به‌نام خدا
تئوری یادگیری ماشین
دکتر سیدصالحی
جلسه چهاردهم

دانشکده ریاضی و علوم کامپیوتر
اردیبهشت ماه ۱۴۰۳

# ۱ ماشین های بردار پشتیبان

$SVM$ یک الگوریتم یادگیری ماشین نظارت شده است که برای وظایف طبقه‌بندی و رگرسیون استفاده می‌شود. این فصل چندین مفهوم کلیدی و فرمول‌بندی‌های مرتبط با $SVM$ ها را تشریح می‌کند. ابتدا مثالی از مسئله بهینه سازی سر کلاس دیدیم . ما داستان ماشین‌های بردار پشتیبان $SVMs$ را با صحبت درباره حاشیه‌ها آغاز می‌کنیم. این بخش، بینشی درباره حاشیه‌ها و درباره «اطمینان» پیش‌بینی‌های ما خواهد داد. میخواهیم خطی را پیدا کنیم که احتمالا در آینده قدرت تعمیم[1] بیشتری خواهد داشت ، ذدر این مرحله حاشیه[2] داریم و هدف ما بیشینه کردن حاشیه یا به نوعی بیشترین حاشیه است . در ادامه شهود ما نسبت به بیشینه سازی حاشیه و شرط خود را نوشتیم . مسئله ی محدب [3] ما یک جواب دارد که به این نوع مسائل ماشین های بردار پشتیبان حاشیه سخت [4] گفته می شود .

## فهرست



[1] generalization
[2] margin
[3] convex
[4] hard margin svm

### ۱-۱ مفهوم حاشیه

حاشیه یک صفحه‌ی تصمیم که نمونه‌های دو کلاس قابل تفکیک خطی را جدا می‌کند، کوچکترین فاصله بین مرز تصمیم و هر یک از نمونه‌های آموزشی است. حاشیه بزرگتر معمولاً بهتر به داده‌های دیده نشده تعمیم می‌یابد.

## ۲ فرمولاسیون دوگانه

[۵]

فرمولاسیون دوگانه ماشین بردار پشتیبان $SVM$ یک روش جایگزین برای حل مسئله اصلی اولیه است. این فرمولاسیون بینش بیشتری به ابرصفحه [۶] بهینه می دهد و استفاده از ترفند کرنل را امکان پذیر می کند.

### ۱-۲ تکنیک ضرایب لاگرانژ

فرمولاسیون دوگانه بر اساس تکنیک مولتی پلیر لاگرانژ است، که یک روش بهینه سازی برای حل مشکلات با محدودیت های مساوی یا نا مساوی است. این تکنیک شامل معرفی متغیرهای جدید به نام مولتی پلیر لاگرانژ است که مسئله بهینه سازی محدود را به یک مسئله بهینه سازی بدون محدودیت تبدیل می کند.

### ۲-۲ مزایای کلیدی

فرمولاسیون دوگانه SVM چندین مزیت را ارائه می دهد، از جمله:

- بینش بیشتری به ابرصفحه بهینه می دهد
- استفاده از ترفند کرنل را امکان پذیر می کند که برای طبقه بندی غیر خطی استفاده می شود
- یک راه حل جایگزین برای مسئله اصلی اولیه را ارائه می دهد که در برخی از موقعیت ها مفید است.

### حاشیه سخت

مسئله بهینه‌سازی: هدف یافتن یک صفحه‌ی تصمیم است که حاشیه بین دو کلاس را به حداکثر برساند.

[5] Dual formulation of the SVM
[6] hyper-plane

$$\max_{M,\mathbf{w},w_0} \frac{2M}{||\mathbf{w}||}$$

$$\text{subject to}$$

$$y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq M \quad \forall i$$

تغییر مقیاس پارامترها: این مسئله بهینه‌سازی را با ثابت کردن حاشیه به ۱ ساده می‌کند.

$$\max_{\mathbf{w}',w_0'} \frac{2}{||\mathbf{w}'||}$$

$$\text{subject to}$$

$$y_i(\mathbf{w}'^T\mathbf{x}_i + w_0') \geq 1 \quad \forall i$$

فرمول‌بندی معادل:

$$\min_{\mathbf{w},w_0} \frac{1}{2}||\mathbf{w}||^2$$

$$\text{subject to}$$

$$y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1 \quad \forall i$$

## فرمول‌بندی دوگانه $SVM$ حاشیه سخت

تکنیک ضرایب لاگرانژ: برای مسائل با محدودیت‌های برابری یا نابرابری مفید است.

$$\begin{aligned} &\min_x \quad f(x) \\ &\text{s.t.} \quad g_i(x) \leq 0, \ i = 1, ..., m \end{aligned}$$

می‌توانیم تابع لاگرانژی زیر را بسازیم:

$$\mathcal{L}(x, a) = f(x) + \sum_{i=1}^{m} a_i g_i(x)$$

که در آن $\mathbf{x}$ متغیر بهینه‌سازی است، $f(\mathbf{x})$ تابع هدف و $g_i(\mathbf{x})$ توابع قیود هستند.
که در آن $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$ ضرایب لاگرانژ هستند.
سپس بهینه‌سازی می‌کنیم:

$$\begin{aligned} &\min_{\mathbf{x}} \max_{\boldsymbol{\alpha} \geq 0} \mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}) \\ &\max_{\boldsymbol{\alpha} \geq 0} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}) \end{aligned} \tag{۱}$$

