

ماشین‌های بردار پشتیبان

سید ناصر رضوی www.snrazavi.ir

۱۳۹۷

فهرست مطالب

- انگیزه. مرز تصمیم‌گیری بهینه
- مفاهیم پایه. بردارهای پشتیبان و بیشینه‌سازی حاشیه
- تابع هدف. مسئله اصلی و مسئله دوگان
- دسته‌بندی خطی و غیرخطی. حاشیه نرم
- دسته‌بندی غیرخطی. ترفند کرنل
- دسته‌بندی چند دسته‌ای. ماشین بردار پشتیبان چند دسته‌ای

□ ماشین‌های بردار پشتیبان [اوپنیک، ۱۹۹۲]

□ یکی از پرطرفدارترین الگوریتم‌های یادگیری ماشین!

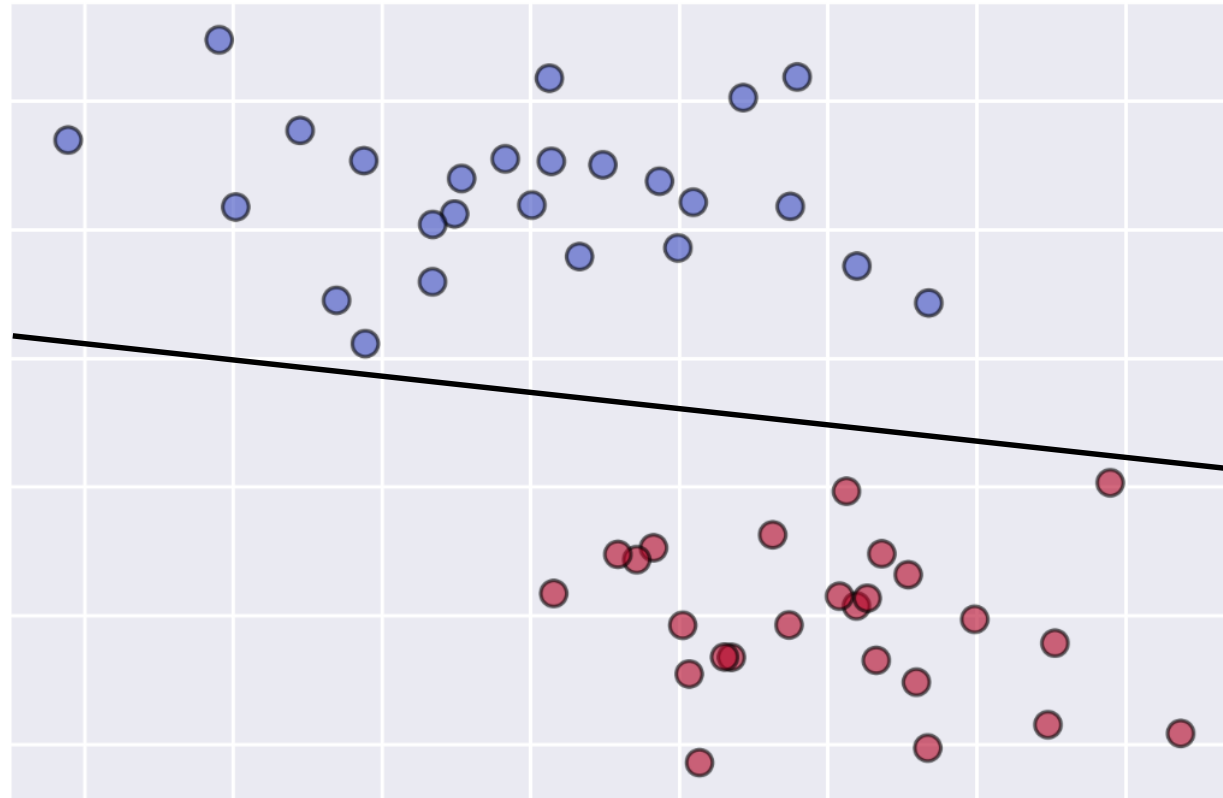
□ جداسازی بهتر داده‌ها نسبت به سایر روش‌های یادگیری ماشین (مسائل دسته‌بندی)!

□ استفاده از آن نسبتاً آسان است!

□ استفاده از **ترفند کرنل**:

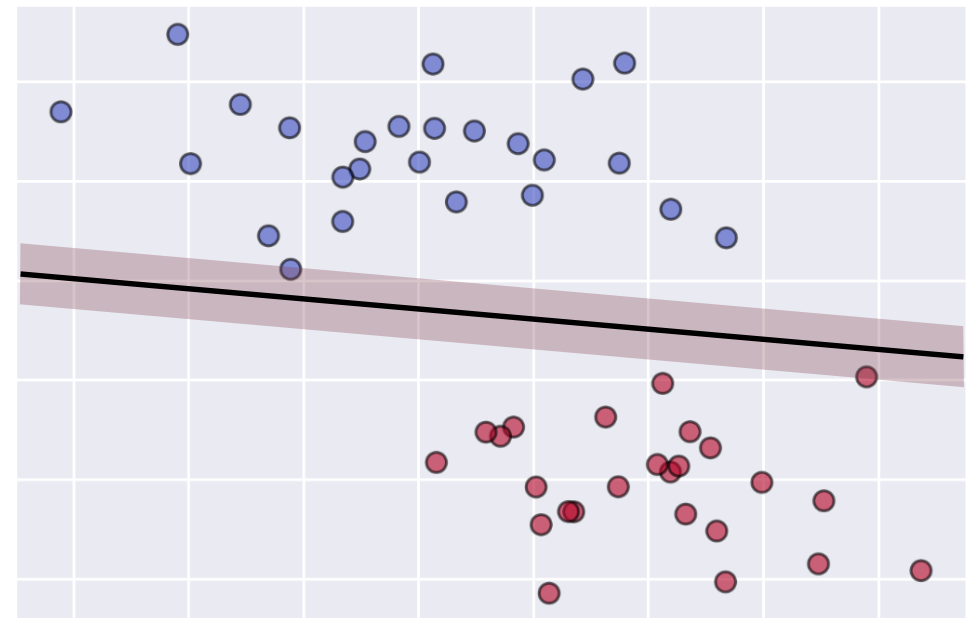
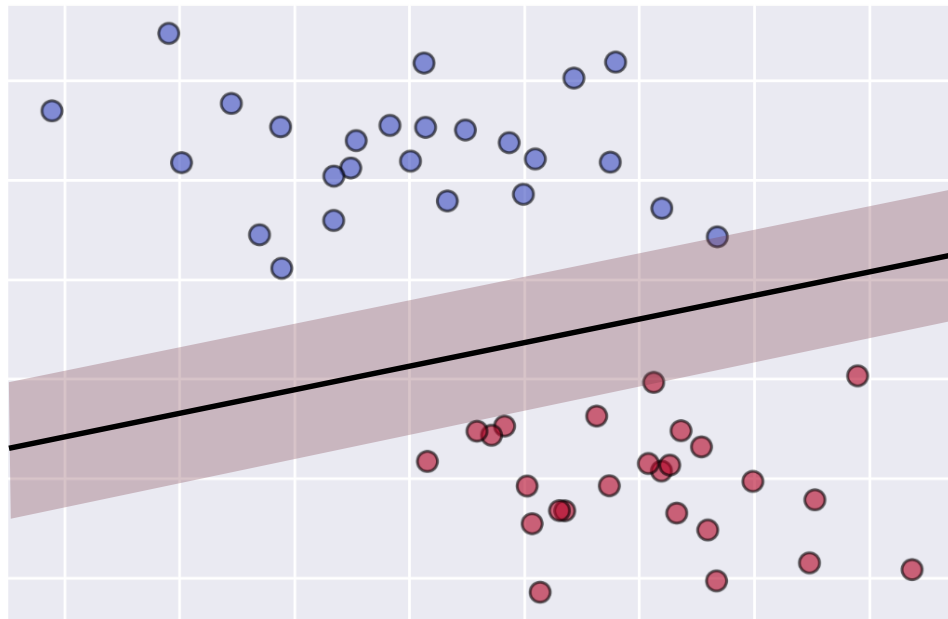
■ دسته‌بندی، رگرسیون، تخمین توزیع، دسته‌بندی تک دسته‌ای و ...

