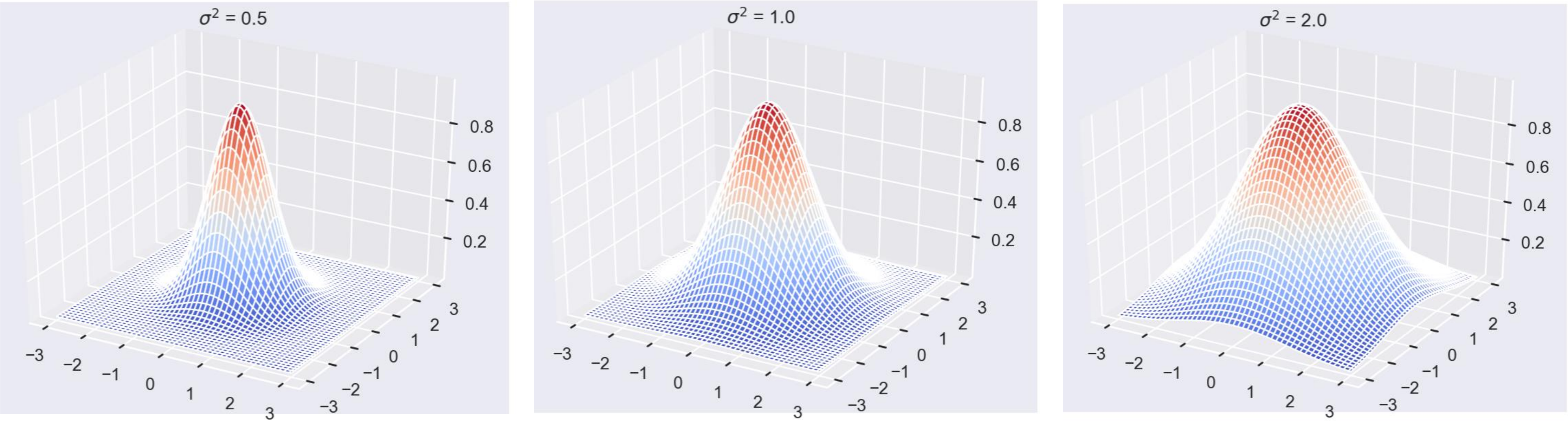
$$f_i \approx \exp\left(-\frac{0}{2\sigma^2}\right) = \exp(0) = 1$$

- حالت دوم. x بسیار دور از $l^{(i)}$

$$f_i \approx \exp\left(-\frac{\infty}{2\sigma^2}\right) = \exp(-\infty) = 0$$

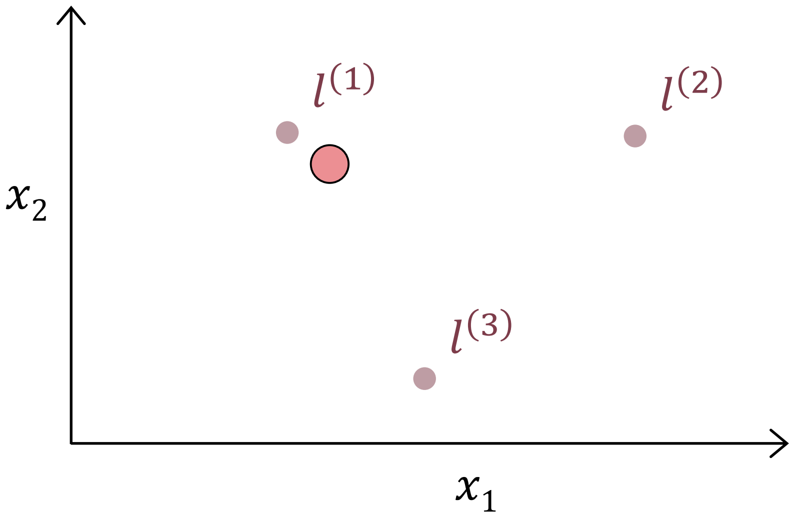
کرنل‌ها به عنوان معیار شباهت

$$f_i = sim\left(x, l^{(i)}\right) = \exp\left(-\frac{\left\|x - l^{(i)}\right\|^2}{2\sigma^2}\right)$$