در ادامه داریم:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}|w|^2 \\ &\text{subject to} \quad y^{(i)}(w^T x^{(i)} + w_0) \geq 1 \quad \text{for } i = 1, \ldots, N \end{aligned}$$

- $\mathbf{w}$: weight vector
- $w_0$: bias term
- $\mathbf{x}_i$: input vectors
- $y_i$: corresponding labels
- $N$: number of samples

برای ترکیب قیود [۷] در تابع هدف، مولتی‌پلیر لاگرانژ $\alpha_n \geq 0$ را معرفی می‌کنیم و تابع لاگرانژ را تعریف می‌کنیم:

$$\min_{w, w_0} \max_{a_n \geq 0} \left( \frac{1}{2}||w||^2 + \sum_{n=1}^{N} a_n (1 - y^{(n)}(w^T x^{(n)} + w_0)) \right) \tag{۲}$$

مسئله دوگانه با تعویض ترتیب عملیات مینیمم و ماکسیمم به دست می‌آید که معادل است با:

$$\max_{a_n \geq 0} \min_{w, w_0} \left( \frac{1}{2}||w||^2 + \sum_{n=1}^{N} a_n (1 - y^{(n)}(w^T x^{(n)} + w_0)) \right) \tag{۳}$$

[7] constraint

تابع لاگرانژ تابع هدف اصلی را با توابع قیود ترکیب می‌کند که توسط مولتی‌پلیر لاگرانژ وزن می‌شوند.

$$\max_{\alpha_n \geq 0} \min_{w, w_0} \mathcal{L}(w, w_0, \alpha) \tag{۴}$$

$$\mathcal{L}(w, w_0, \alpha) = \frac{1}{2}||w||^2 + \sum_{n=1}^{N} \alpha_n (1 - y^{(n)}(w^T x^{(n)} + w_0)) \tag{۵}$$

$$\nabla_w \mathcal{L}(w, w_0, \alpha) = 0 \Rightarrow w = \sum_{n=1}^{N} \alpha_n y^{(n)} x^{(n)} \tag{۶}$$

$$\frac{\partial}{\partial w_0} \mathcal{L}(w, w_0, \alpha) = 0 \Rightarrow \sum_{n=1}^{N} \alpha_n y^{(n)} = 0 \tag{۷}$$

این مسئله معادل مسئله اصلی است. مسئله ماکسیمم داخلی برای یافتن مولتی‌پلیر لاگرانژ بهینه است که محدودیت‌ها را برآورده می‌کند، در حالی که مسئله مینیمم خارجی برای یافتن مقادیر بهینه $\mathbf{w}$ و $w_0$ است که تابع لاگرانژ را به حداقل می‌رسانند.

$$w = \sum_{n=1}^{N} a_n y^{(n)} x^{(n)} \sum_{n=1}^{N} a_n y^{(n)} = 0 \tag{۸}$$

$$\mathcal{L}(w, w_0, \alpha) = \frac{1}{2} w^T w + \sum_{n=1}^{N} \alpha_n (1 - y^{(n)}(w^T x^{(n)} + w_0)) \tag{۹}$$

$$\mathcal{L}(\alpha) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{m=1}^{N} \sum_{n=1}^{N} \alpha_n \alpha_m y^{(m)} y^{(n)} x^{(m)T} x^{(n)} \tag{۱۰}$$

$$\text{Maximize with respect to } \alpha \quad \sum_{n=1}^{N} \alpha_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y^{(n)} y^{(m)} x^{(n)T} x^{(m)}$$

$$\text{subject to} \quad \alpha_n \geq 0 \quad \text{for } n = 1, ..., N$$

$$\sum_{n=1}^{N} \alpha_n y^{(n)} = 0$$

این معادلات بخشی از فرمول‌بندی برای مسئله بهینه‌سازی ماشین بردار پشتیبان (SVM) هستند. معادله اول بردار وزن ($w$) را بر حسب ضریب‌های لاگرانژ ($\alpha_n$)، برچسب‌ها ($y^{(n)}$) و بردارهای ورودی ($x^{(n)}$) تعریف می‌کند. معادله دوم شروط را نشان می‌دهد که مجموع حاصل‌ضرب مولتی‌پلیرهای لاگرانژ و برچسب‌ها باید برابر با صفر باشد. معادله سوم لاگرانژ ($L$) است که باید نسبت به ضریب‌های لاگرانژ ($\alpha$) حداکثر شود، مشروط به قیدهایی که در آخرین معادله داده شده است.

$$\begin{aligned}
&\max_{\alpha} \left( \sum_{n=1}^{N} \alpha_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y^{(n)} y^{(m)} (x^{(n)})^T x^{(m)} \right) \\
&subject\ to \sum_{n=1}^{N} \alpha_n y^{(n)} = 0, \\
&\alpha_n \geq 0, \quad n = 1, \ldots, N
\end{aligned} \tag{۱۱}$$

مسئله دوگانه یک مسئله $QP$ محدب است که می‌تواند با استفاده از الگوریتم‌های بهینه‌سازی مختلف به‌طور موثر حل شود.