انگیزه: داده‌های تفکیک‌پذیر خطی



انگیزه: مرز تصمیم‌گیری بهینه

□ پرسش. کدام مرز تصمیم‌گیری بهتر است؟

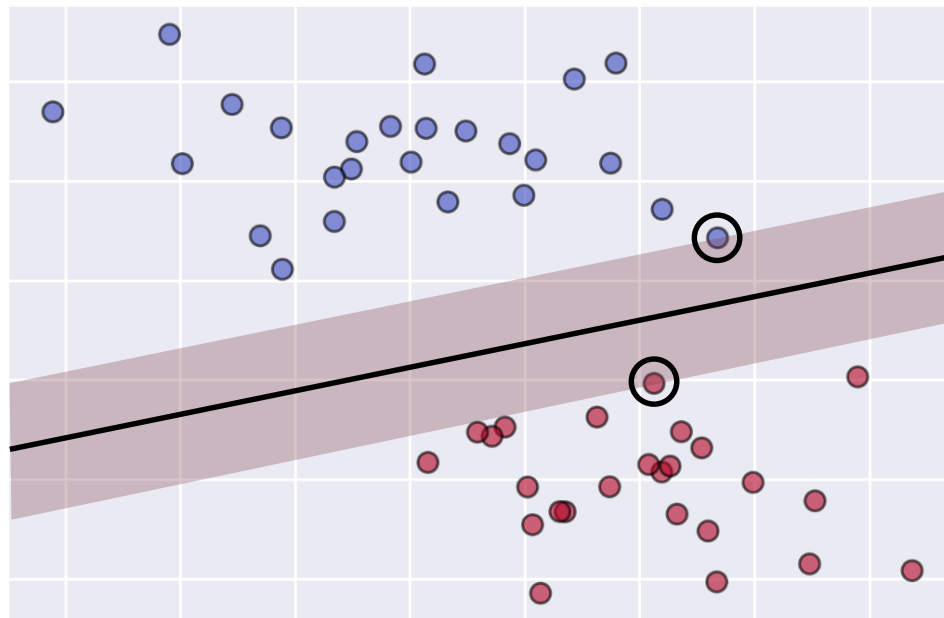


□ راه‌حل بیشترین حاشیه. بیشترین پایداری در برابر تخریب داده‌ها. [افزایش قابلیت تعمیم]

انگیزه: بردارهای پشتیبان

۶

□ بردار پشتیبان. نزدیک‌ترین داده‌ها به مرز تصمیم‌گیری.

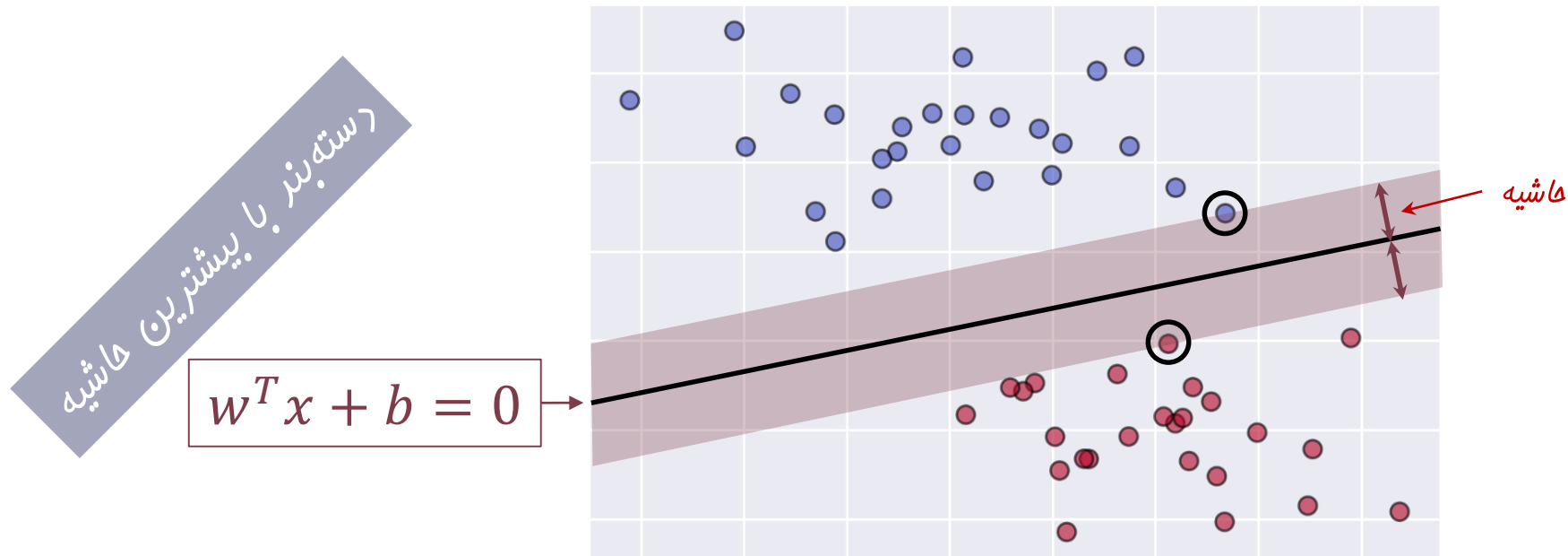


□ هدف. پیشینه کردن فاصله بردارهای پشتیبان از مرز تصمیم‌گیری.

انگیزه: بیشینه‌سازی حاشیه

۷

□ حاشیه. فاصله بردارهای پشتیبان تا مرز تصمیم‌گیری.



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

مرز تصمیم‌گیری بهینه: نمادها

۸

□ نمونه‌های آموزشی.

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

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

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

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



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

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

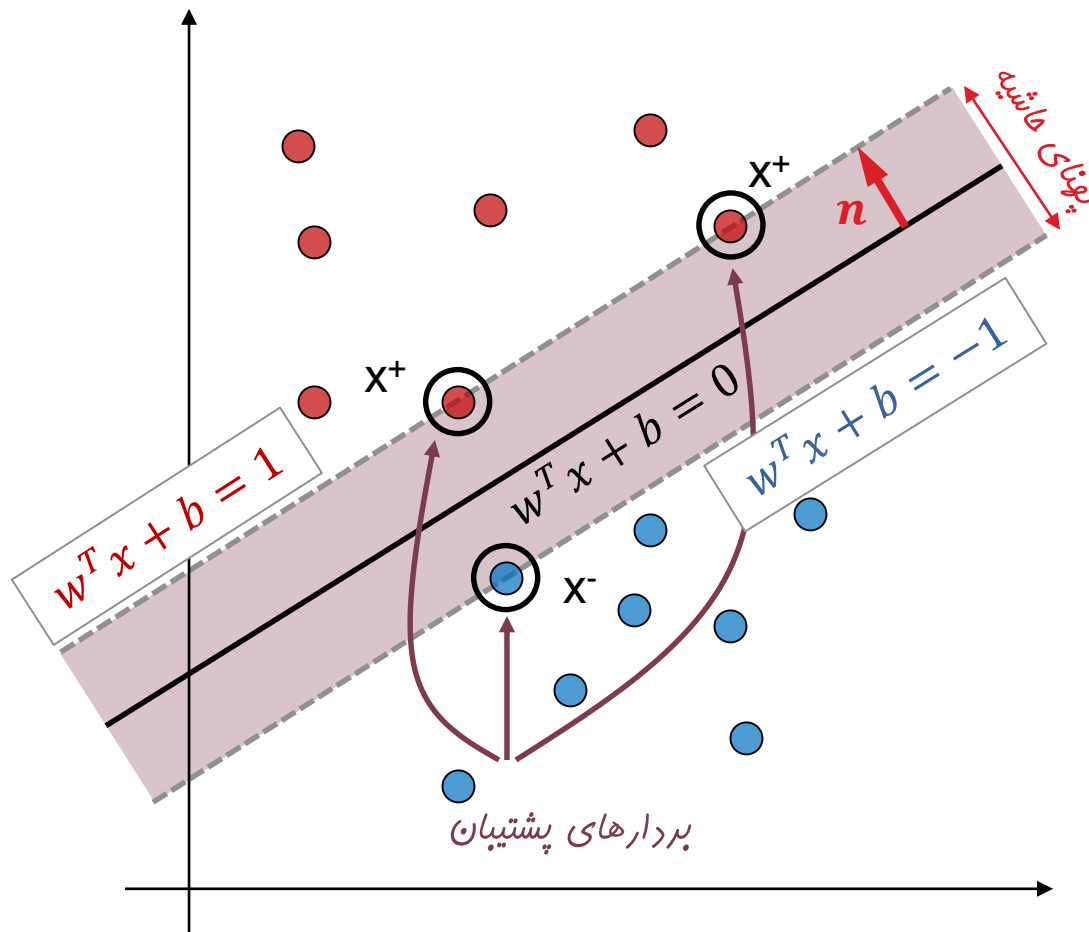
تابع هدف: محاسبه ماشیه

۹

□ می دانیم:

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

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



تابع هدف: محاسبه ماشیه

□ می دانیم:

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

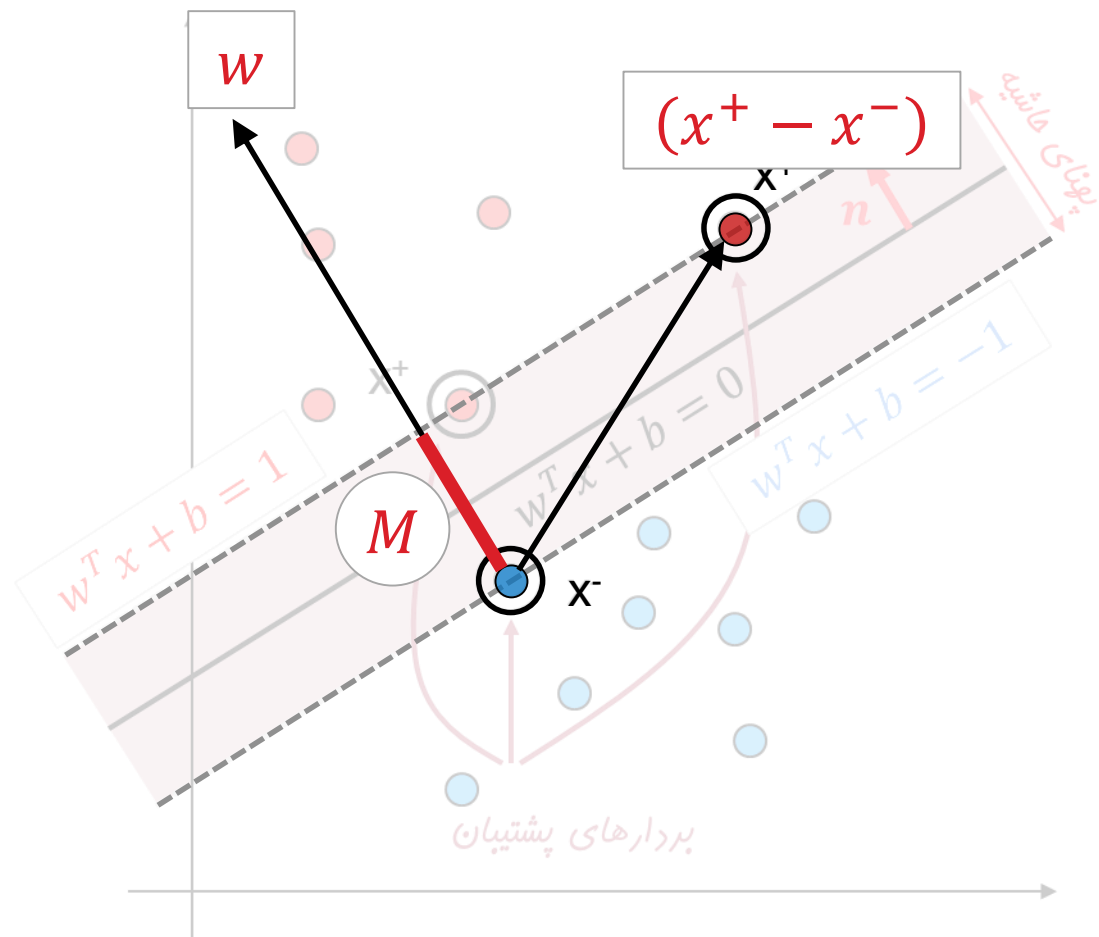
$$w^T x^- + b = -1$$

□ بنابراین:

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

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

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



تابع هدف

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

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

□ 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

۱۲

□ تابع هدف.

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

□ ساده‌سازی.

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

تابع هدف: بیان رسمی

۱۳

□ تابع هدف.

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

← بهینه‌سازی مرب

□ حل مسئله با استفاده از ضرایب لاگرانژ.

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



جوزف لویی لاگرانژ

تابع هدف: بیان رسمی

۱۴

□ تابع هدف.

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

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

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

مرز تصمیم‌گیری یک ترکیب فطی از داده‌های آموزشی

تابع هدف: شکل دوگان

□ تابع هدف.

$$\begin{aligned} L_d &= \frac{1}{2} (\mathbf{w}^T \mathbf{w}) - \mathbf{w}^T \sum_{t=1}^m \alpha^t y^t \mathbf{x}^t - b \sum_{t=1}^m \alpha^t y^t + \sum_{t=1}^m \alpha^t \\ &= -\frac{1}{2} (\mathbf{w}^T \mathbf{w}) + \sum_{t=1}^m \alpha^t \\ &= -\frac{1}{2} \sum_{t=1}^m \sum_{s=1}^m \alpha^t \alpha^s y^t y^s (\mathbf{x}^t)^T \mathbf{x}^s + \sum_{t=1}^m \alpha^t \end{aligned}$$

الگوریتم
بهینه‌سازی ترتیبی مینیمال
پلت (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.

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

الگوریتم
بهینه‌سازی ترتیبی مینیمال
پلّت (۱۹۹۹) ←

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

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: ماشیه نرم

□ اگر داده‌ها به صورت خطی تفکیک‌پذیر نباشند چه می‌شود؟

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

□ حاشیه نرم. اجازه دادن اندکی خطا در جداسازی!

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

□ خطای نرم.

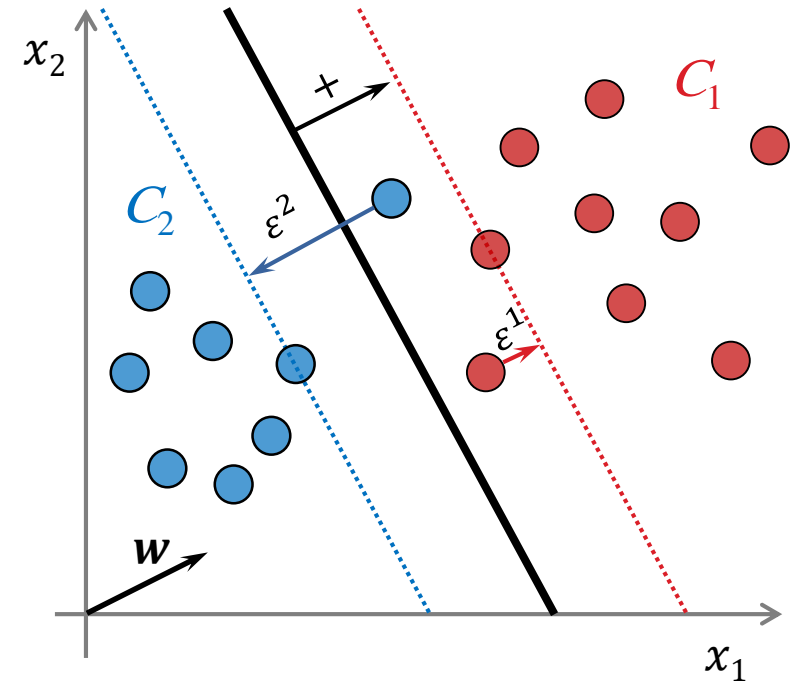
ضریب جریمه

□ تابع هدف جدید.

$$\begin{aligned} & \min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{t=1}^m \varepsilon^t \\ \text{s.t.} \quad & y^t(\mathbf{w}^T \mathbf{x}^t + b) \geq 1 - \varepsilon^t \\ & \varepsilon^t \geq 0 \end{aligned}$$

داده‌های تفکیک‌ناپذیر خطی: ماشین بردار پشتیبان

$$\begin{aligned} & \min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{t=1}^m \varepsilon^t \\ \text{s.t.} \quad & y^t (\mathbf{w}^T \mathbf{x}^t + b) \geq 1 - \varepsilon^t \\ & \varepsilon^t \geq 0 \end{aligned}$$



ضرایب لاگرانژ

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

داده‌های تفکیک‌ناپذیر خطی: ماشین بردار

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

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

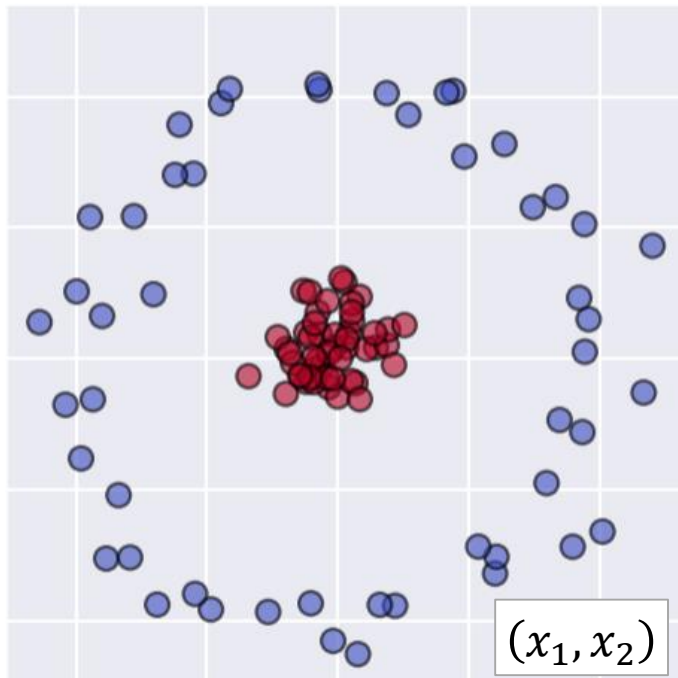
$$\frac{\partial L_p}{\partial b} = 0 \Rightarrow \sum_{t=1}^m \alpha^t y^t = 0$$