کرنل‌ها به عنوان معیار شباهت

- پیش‌بینی. $y = 1$ اگر:



$$b + w_1 f_1 + w_2 f_2 + w_3 f_3 \geq 0$$

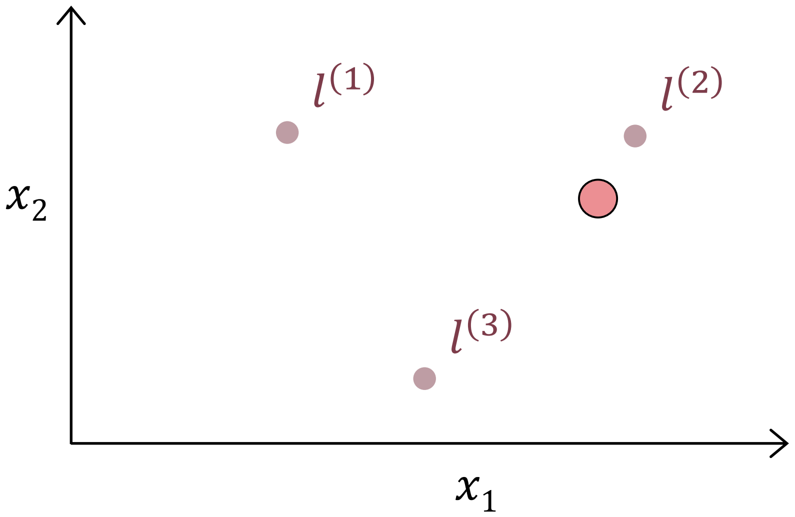
$b = -0.5$, $w_1 = 1.0$, $w_2 = 1.0$, $w_3 = 0.0$

$$f_1 \approx 1, f_2 \approx f_3 \approx 0$$

$$h(f) \approx -0.5 + (1.0)(1.0) + (1.0)(0.0) + (0.0)(0.0) = 0.5 \geq 0 \quad \Rightarrow \quad \boxed{y = 1}$$

کرنل‌ها به عنوان معیار شباهت

- پیش‌بینی. $y = 1$ اگر:



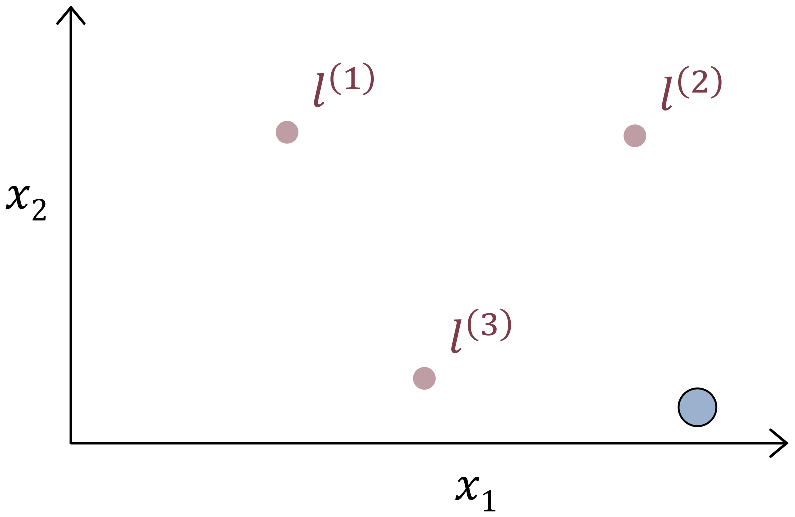
$$\underbrace{b}_{-0.5} + \underbrace{w_1}_{1.0} f_1 + \underbrace{w_2}_{1.0} f_2 + \underbrace{w_3}_{0.0} f_3 \geq 0$$

$$f_1 \approx f_3 \approx 0, f_2 \approx 1$$

$$h(f) \approx -0.5 + (1.0)(0.0) + (1.0)(1.0) + (0.0)(0.0) = 0.5 \geq 0 \quad \Rightarrow \quad \boxed{y = 1}$$

کرنل‌ها به عنوان معیار شباهت

- پیش‌بینی $y = 1$ اگر:



$$\boxed{b + w_1 f_1 + w_2 f_2 + w_3 f_3 \geq 0}$$

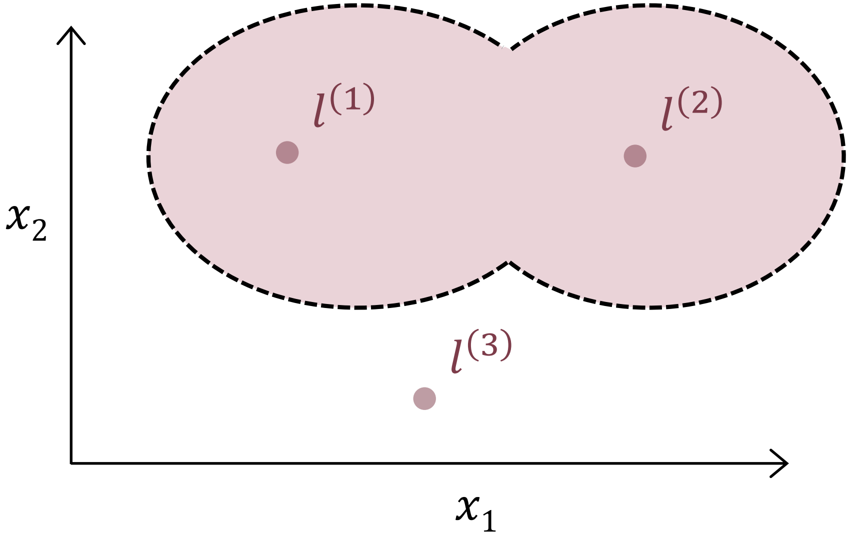
-0.5 1.0 1.0 0.0

$$f_1 \approx f_2 \approx f_3 \approx 0$$

$$h(f) \approx -0.5 + (1.0)(0.0) + (1.0)(0.0) + (0.0)(0.0) = -0.5 \geq 0 \quad \Rightarrow \quad \boxed{y = 0}$$

کرنل‌ها به عنوان معیار شباهت

- پیش‌بینی. $y = 1$ اگر:



$$b + w_1 f_1 + w_2 f_2 + w_3 f_3 \geq 0$$

-0.5 1.0 1.0 0.0

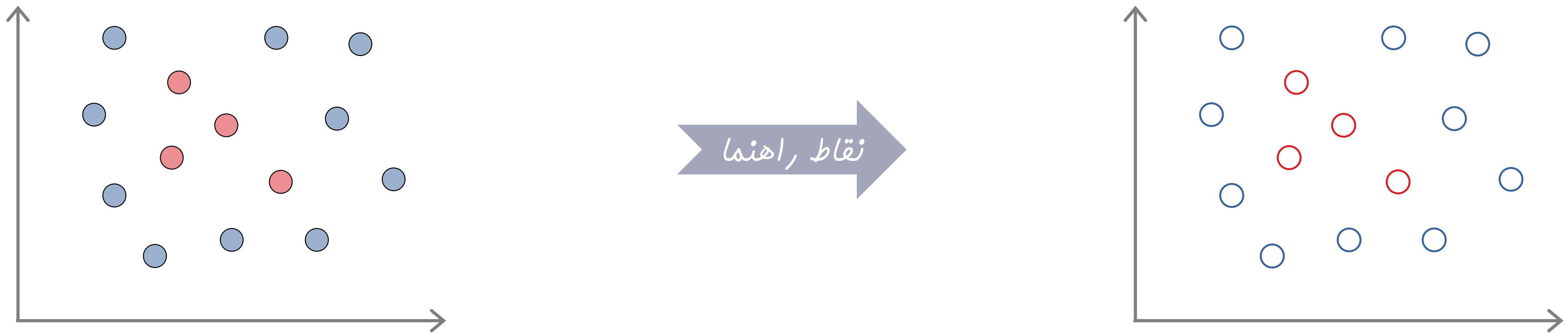
- مرز تصمیم‌گیری. نقاط نزدیک به $l^{(1)}$ و $l^{(2)}$ را در دسته ۱ و سایر نقاط را در دسته صفر دسته‌بندی می‌کند.

چند پرسش

- الگوریتم یادگیری نقاط راهنما را چگونه به صورت خودکار انتخاب می‌کند؟
- مقدار مناسب برای پارامترهای تابع کرنل چگونه تعیین می‌شوند؟
- آیا انواع دیگری از کرنل‌ها وجود دارد؟

انتخاب نقاط راهنما

- الگوریتم یادگیری نقاط راهنما را چگونه به صورت خودکار انتخاب می‌کند؟
  - به ازای هر نمونه در مجموعه آموزشی، یک نقطه راهنما مساوی با آن نمونه انتخاب می‌شود.



نگاشت ویژگی‌ها

- مجموعه آموزشی.

$$\{(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),\cdots,(x^{(m)},y^{(m)})\}$$

- نقاط راهنما.

$$l^{(1)}=x^{(1)},\ l^{(2)}=x^{(2)},\cdots,\ l^{(m)}=x^{(m)}$$

- نگاشت فضای ویژگی.

$$x=\begin{bmatrix} x_0=1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

نگاشت فضای ویژگی

$$f=\begin{bmatrix} f_0=1 \\ f_1=K(x,l^{(1)}) \\ f_2=K(x,l^{(2)}) \\ \vdots \\ f_m=K(x,l^{(m)}) \end{bmatrix}$$

ترفند کرنل

- تابع کرنل. پیش‌پردازش داده x با استفاده از توابع کرنل:

$$\boldsymbol{z} = \varphi(\boldsymbol{x}) = (\varphi_1(\boldsymbol{x}), \varphi_2(\boldsymbol{x}), \dots, \varphi_k(\boldsymbol{x}))$$

ممکن است بینهایت باشد!!!

$$g(\boldsymbol{z}) = \boldsymbol{w}^T\boldsymbol{z} + b$$

$$g(\boldsymbol{x}) = \boldsymbol{w}^T\varphi(\boldsymbol{x}) + b$$

- مرز تصمیم‌گیری.