: $Quadratic\, programming$

$$\min_{\alpha} \frac{1}{2}\alpha^T \begin{bmatrix} y^{(1)}y^{(1)}x^{(1)T}x^{(1)} & \dots & y^{(1)}y^{(N)}x^{(1)T}x^{(N)} \\ \vdots & \ddots & \vdots \\ y^{(N)}y^{(1)}x^{(N)T}x^{(1)} & \dots & y^{(N)}y^{(N)}x^{(N)T}x^{(N)} \end{bmatrix} \alpha + (-1)^T \alpha$$

$$\text{s.t.} \quad -\alpha \leq 0, \quad y^T \alpha = 0$$

$$w = \sum_{n=1}^{N} \alpha_n y^{(n)} x^{(n)}$$

- شروط مهمی که باید برقرار باشد [$w^*, w_0^*, \alpha^*$]:

۱. $\alpha_n^* \geq 0, \quad n = 1, ..., N$

۲. $y^{(n)}(w^{*T}x^{(n)} + w_0^*) \geq 1, \quad n = 1, ..., N$

۳. $\alpha_n^*(1 - y^{(n)}(w^{*T}x^{(n)} + w_0^*)) = 0, \quad n = 1, ..., N$

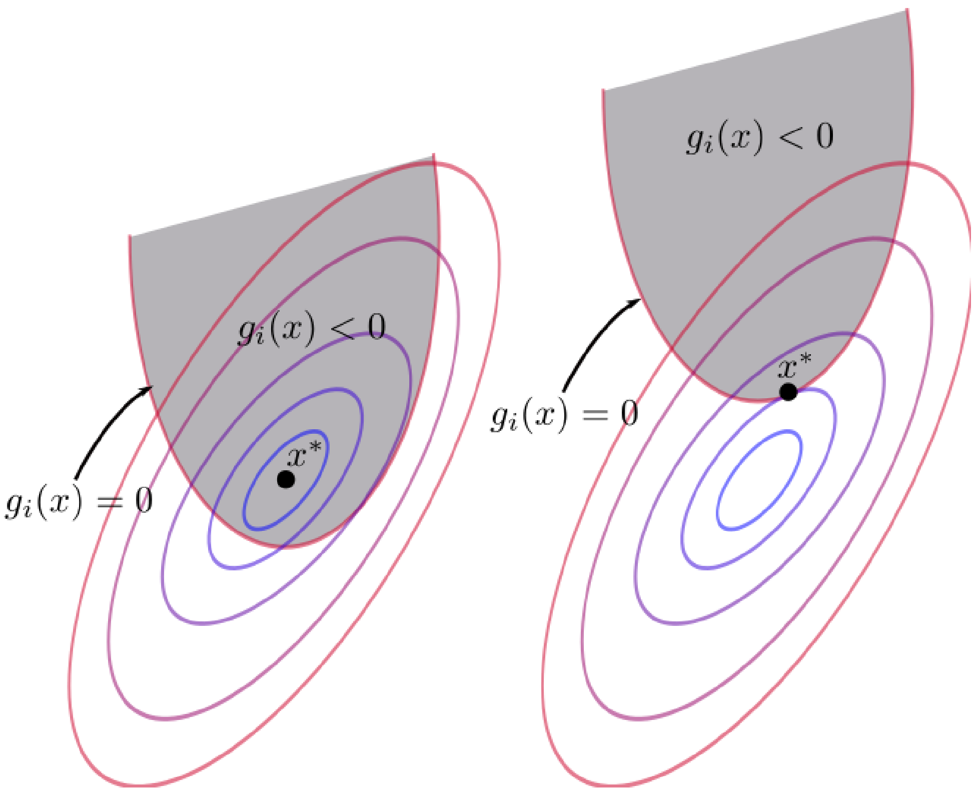
نقاط با $\alpha \neq 0$ نقاطی هستند که حتما روی خط می‌افتند.

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$
$$\text{s.t. } g_i(\boldsymbol{x}) \leq 0 \quad i = 1, \dots, m$$

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\alpha}) = f(\boldsymbol{x}) + \sum \alpha_i \, g_i(\boldsymbol{x})$$

In general, the optimal $\boldsymbol{x}^*, \boldsymbol{\alpha}^*$ satisfies KKT conditions:

$$\alpha_i^* \geq 0 \quad i = 1, \dots, m$$
$$g_i(\boldsymbol{x}^*) \leq 0 \quad i = 1, \dots, m$$
$$\alpha_i^* g_i(\boldsymbol{x}^*) = 0 \quad i = 1, \dots, m$$




## ۲-۳ قیود غیرفعال:

این قیودی است که بر حل بهینه مسئله تأثیر نمی‌گذارد. این قیود در سمت چپ اسلاید شما نمایش داده شده است. شرط برای قیود غیرفعال این است که در نقطه بهینه $(x^*)$ برابر با صفر نباشد، که به صورت ریاضی به صورت $g_i(x) \neq 0$ نمایش داده می‌شود. در نمودار، نقطه بهینه $(x^*)$ در داخل منطقه مجاز اما نه

بر روی مرز تعریف شده توسط قیود قرار دارد، و ضریب لاگرانژ مربوطه () برابر با صفر است.