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

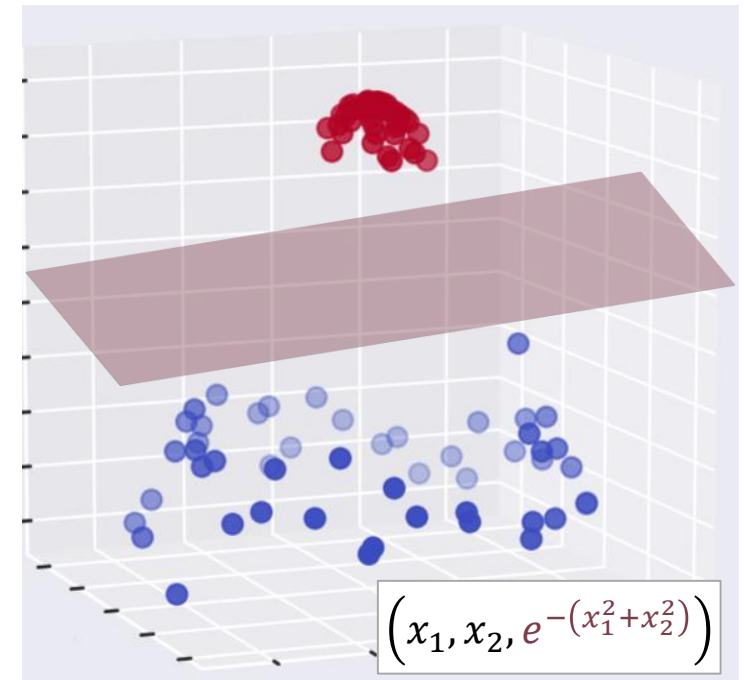
ترفند کرنل و دسته‌بندی غیرفقطی

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.

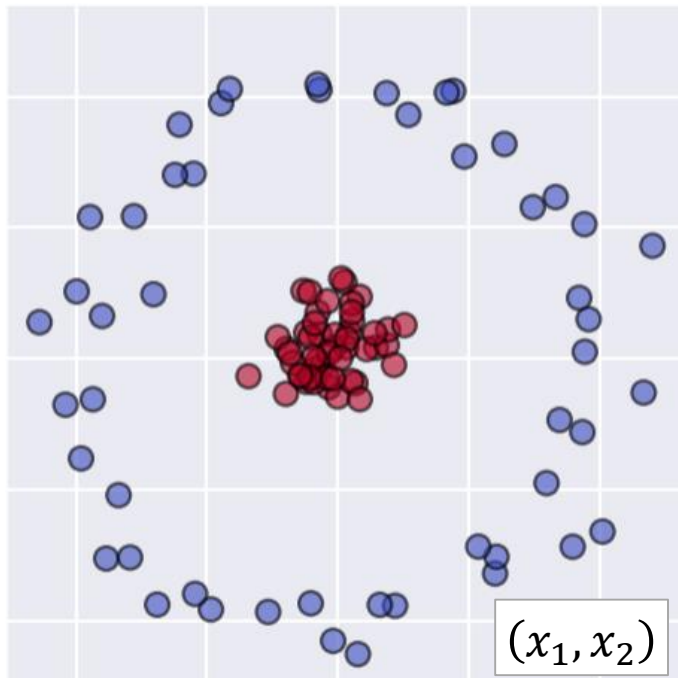


$$\Phi: x \rightarrow \varphi(x)$$

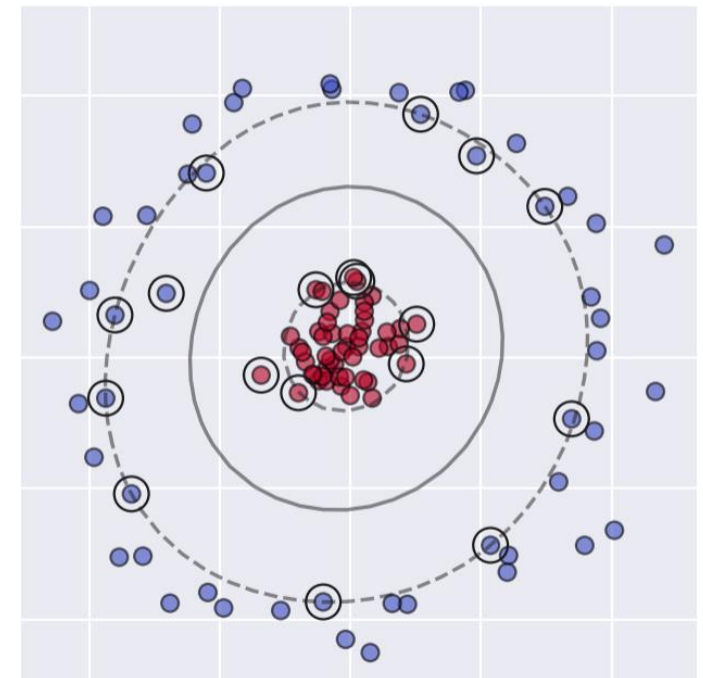


توابع کرنل

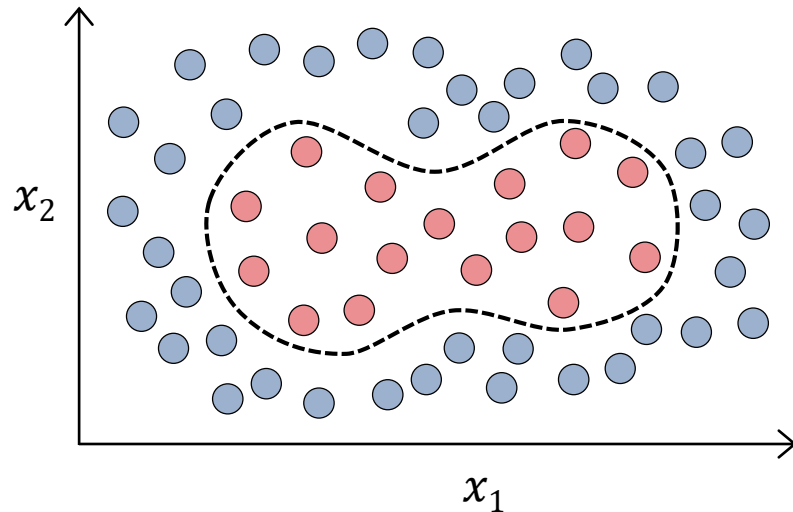
- ایده. نگاشت مسئله به یک فضای ویژگی جدید با استفاده از تبدیلات غیرخطی.
- استفاده از یک مدل خطی در فضای جدید به منظور دسته‌بندی داده‌ها.
- مدل خطی در فضای جدید متناظر با یک مدل غیرخطی در فضای اصلی است.



$$\Phi: x \rightarrow \varphi(x)$$



مرز تصمیم‌گیری غیرخطی



□ پیش‌بینی. $y = 1$ اگر:

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

□ ویژگی‌ها.

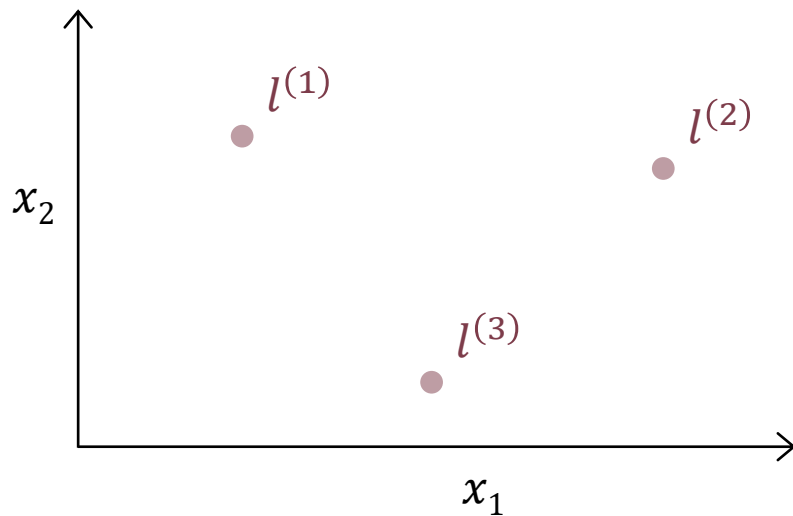
$$f_1 = x_1, \quad f_2 = x_2, \quad f_3 = x_1^2, \quad f_4 = x_2^2, \quad f_5 = x_1x_2, \quad \dots$$

$$h(f) = b + w_1f_1 + w_2f_2 + w_3f_3 + w_4f_4 + w_5f_5 + \dots \leftarrow \text{مرز تصمیم‌گیری فطی}$$

□ 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.



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

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

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

کرنل (کرنل گوسی)

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

کرنل‌ها به عنوان معیار شباهت

□ تابع کرنل.