$$\boldsymbol{w} = \sum_{t=1}^{m} \alpha^t y^t \boldsymbol{z}^t = \sum_{t=1}^{m} \alpha^t y^t \varphi(\boldsymbol{x}^t)$$

- دسته‌بندی داده جدید.

$$g(\boldsymbol{x}) = \boldsymbol{w}^T\varphi(\boldsymbol{x}) + b = \left(\sum_{t=1}^{m} \alpha^t y^t \varphi(\boldsymbol{x}^t)^T\right)\varphi(\boldsymbol{x}) + b = \left(\sum_{t=1}^{m} \alpha^t y^t \varphi(\boldsymbol{x}^t)^T\varphi(\boldsymbol{x})\right) + b = \left(\sum_{t=1}^{m} \alpha^t y^t k(\boldsymbol{x}^t, \boldsymbol{x})\right) + b$$

توابع کرنل

$$L_p = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{t=1}^{m}\varepsilon^t$$

$$\text{s.t.} \qquad y^t\boldsymbol{w}^T\varphi(\boldsymbol{x}^t) \ge 1 - \varepsilon^t$$

$$\varepsilon^t \ge 0$$

$$L_p = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{t=1}^{m}\varepsilon^t - \sum_{t=1}^{m}\alpha^t[y^t\boldsymbol{w}^T\varphi(\boldsymbol{x}^t) - 1 + \varepsilon^t] - \sum_{t=1}^{m}\mu^t\varepsilon^t$$

ضرایب لاگرانژ ضرایب لاگرانژ

توابع کرنل: مسئله اصلی

$$L_p = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{t=1}^{m}\varepsilon^t - \sum_{t=1}^{m}\alpha^t[y^t\boldsymbol{w}^T\varphi(\boldsymbol{x}^t) - 1 + \varepsilon^t] - \sum_{t=1}^{m}\mu^t\varepsilon^t$$

$$\frac{\partial L_p}{\partial \boldsymbol{w}} = 0 \Rightarrow \boldsymbol{w} = \sum_{t=1}^{m}\alpha^t y^t \varphi(\boldsymbol{x}^t)$$

$$\frac{\partial L_p}{\partial \varepsilon^t} = 0 \Rightarrow C - \alpha^t - \mu^t = 0 \Rightarrow 0 \le \alpha^t \le C$$

توابع کرنل: مسئله دوگان

$$L_d = -\frac{1}{2}\sum_{t=1}^{m}\sum_{s=1}^{m}\alpha^t\alpha^s y^t y^s \varphi(\boldsymbol{x}^t)^T\varphi(\boldsymbol{x}^s) + \sum_{t=1}^{m}\alpha^t$$

subject to $\sum_{t=1}^{m}\alpha^t y^t = 0$ and $0 \le \alpha^t \le C \ \forall t$

- ایده ماشین‌های کرنل. [ترفند کرنل]
  - جایگزینی حاصل‌ضرب داخلی توابع پایه با یک تابع کرنل به صورت $K(\boldsymbol{x}^t, \boldsymbol{x}^s)$

$$L_d = -\frac{1}{2}\sum_{t=1}^{m}\sum_{s=1}^{m}\alpha^t\alpha^s y^t y^s K(\boldsymbol{x}^t, \boldsymbol{x}^s) + \sum_{t=1}^{m}\alpha^t$$

ماتریس گِرَم: یک ماتریس متقارن و مثبت معین (برای تفکیک‌پذیری خطی)

توابع کرنل: کرنل چندجمله‌ای

- کرنل چند جمله‌ای. یک چندجمله‌ای از درجه q.

$$K(x^t, x) = (x^T x^t + 1)^q$$

- مثال. $[q = 2, d = 2]$

۳ ضرب، ۲ جمع

$$K(x, y) = (x^T y + 1)^2$$
$$= (x_1 y_1 + x_2 y_2 + 1)^2$$
$$= 1 + 2x_1 y_1 + 2x_2 y_2 + 2x_1 x_2 y_1 y_2 + x_1^2 y_1^2 + x_2^2 y_2^2$$