$$y^{(n)}(w^T x^{(n)} + w_0) > 1$$

$$\Rightarrow \alpha_n = 0, x^{(n)}$$

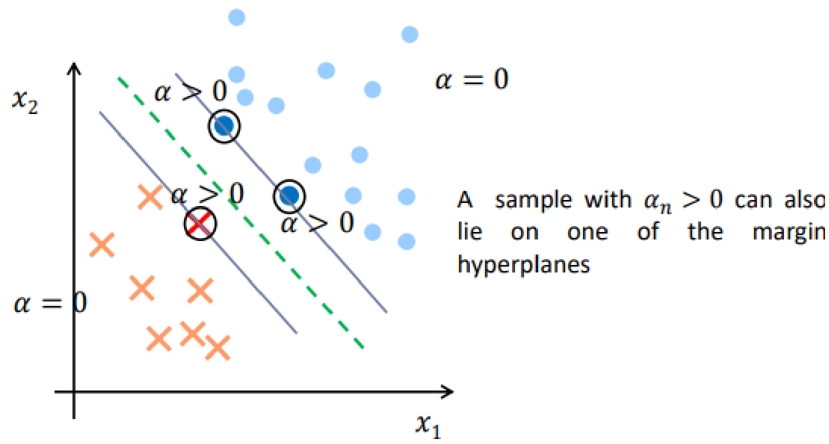
بردار پشتیبان نیست .

## ۴-۲ قیود فعال:

این قیودی است که بر حل بهینه مسئله تأثیر مستقیم دارد. این قیود در سمت راست اسلاید شما نمایش داده شده است. برای قیود فعال، شرط این است که در نقطه بهینه $(x^*)$ برابر با صفر باشد، که به صورت ریاضی به صورت $(g_i(x) = 0)$ نمایش داده می‌شود. اینجا، نقطه بهینه $(x^*)$ بر روی مرز منطقه مجاز قرار دارد، که لبه تعریف شده توسط قیود است.

$$y^{(n)}(w^T x^{(n)} + w_0) = 1$$

نقاطی که روی مارجین می افتند برای ما قیود فعال هستند که می توانند جای بهینه را عوض کنند . (همانطور که در فرمول مشخص است )




$(SVs) = x^{(n)}|\alpha_n > 0$ این بدان معنی است که یک بردار پشتیبان یک نقطه داده $x^{(n)}$ است که ضریب لاگرانژ مربوطه برای آن بزرگتر از صفر است.

جهت ابر صفحه (مرز تصمیم) را می توان تنها بر اساس بردارهای پشتیبانی تعیین کرد. این به این دلیل

است که بردارهای پشتیبان، نقاط داده ای هستند که نزدیک ترین نقطه به ابر صفحه قرار دارند و بیشترین تأثیر را در جهت آن دارند.

جهت ابرصفحه $W$ را می توان با استفاده از فرمول زیر محاسبه کرد:

$$w = \sum_{\alpha_n > 0} \alpha_n y^{(n)} x^{(n)}$$

پس از حل مسئله برنامه نویسی درجه دوم (□□) برای یافتن ضرایب لاگرانژ، می توانیم بردار وزن را با استفاده از فرمول زیر پیدا کنیم:

$$w = \sum_{n=1}^{N} \alpha_n y^{(n)} x^{(n)}$$

هر یک از نمونه هایی که دارای $\alpha_s > 0$ هستند در حاشیه قرار می گیرند ، به این معنی که آنها به ابرصفحه نزدیک ترین هستند. بنابراین، ما می‌توانیم از هر یک از این $SV$ ها برای حل $w_0$ استفاده کنیم.

$$y(s)(\mathbf{w}^T \mathbf{x}^(s) + w_0) = 1$$

$$\Rightarrow w_0 = y^{(s)} - \mathbf{w}^T \mathbf{x}^{(s)}$$

## ۲-۵ طبقه بندی نمونه های جدید

هنگام طبقه‌بندی یک نمونه جدید، $SVM$ با حاشیه سخت از فرمول زیر استفاده می‌کند:

$$\hat{y} = sign(w_0 + w^T x)$$

این نشان دهنده تابع تصمیم گیری یک $SVM$ است. برچسب پیش بینی شده $(\hat{y})$ را محاسبه می کند.

$$\hat{y} = sign\left(w_0 + \left(\sum_{a_n > 0} a_n y^{(n)} x^{(n)}\right)^T x\right)$$

تابع تصمیم‌گیری به شکل گسترش یافتهبه صورت بالا است ، مجموع بر روی همه بردارهای پشتیبان (نمونه‌هایی که$a_n > 0$) اجرا می‌شود.

$$\hat{y} = sign\left(y^{(s)} - w_0 - \sum_{a_n > 0} a_n y^{(n)} (x^{(n)})^T x^{(s)}\right) + \sum_{a_n > 0} a_n y^{(n)} (x^{(n)})^T x$$

با جابجایی و کم و زیاد کردن واریانس کرنل گوشی ،مرز های ما بیشتر و بیشتر جابجا می شوند و اورفیت ایجاد می کنند و دور هر دیتای ترین جمع و غیر خطی می شوند .

$$\max_{\alpha} \left( \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y^{(n)} y^{(m)} \left( x^{(n)} \right)^T x^{(m)} \right)$$

Subject to the constraints:

$$\sum_{n=1}^{N} \quad \alpha_n y^{(n)} = 0$$

and

$$\alpha_n \geq 0 \quad \text{for } n = 1, \ldots, N$$

# ۳ ماشین بردار پشتیبان در مسائل خطی تفکیک ناپذیر($Soft-margin\ SVM$)

در بسیاری از مسائل دنیای واقعی، داده‌ها نمی‌توانند به طور کامل توسط یک مرز خطی جدا شوند. تقریبا جداپذیری خطی به داده‌هایی اشاره دارد که تقریباً می‌توانند توسط یک مرز خطی جدا شوند اما ممکن است چند نمونه به اشتباه طبقه‌بندی شده یا درون حاشیه قرار گیرند. گاهی داده ها انقدر تمیز نیستند و ممکن است داده ها با آنکه به صورت خطی جداپذیر هستند، ولیدر هم رفته اند و کاملا تفکیک شده نباشند. برای همچین حالاتی از $Soft - margin\ SVM$ استفاده می کنیم .