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

□ حالت اول. $x \approx l^{(i)}$

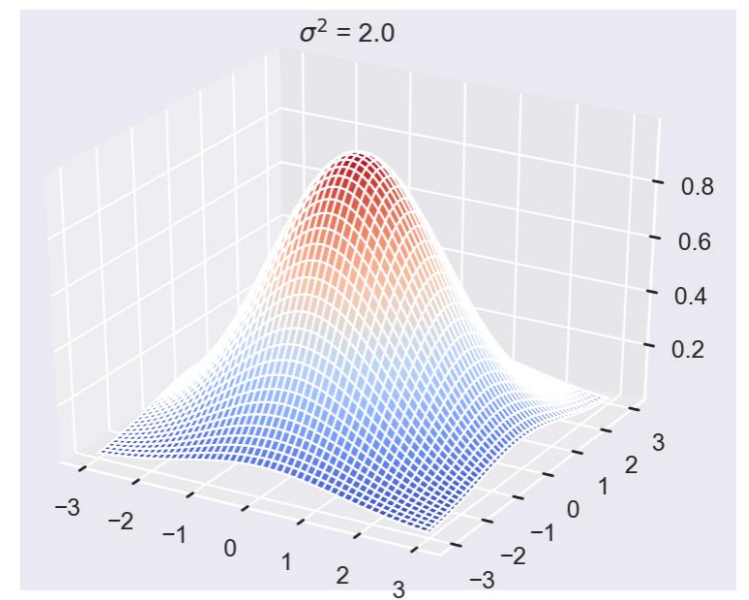
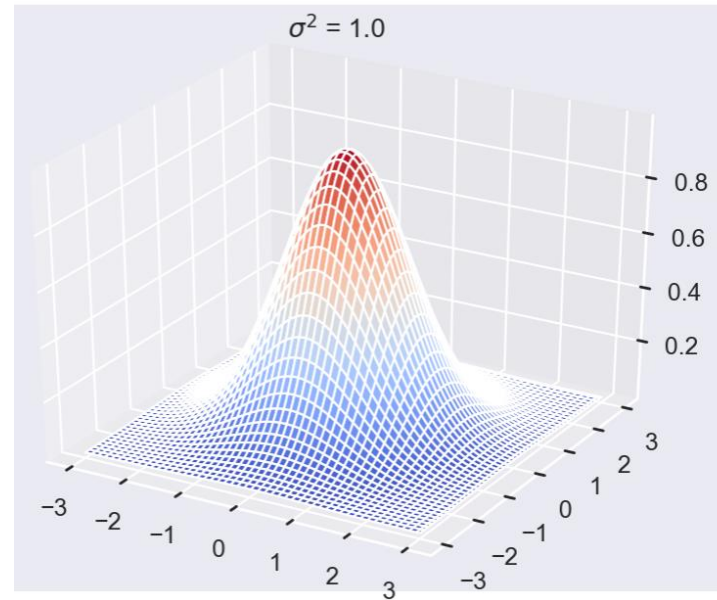
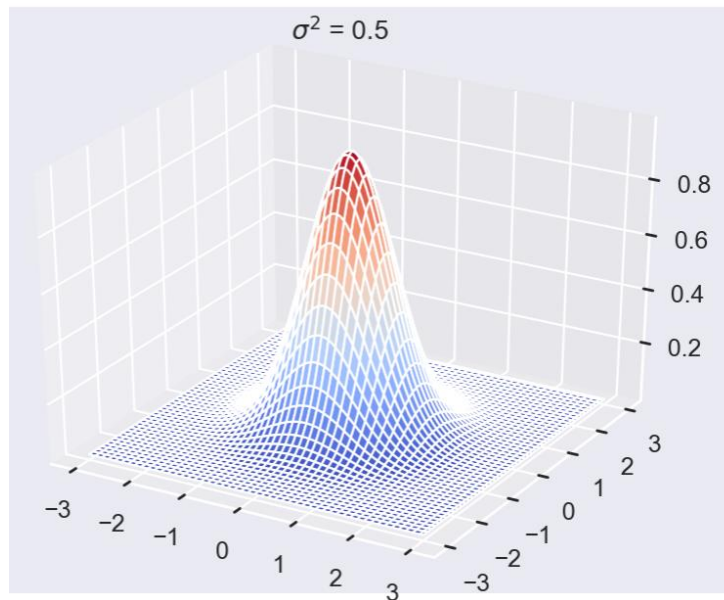
$$f_i \approx \exp\left(-\frac{0}{2\sigma^2}\right) = \exp(0) = 1$$

□ حالت دوم. x بسیار دور از $l^{(i)}$

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

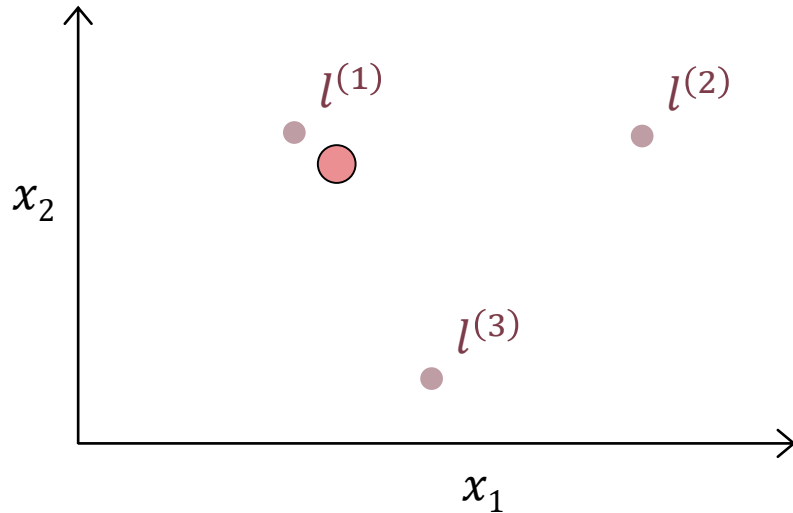
کرنل‌ها به عنوان معیار شباهت

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



کرنل‌ها به عنوان معیار شباهت

□ پیش‌بینی. $y = 1$ اگر:



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

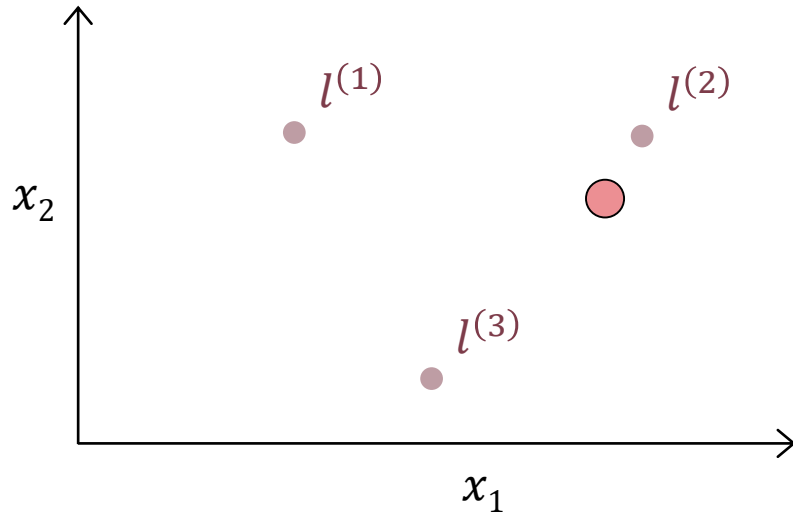
$$\begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ -0.5 & 1.0 & 1.0 & 0.0 \end{array}$$

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

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

کرنل‌ها به عنوان معیار شباهت

□ پیش‌بینی. $y = 1$ اگر:



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

$$\begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ -0.5 & 1.0 & 1.0 & 0.0 \end{array}$$

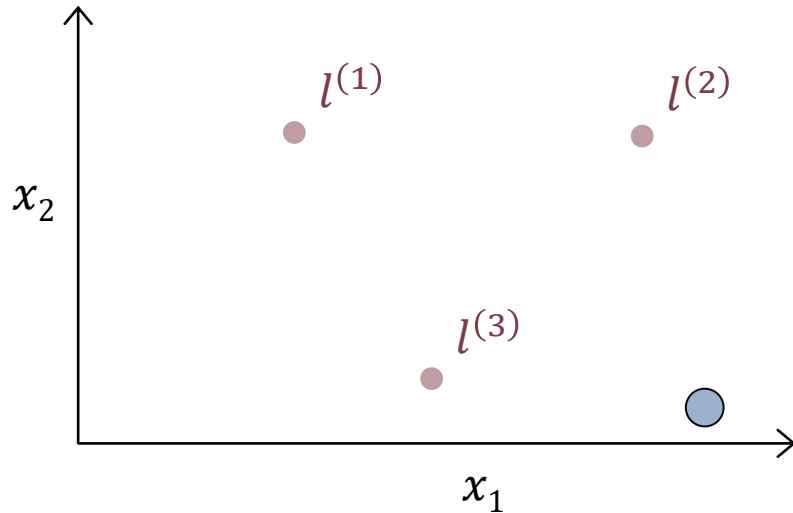
$$f_1 \approx f_3 \approx 0, f_2 \approx 1$$

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

کرنل‌ها به عنوان معیار شباهت

۲۹

□ پیش‌بینی. $y = 1$ اگر:



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

$$\begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ -0.5 & 1.0 & 1.0 & 0.0 \end{array}$$

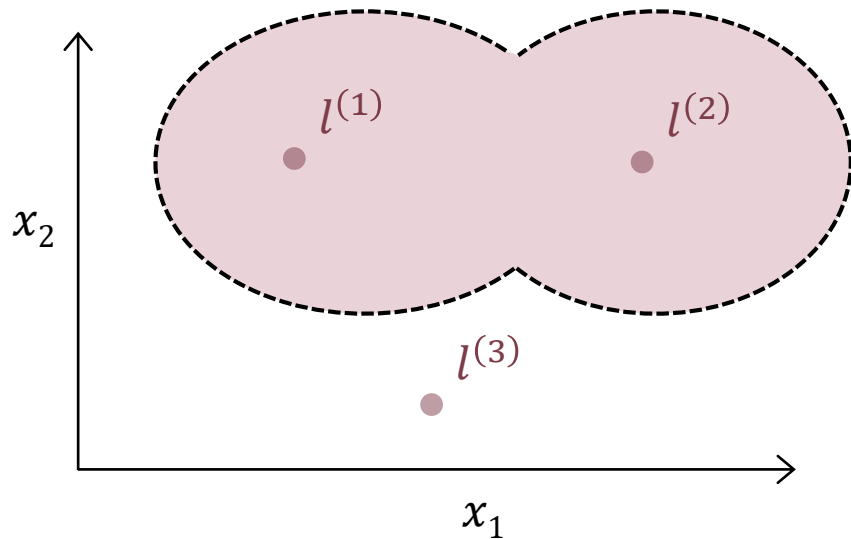
$$f_1 \approx f_2 \approx f_3 \approx 0$$

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

کرنل‌ها به عنوان معیار شباهت

۳۰

□ پیش‌بینی. $y = 1$ اگر:



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

$$\begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ -0.5 & 1.0 & 1.0 & 0.0 \end{array}$$

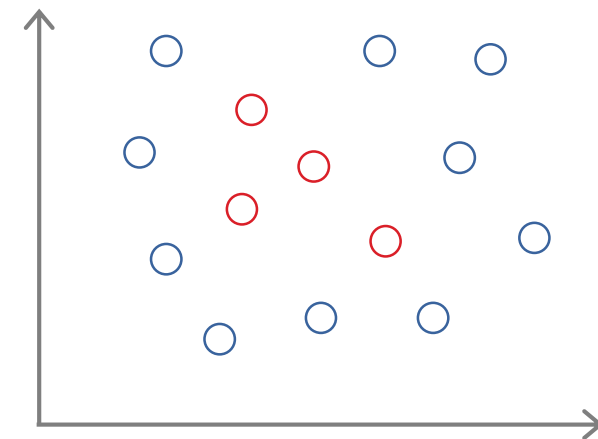
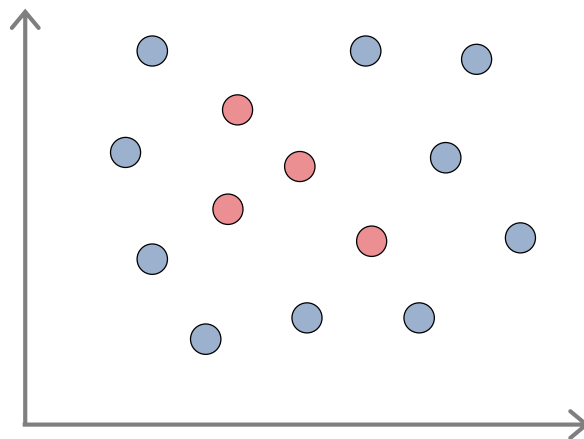
□ مرز تصمیم‌گیری. نقاط نزدیک به $l^{(1)}$ و $l^{(2)}$ را در دسته ۱ و سایر نقاط را در دسته صفر دسته‌بندی می‌کند.

چند پرسش

- الگوریتم یادگیری نقاط راهنما را چگونه به صورت خودکار انتخاب می کند؟
- مقدار مناسب برای پارامترهای تابع کرنل چگونه تعیین می شوند؟
- آیا انواع دیگری از کرنل ها وجود دارد؟

انتخاب نقاط راهنما

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



نگاشت ویژگی‌ها

۳۳

□ مجموعه آموزشی.

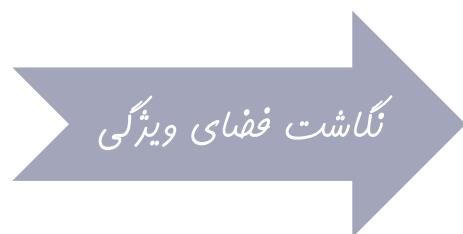
$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$$

□ نقاط راهنما.

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

□ نگاشت فضای ویژگی.

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



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

ترفند کرنل

۳۴

□ تابع کرنل. پیش پردازش داده x با استفاده از توابع کرنل:

$$\begin{aligned} \mathbf{z} &= \varphi(\mathbf{x}) \\ &= (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \dots, \varphi_k(\mathbf{x})) \end{aligned}$$

ممکن است بینهایت باشد!!!

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

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

□ مرز تصمیم گیری.

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

□ دسته بندی داده جدید.

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

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

s.t. $y^t \mathbf{w}^T \varphi(\mathbf{x}^t) \geq 1 - \varepsilon^t$

$$\varepsilon^t \geq 0$$

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

ضرایب لاگرانژ

ضرایب لاگرانژ

توابع کرنل: مسئله اصلی

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

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

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

توابع کرنل: مسئله دوگان

۳۷

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

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

□ ایده ماشین‌های کرنل. [ترفند کرنل]

□ جایگزینی حاصل ضرب داخلی توابع پایه با یک تابع کرنل به صورت $K(\mathbf{x}^t, \mathbf{x}^s)$

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

ماتریس گرام: یک ماتریس متقارن و مثبت معین (برای تفکیک‌پذیری قطعی)

توابع کرنل: چند جمله‌ای

□ کرنل چند جمله‌ای. یک چند جمله‌ای از درجه q .

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

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

□ مثال. $[q = 2, d = 2]$

$$K(x, y) = (x^T y + 1)^2$$

$$= (x_1 y_1 + x_2 y_2 + 1)^2$$

$$= 1 + 2x_1 y_1 + 2x_2 y_2 + 2x_1 x_2 y_1 y_2 + x_1^2 y_1^2 + x_2^2 y_2^2$$

۳ ضرب، ۲ جمع

۶ ضرب، ۵ جمع

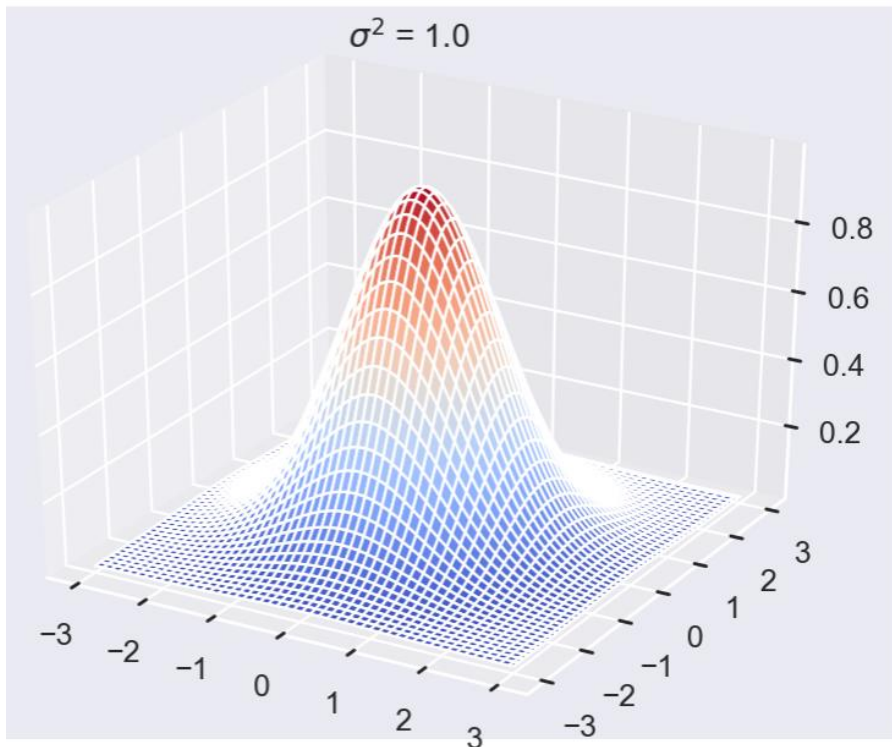
$$\varphi(x) = [1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2, x_1^2, x_2^2]^T$$

$$\varphi(y) = [1, \sqrt{2}y_1, \sqrt{2}y_2, \sqrt{2}y_1 y_2, y_1^2, y_2^2]^T$$

توابع کرنل: کرنل گوسی

۳۹

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



□ کرنل گوسی.