۶ ضرب، ۵ جمع

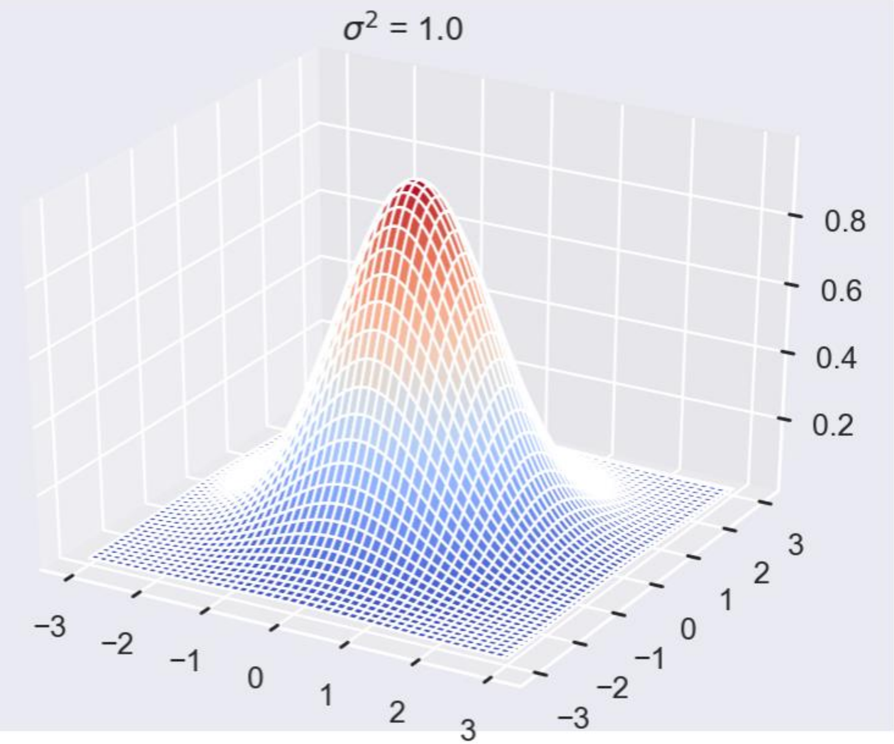
$$\varphi(x) = \left[1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2, x_1^2, x_2^2\right]^T$$
$$\varphi(y) = \left[1, \sqrt{2}y_1, \sqrt{2}y_2, \sqrt{2}y_1 y_2, y_1^2, y_2^2\right]^T$$

توابع کرنل: کرنل گوسی

- کرنل گوسی.

$$K(\boldsymbol{x}^t, \boldsymbol{x}) = \exp\left(-\frac{\|\boldsymbol{x}^t - \boldsymbol{x}\|^2}{2\sigma^2}\right)$$

- یافتن یک مقدار مناسب برای σ.
  - با استفاده از مجموعه اعتبارسنجی [انتخاب مدل]
  - مقادیر بزرگ‌تر: مرز تصمیم‌گیری هموارتر

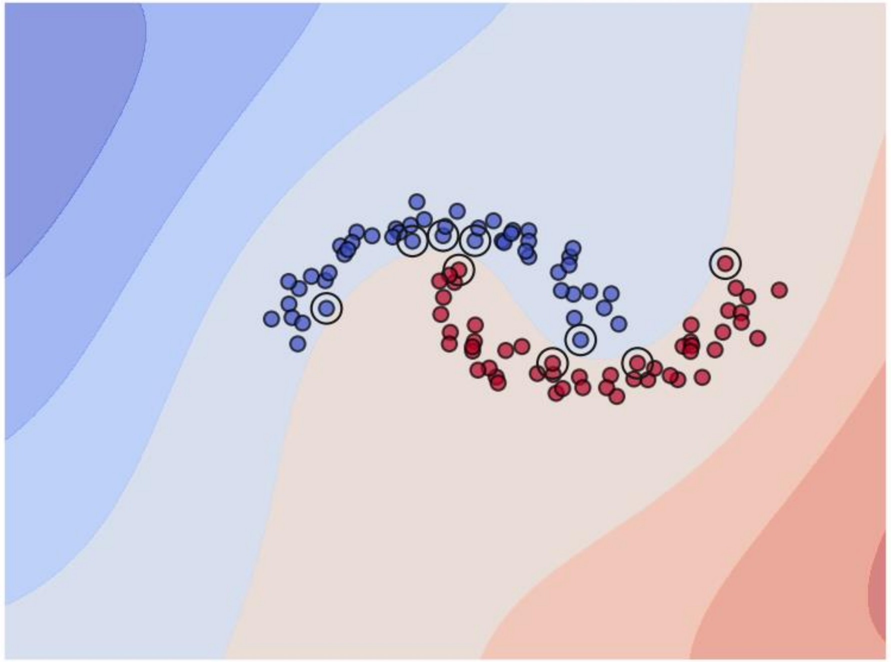


توابع کرنل: کرنل گوسی

- کرنل گوسی.

$$K(\boldsymbol{x}^t, \boldsymbol{x}) = \exp\left(-\frac{\|\boldsymbol{x}^t - \boldsymbol{x}\|^2}{2\sigma^2}\right)$$

- یافتن یک مقدار مناسب برای σ.
  - با استفاده از مجموعه اعتبارسنجی [انتخاب مدل]
  - مقادیر بزرگ‌تر: مرز تصمیم‌گیری هموارتر



ابر پارامترها: انتخاب مدل

- راه‌حل اول. [یک ایده بسیار بد]
  - انتخاب مقداری که منجر به بالاترین دقت دسته‌بندی بر روی مجموعه آزمایشی می‌شود.