## ۳-۱ مفهوم حاشیه نرم

هدف: حداکثر کردن حاشیه در حالی که اجازه می‌دهد برخی نقاط در طرف اشتباه حاشیه یا حتی اشتباه طبقه‌بندی شوند. متغیرهای شل ($\xi_i$) برای هر نقطه داده برای اندازه‌گیری درجه اشتباه طبقه‌بندی یا نقض حاشیه معرفی شده‌اند.

## ۳-۲ کمینه کردن تعداد نقاط طبقه‌بندی شده اشتباه:

ایده‌آل این است که یک $SVM$ تمام نقاط آموزشی را به طور کامل طبقه‌بندی کند. با این حال، این همیشه ممکن نیست، به ویژه با داده‌های پر سر و صدا. هدف کمینه کردن این طبقه‌بندی‌های اشتباه است.

در این حالت ما اجازه میدهیم برخی از داده ها در آموزش دچار خطا بشوند .

$SVM$ با حاشیه نرم یک متغیر اسلک $(slack\ variable)(\xi)$ را برای هر نقطه داده معرفی می‌کند تا

اجازه دهد برخی از طبقه‌بندی‌های اشتباه وجود داشته باشند. هدف یافتن تعادل بین حداکثر کردن حاشیه (فاصله بین صفحه جداکننده و نزدیک‌ترین نقاط داده از هر کلاس) و کمینه کردن مجموع این متغیرهای اسلک است که نمایانگر جریمه کلی برای هر گونه طبقه‌بندی اشتباه است.

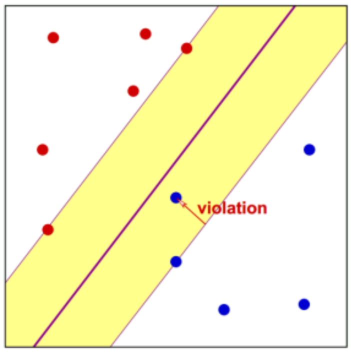
$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i$$

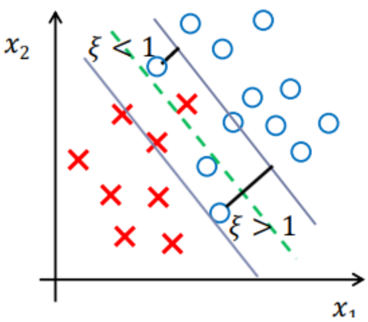
Subject to the constraints:

$$y^{(n)}(w^T x^{(n)} + w_0) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0 \quad n = 1, \cdot, N$$

## Error measure

- Margin violation amount $\xi_n$ ($\xi_n \geq 0$):
  - $y^{(n)}\left(\boldsymbol{w}^T \boldsymbol{x}^{(n)} + w_0\right) \geq 1 - \xi_n$
- Total violation: $\sum_{n=1}^{N} \xi_n$




$\xi_n$: **slack** variables

$0 < \xi_n < 1$: if $\boldsymbol{x}^{(n)}$is correctly classified but inside margin

$\xi_n > 1$: if $\boldsymbol{x}^{(n)}$ is misclassifed

دو دسته قید داریم یکی مانند قبل و دیگری روی اسلک ها .

لاگرانژ روشی برای یافتن ماکزیمم و مینیمم های محلی یک تابع تحت قیود تساوی است، یعنی حداکثر $f(x,y)$ تحت قید $g(x,y) = c$. توابع $f(\cdot)$ و $g(\cdot)$ باید دارای مشتقات جزئی اول پیوسته باشند.

لگرانژیان، نمایش داده شده با $\mathcal{L}(x,y,\lambda)$، با ترکیب تابع هدف $f(x,y)$ و تابع قید $g(x,y)$ به همراه ضریب لاگرانژ $\lambda$ به دست می آید که می تواند مثبت یا منفی باشد. به طور ریاضی، $\mathcal{L}(x,y,\lambda) = f(x,y) + \lambda(g(x,y)-c)$.

اگر $f(x_0,y_0)$ ماکزیمم تابع $f(x,y)$ برای مسئله اصلی تحت قید باشد، سپس وجود دارد $\lambda$ به گونه ای که $(x_0,y_0,\lambda_0)$ یک نقطه ایستایی برای تابع لاگرانژ است (بنابراین $\partial\mathcal{L}/\partial x = \partial\mathcal{L}/\partial y = 0$).