□ یافتن یک مقدار مناسب برای σ .

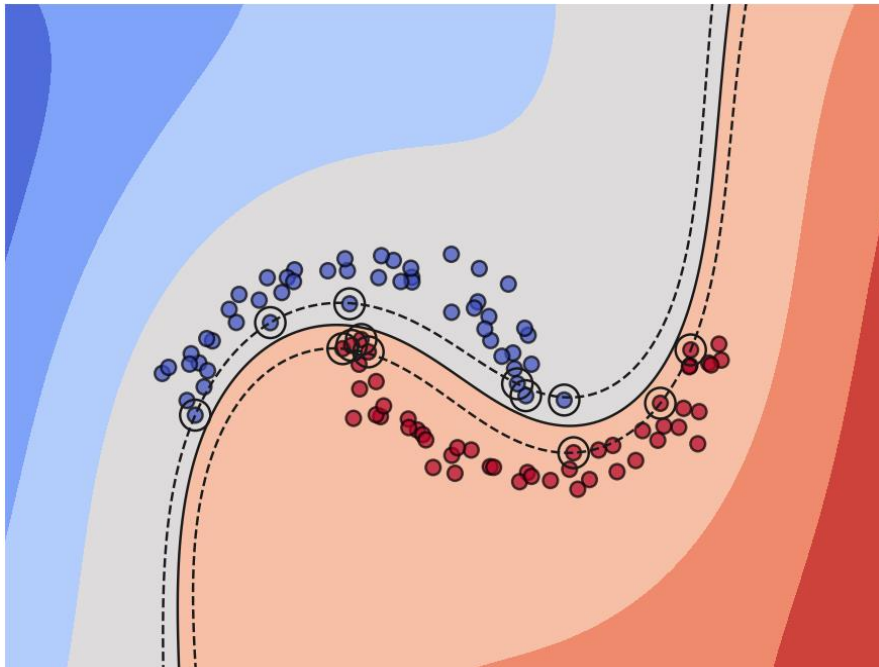
□ با استفاده از مجموعه اعتبارسنجی [انتخاب مدل]

□ مقادیر بزرگتر: مرز تصمیم‌گیری هموارتر

توابع کرنل: کرنل گوسی

۴۰

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



□ کرنل گوسی.

□ یافتن یک مقدار مناسب برای σ .

□ با استفاده از مجموعه اعتبارسنجی [انتخاب مدل]

□ مقادیر بزرگتر: مرز تصمیم‌گیری هموارتر

ابر پارامترها: انتخاب مدل

□ راه حل اول. [یک ایده بسیار بد]

□ انتخاب مقداری که منجر به بالاترین دقت دسته‌بندی بر روی **مجموعه آزمایشی** می‌شود.

داده‌های آموزشی	داده‌های آزمایشی
-----------------	------------------

□ توجه. [بسیار مهم]

□ از **مجموعه آزمایشی** در انتهای مراحل و تنها برای **تخمین قابلیت تعمیم** دسته‌بند استفاده کنید.

ابَر پارامترها: انتخاب مدل

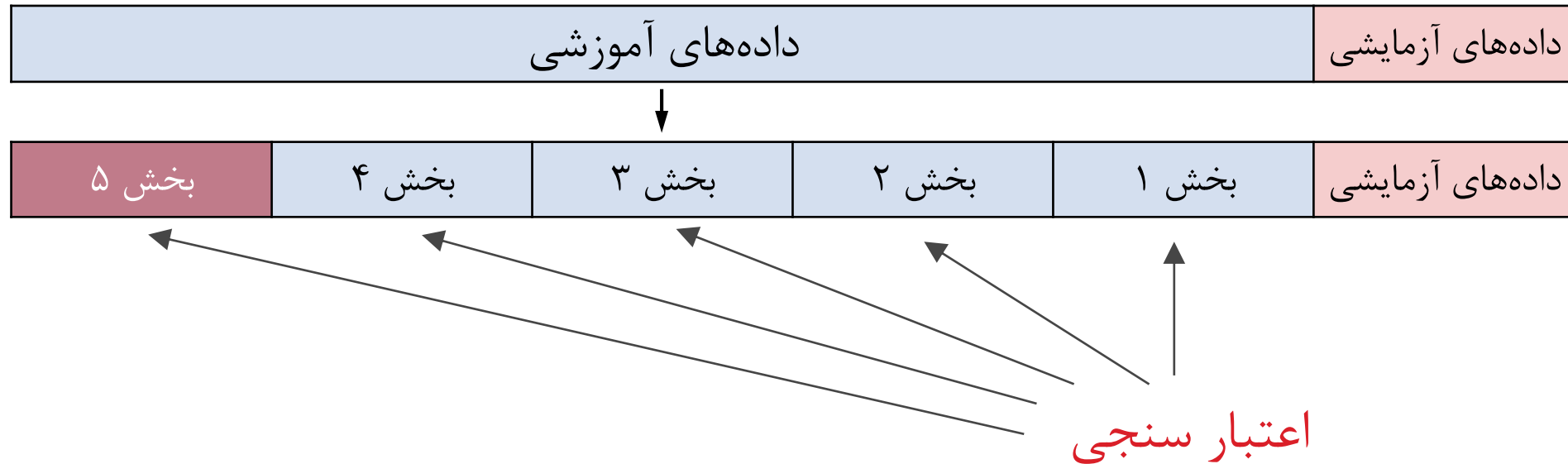
□ راه حل دوم. [اعتبارسنجی چند بخشی]

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

داده‌های اعتبارسنجی [برای تعیین مقدار ابرپارامترها]

ابزار پارامترها: انتخاب مدل

□ راه حل دوم. [اعتبارسنجی چند بخشی]



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

ابَر پارامترها: انتخاب مدل

۴۴

```
cv = StratifiedKFold(n_splits=5, shuffle=True)
```

```
C_range = np.logspace(-3, 5, 9)
```

```
gamma_range = np.logspace(-3, 5, 9)
```

```
pgrid = dict(gamma=gamma_range, C=C_range)
```

```
grid = GridSearchCV(SVC(), param_grid=pgrid, cv=cv)
```

```
grid.fit(X, y)
```

تقسیم داده‌های آموزشی
به داده‌های آموزشی و
داده‌های اعتبارسنجی

ابری پارامترها: انتخاب مدل

۴۵

```
cv = StratifiedKFold(n_splits=5, shuffle=True)
```

```
C_range = np.logspace(-3, 5, 9)
```

```
gamma_range = np.logspace(-3, 5, 9)
```

```
pgrid = dict(gamma=gamma_range, C=C_range)
```

```
grid = GridSearchCV(SVC(), param_grid=pgrid, cv=cv)
```

```
grid.fit(X, y)
```

مشخص کردن بازه
جستجو برای ابری پارامترها



$\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5\}$

ابَر پارامترها: انتخاب مدل

۴۶

```
cv = StratifiedKFold(n_splits=5, shuffle=True)
```

```
C_range = np.logspace(-3, 5, 9)
```

```
gamma_range = np.logspace(-3, 5, 9)
```

```
pgrid = dict(gamma=gamma_range, C=C_range)
```

```
grid = GridSearchCV(SVC(), param_grid=pgrid, cv=cv)
```

```
grid.fit(X, y)
```