داده‌های آزمایشی	داده‌های آموزشی

- توجه. [بسیار مهم]
  - از مجموعه آزمایشی در انتهای مراحل و تنها برای تخمین قابلیت تعمیم دسته‌بند استفاده کنید.

ابر پارامترها: انتخاب مدل

- راه‌حل دوم. [اعتبارسنجی چند بخشی]

داده‌های آزمایشی	داده‌های آموزشی

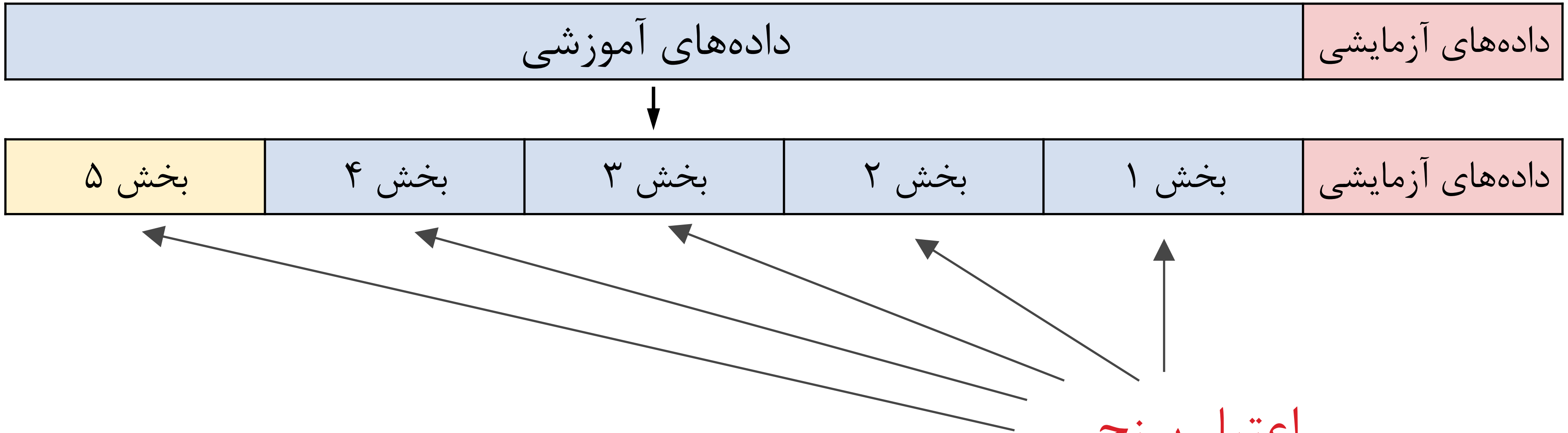
↓

داده‌های آزمایشی	بخش ۱	بخش ۲	بخش ۳	بخش ۴	بخش ۵
داده‌های آزمایشی	بخش ۱	بخش ۲	بخش ۳	بخش ۴	بخش ۵
داده‌های آزمایشی	بخش ۱	بخش ۲	بخش ۳	بخش ۴	بخش ۵
داده‌های آزمایشی	بخش ۱	بخش ۲	بخش ۳	بخش ۴	بخش ۵
داده‌های آزمایشی	بخش ۱	بخش ۲	بخش ۳	بخش ۴	بخش ۵

داده‌های اعتبار سنجی [برای تعیین مقدار ابرپارامترها]

ابر پارامترها: انتخاب مدل

- راه‌حل دوم. [اعتبارسنجی چند بخشی]