تابع لاگرانژین را تشکیل میدهیم :

$$\mathcal{L}(w,w_0,\xi,\alpha,\beta) = \frac{1}{2}||w||^2 + C\sum_{n=1}^{N}\bar{\xi}_n + \sum_{n=1}^{N}\alpha_n(1-\xi_n - y^{(n)}(w^T x^{(n)} + w_0)) - \sum_{n=1}^{N}\beta_n\xi_n$$

$$\begin{aligned} \max_{\alpha} \quad & \sum_{n=1}^{N}\alpha_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y^{(n)}y^{(m)}(x^{(n)})^T x^{(m)} \\ s.t. \quad & 0 \le \alpha_n \le C, \quad n = 1,\ldots,N \\ & 0 = \sum_{n=1}^{N} y^{(n)}\alpha_n \end{aligned} \tag{۱۲}$$

پس از حل مسئله درجه دوم بالا، $w$ به صورت زیر پیدا می شود:

$$w = \sum_{n=1}^{N}\alpha_n y^{(n)} x^{(n)}$$

**شرایط کاروش-کان-تاکر** $(KKT)$

شرایط $KKT$ مجموعه‌ای از شرایط ریاضی هستند که به ما کمک می‌کنند تا تعیین کنیم آیا یک راه‌حل برای یک مسئله بهینه‌سازی، بهینه است یا خیر. این شرایط ضروری هستند، به این معنا که اگر یک راه‌حل شرایط $KKT$ را برآورده کند، آنگاه می‌تواند یک راه‌حل بهینه باشد، اما ممکن است تنها راه‌حل بهینه نباشد.

شرایط $KKT$ به مسائل بهینه‌سازی که دارای یک تابع هدف و مجموعه‌ای از محدودیت‌ها هستند اعمال می‌شوند. تابع هدف، تابعی است که می‌خواهیم آن را کمینه یا بیشینه کنیم، و محدودیت‌ها شرایطی هستند که راه‌حل ما باید آن‌ها را برآورده کند.

در زمینه ماشین‌های بردار پشتیبان $(SVM)$، شرایط $KKT$ برای یافتن مقادیر بهینه وزن‌ها، بایاس، و ضرایب لاگرانژ که مرز تصمیم‌گیری $SVM$ را تعریف می‌کنند، استفاده می‌شود.

شرایط $KKT$ برای مسئله دوگان $SVM$ شامل موارد زیر است:

- عدم منفی بودن ضرایب لاگرانژ: $\alpha_i \geq 0$ برای همه $i$.
- شرط حاشیه: $y^{(n)}(w^{*T}x^{(n)} + w_0^*) \geq 1 - \xi_n^*$ برای همه $n$. این شرط اطمینان می‌دهد که نقاط داده روی حاشیه دارای حاشیه تابعی حداقل ۱ هستند.
- مکمل اسلک: $\alpha_i^*(1 - y^{(n)}(w^{*T}x^{(n)} + w_0^*) - \xi_n^*) = 0$ برای همه $n$. این شرط اطمینان می‌دهد که یا ضریب لاگرانژ $\alpha_i^*$ صفر است یا محدودیت $y^{(n)}(w^{*T}x^{(n)} + w_0^*) \geq 1 - \xi_n^*$ به صورت مساوی برآورده شده است.
- عدم منفی بودن متغیرهای اسلک: $\xi_n \geq 0$ برای همه $n$. متغیرهای اسلک میزان مجاز نقض محدودیت حاشیه را اندازه‌گیری می‌کنند.
- عدم وجود شکاف دوگان: $\xi_n\beta_n = 0$ برای همه $n$. این شرط اطمینان می‌دهد که شکاف دوگان بین توابع هدف اولیه و دوگان مسئله بهینه‌سازی $SVM$ وجود ندارد.

**$SVM$. با حاشیه نرم: بردارهای پشتیبان**

- بردارهای پشتیبان نقاط داده آموزشی هستند که مرز تصمیم‌گیری (فراصفحه) $SVM$ را تعریف می‌کنند.
- دو نوع بردار پشتیبان وجود دارد که با مقدار ضریب لاگرانژ $\alpha_n$ مشخص می‌شوند:

۱ **بردارهای پشتیبان حاشیه:** این‌ها نقاط داده‌ای هستند که دقیقا روی حاشیه در اطراف مرز تصمیم‌گیری قرار دارند. این‌ها دارای $0 < \alpha_n < C$ هستند.

- ❋ معادله $y^{(n)}(w^Tx^{(n)} + w_0) = 1$ برای این بردارهای پشتیبان برقرار است، که به این معناست که این‌ها به درستی طبقه‌بندی شده‌اند و روی حاشیه قرار دارند.
- ❋ آن‌ها با $(\xi_n = 0)$ نمایش داده می‌شوند زیرا متغیر اسلک $\xi_n$ برای این نقاط صفر است.