ایجاد دسته‌بند

ابَر پارامترها: انتخاب مدل

۴۷

```
cv = StratifiedKFold(n_splits=5, shuffle=True)

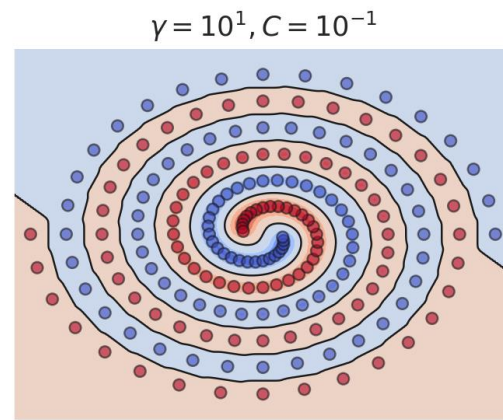
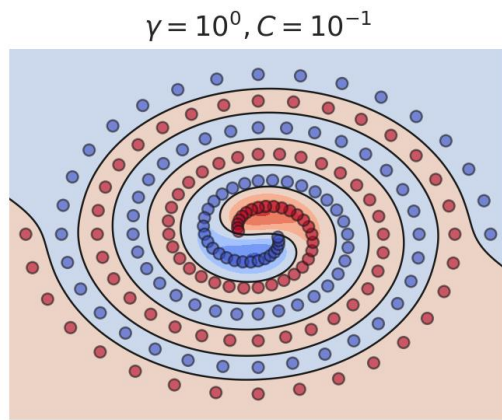
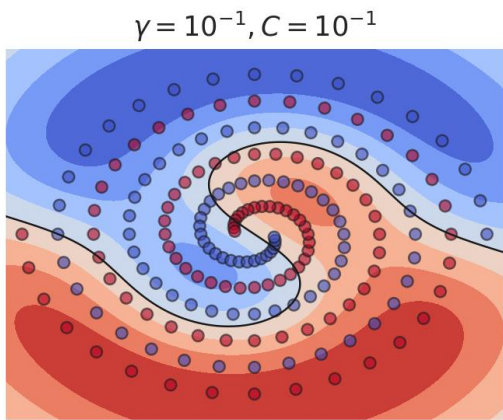
C_range = np.logspace(-3, 5, 9)
gamma_range = np.logspace(-3, 5, 9)

pgrid = dict(gamma=gamma_range, C=C_range)
grid = GridSearchCV(SVC(), param_grid=pgrid, cv=cv)

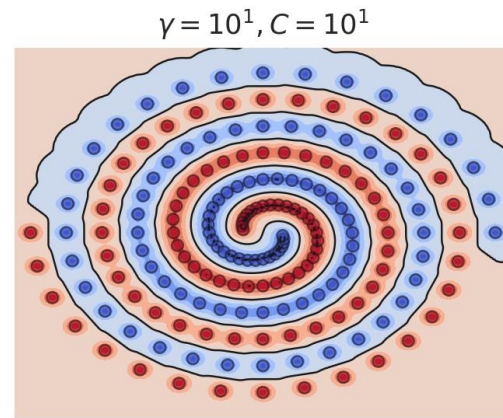
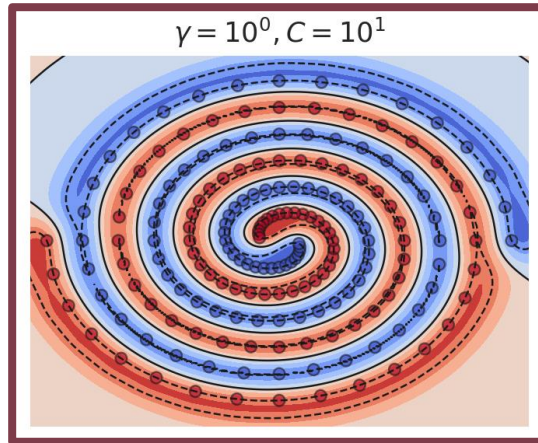
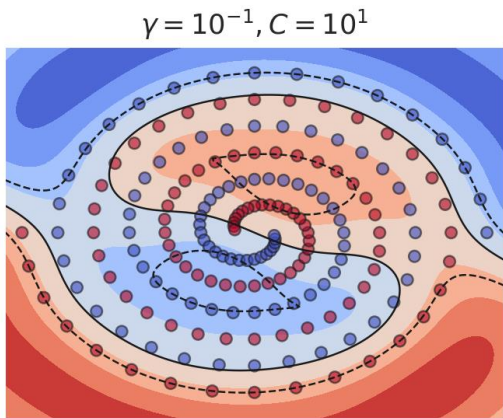
grid.fit(X, y)
```

آموزش دسته‌بند

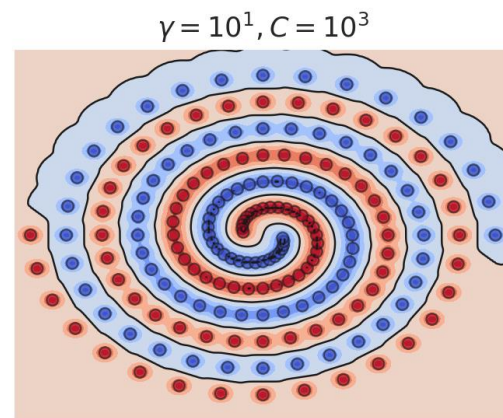
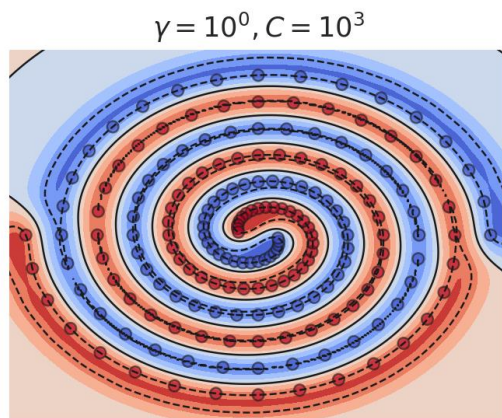
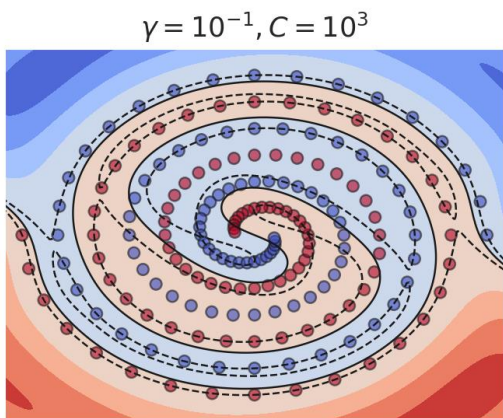
جستجوی توری



$C = 10^{-1}$



$C = 10^1$



$C = 10^3$

$\gamma = 10^{-1}$

$\gamma = 10^0$

$\gamma = 10^1$

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

دقت دسته‌بندی بر روی داده‌های اعتبارسنجی

راهنمای استفاده از ماشین بردار پشتیبان

- پیاده‌سازی. استفاده از بسته‌های نرم‌افزاری موجود مانند LIBSVM و SVM^{light}
- تعیین تابع کرنل.
 - کرنل خطی (عدم استفاده از کرنل): وقتی که $n \gg m$
 - گوسی، چند جمله‌ای، 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

۵۱

□ حالت ۱. $[n \gg m]$

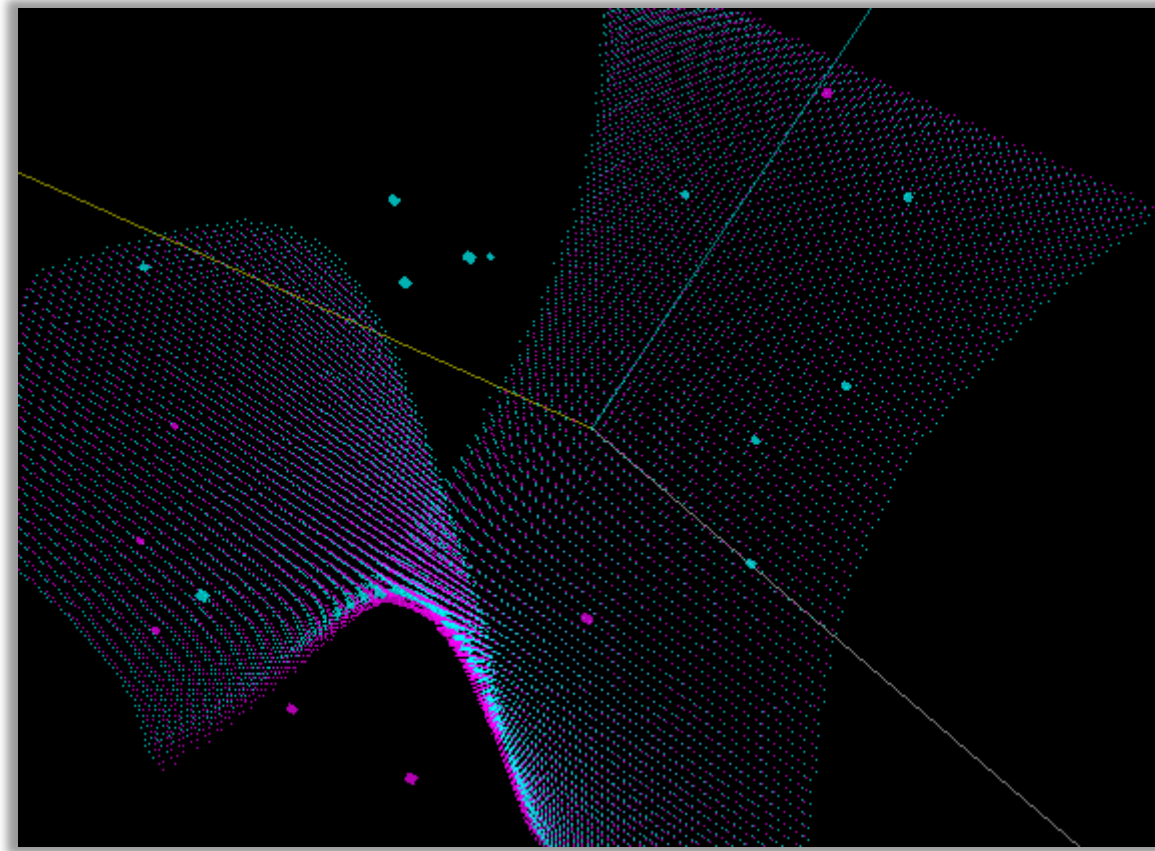
□ مثال: تشخیص هرزنامه (۱۰۰۰ نمونه آموزشی، ۵۰۰۰۰ ویژگی)

□ 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.



<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 $\phi(x^{(i)})$

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

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

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

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