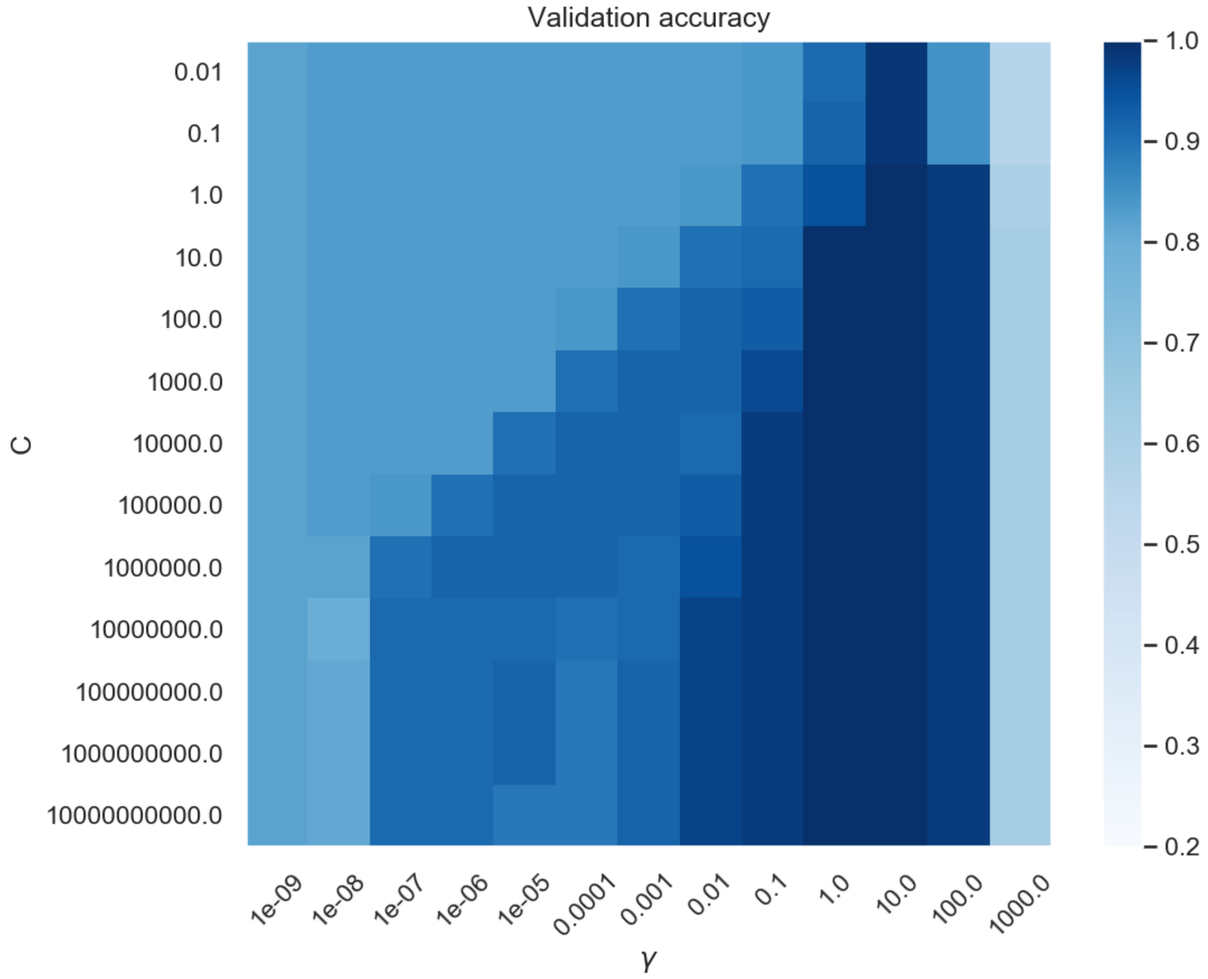


اعتبار سنجی

هر بار یک بخش را به عنوان داده‌های اعتبارسنجی انتخاب کن و سپس از نتایج به دست آمده میانگین بگیر

پارامترهای ماشین بردار پشتیبان

- پرسش. مقدار مناسب برای پارامترهای تابع کرنل چگونه تعیین می‌شوند؟



- پارامتر C.
  - مقادیر کوچک‌تر: بایاس بیشتر، واریانس کمتر
  - مقادیر بزرگ‌تر: بایاس کمتر، واریانس بیشتر

- پارامتر σ.
  - مقادیر کوچک‌تر: بایاس کمتر، واریانس بیشتر
  - مقادیر بزرگ‌تر: بایاس بیشتر، واریانس کمتر

راهنمای استفاده از ماشین بردار پشتیبان

- پیاده‌سازی. استفاده از بسته‌های نرم‌افزاری موجود مانند $\mathrm{SVM}^{\mathrm{light}}$ و LIBSVM

- تعیین تابع کرنل.
  - کرنل خطی (عدم استفاده از کرنل): وقتی که $n \gg m$
  - گوسی، چندجمله‌ای، رشته‌ای و ...