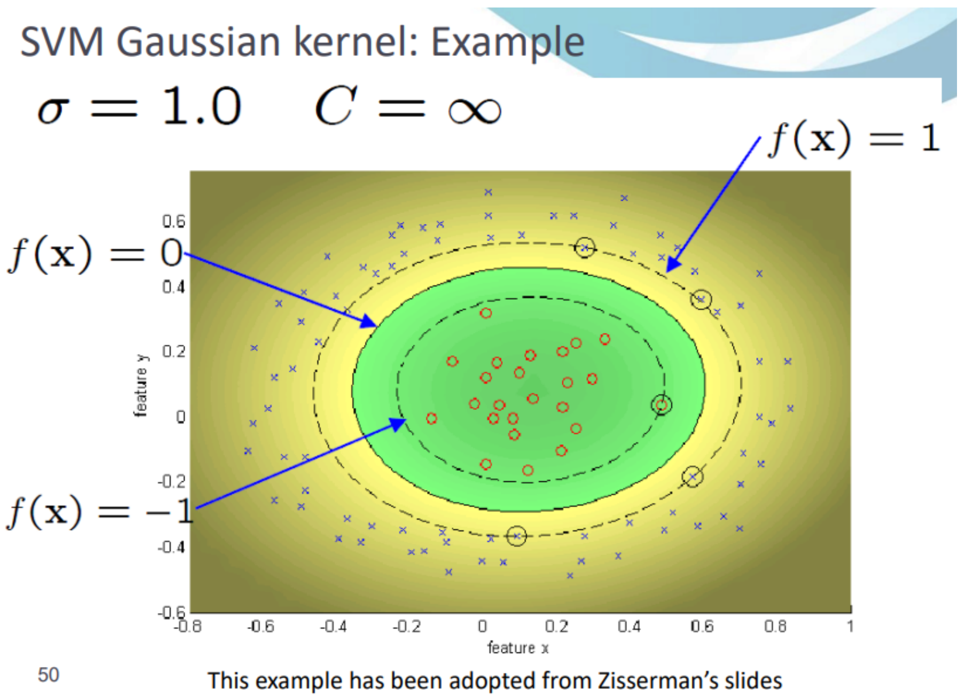
۲ **بردارهای پشتیبان غیر حاشیه** $non - marginsupportvector$: این‌ها نقاط داده‌ای هستند که در طرف اشتباه حاشیه یا داخل حاشیه قرار دارند. این‌ها دارای $\alpha_n = C$ هستند.

- ❋ معادله $y^{(n)}(w^Tx^{(n)} + w_0) < 1$ برای این بردارهای پشتیبان برقرار است، که به این معناست که این‌ها یا به اشتباه طبقه‌بندی شده‌اند یا به درستی طبقه‌بندی شده‌اند اما در داخل حاشیه قرار دارند.

* آن‌ها با $(\xi_n > 0)$ نمایش داده می‌شوند زیرا متغیر اسلک $\xi_n$ برای این نقاط بزرگ‌تر از صفر است.
* معادله $C - \alpha_n - \beta_n = 0$ نیز برای این بردارهای پشتیبان صدق می‌کند.

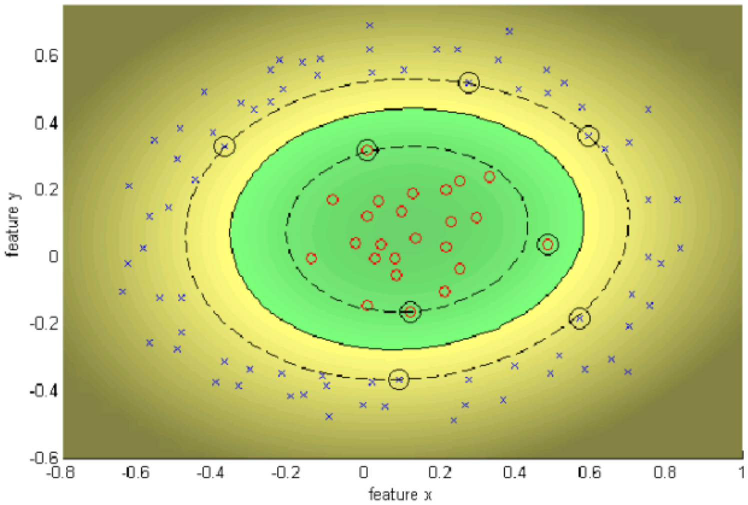
به طور کلی، بردارهای پشتیبان حاشیه عرض حاشیه را تعریف می‌کنند، در حالی که بردارهای پشتیبان غیر حاشیه اجازه می‌دهند که برخی از طبقه‌بندی‌ها اشتباه باشد اما بر اساس پارامتر هزینه $C$ جریمه می‌شوند. پارامتر هزینه $C$ تعادلی بین حداکثرسازی حاشیه و اجازه دادن به برخی اشتباهات طبقه‌بندی را کنترل می‌کند.

تصاویر نمونه‌ای از یک ماشین بردار پشتیبان $(SVM)$ را با استفاده از هسته گاوسی $(RBF)$ برای طبقه‌بندی نشان می‌دهد که مرز تصمیم و نواحی حاشیه را نشان می‌دهد.



## SVM Gaussian kernel: Example

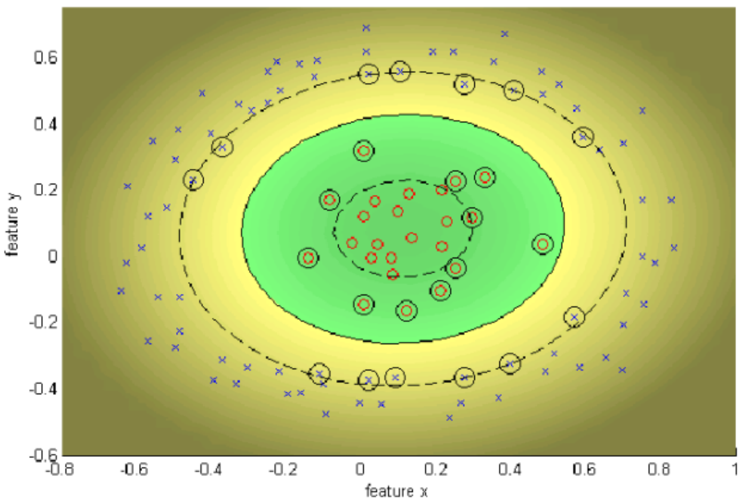
$$\sigma = 1.0 \quad C = 100$$




This example has been adopted from Zisserman's slides

## SVM Gaussian kernel: Example

$$\sigma = 1.0 \quad C = 10$$

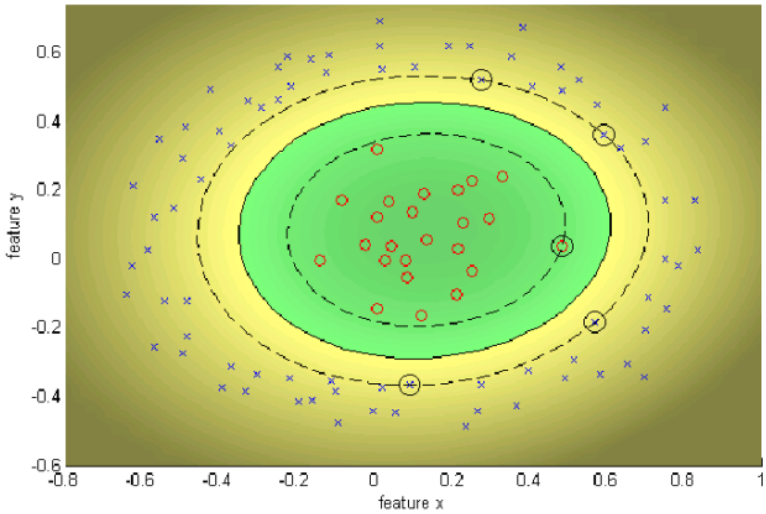



52

This example has been adopted from Zisserman's slides

## SVM Gaussian kernel: Example

$$\sigma = 1.0 \quad C = \infty$$

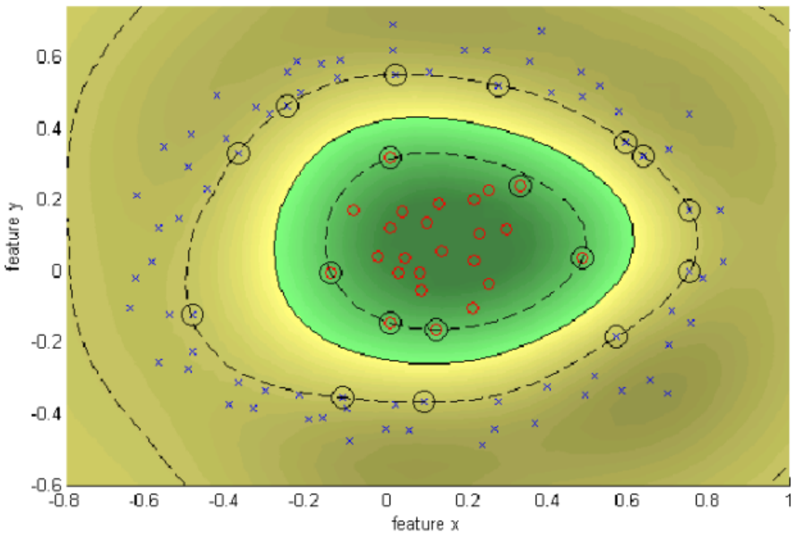



53

This example has been adopted from Zisserman's slides

## SVM Gaussian kernel: Example

$$\sigma = 0.25 \quad C = \infty$$

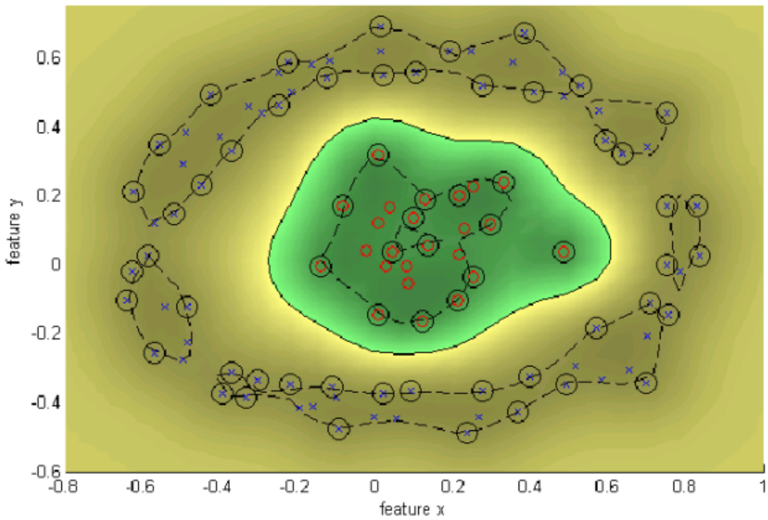



54

This example has been adopted from Zisserman's slides

## SVM Gaussian kernel: Example

$$\sigma = 0.1 \quad C = \infty$$




55

This example has been adopted from Zisserman's slides