- تعیین مقدار پارامترها. جستجوی توری
  - انتخاب مقدار برای پارامتر C
  - انتخاب مقدار برای پارامترهای تابع کرنل (مانند σ)

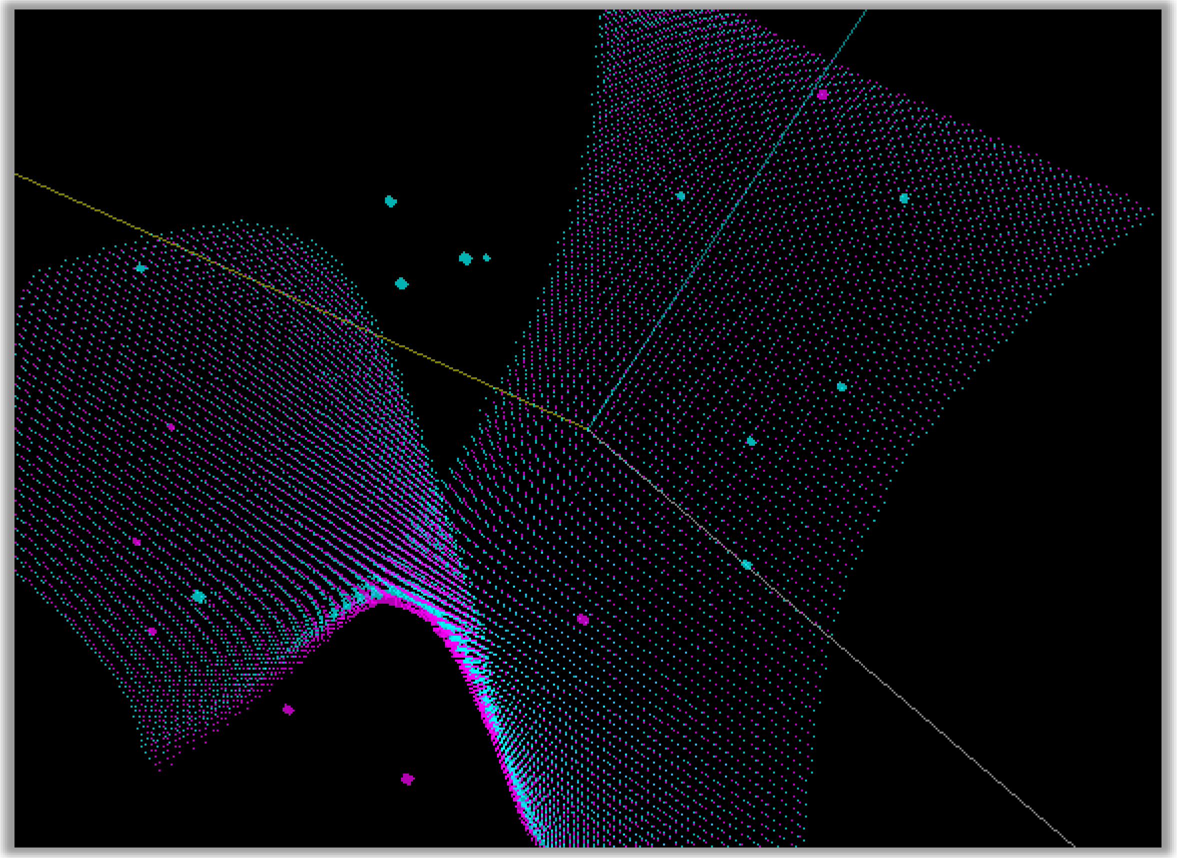
ماشین بردار پشتیبان، شبکه عصبی یا رگرسیون لجستیک

- حالت ۱. $[n \gg m]$
  - مثال: تشخیص هرزنامه (۱۰۰۰ نمونه آموزشی، ۵۰۰۰۰ ویژگی)
  - رگرسیون لجستیک یا ماشین بردار پشتیبان خطی

- حالت ۲. [تعداد ویژگی‌ها کم، تعداد نمونه‌های آموزشی زیاد]
  - ماشین بردار پشتیبان با کرنل گوسی

- توجه. شبکه‌های عصبی در تمامی حالت‌های فوق قابل استفاده هستند، اما ممکن است به زمان بیشتری برای آموزش نیاز داشته باشند.

اجرای نمایشی



http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/svmtoy3d/

پیوست: مطالب بیشتر در مورد کرنل‌ها

- پرسش. چگونه می‌دانیم استفاده از کرنل‌ها در جداسازی داده‌ها به ما کمک می‌کند؟
  - در فضای n-بعدی، هر مجموعه از n بردار مستقل، به صورت خطی تفکیک‌پذیر هستند.
  - اگر ماتریس K یک ماتریس مثبت معین باشد، آنگاه داده‌ها به صورت خطی تفکیک‌پذیر هستند.

- قضیه. ماتریس K یک ماتریس مثبت معین است، زیرا $K = L^T L$
  - ستون i در ماتریس L برابر است با بردار $\phi\left(x^{(i)}\right)$

- اثبات. بردار غیر صفر v را در نظر بگیرید. در این صورت:

$$v^T K v = v^T L^T L v = (Lv)^T (Lv) = w^T w = \|w\|^2 \geq 0$$

  - و چون L و v هر دو مخالف صفر هستند، بردار w نیز مخالف صفر است. یعنی:

$$\|w\|^2 > 0 \Rightarrow v^T K v > 0 \Rightarrow K \text{ is positive definite}$$