بسم الله الرحمن الرحيم

دانشگاه حکیم سبزواری

دانشکده ریاضی وعلوم کامپیوتر

استاد راهنما :
دکتر مهدی زعفرانیه

پژوهشگر:
سیّده افسانه صالحی ساداتی

تابستان 1399

ماشین‌های بردار پشتیبان

Support Vector Machines
SVM

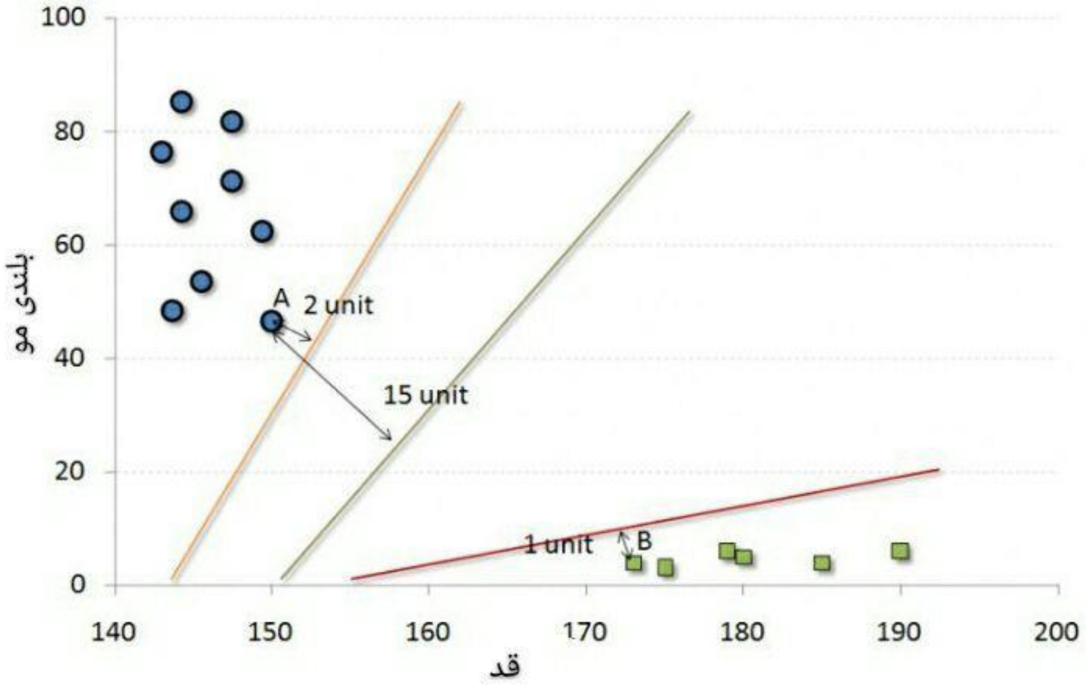
فهرست

- **مقدمه**
- **تاریخچه**
- **انواع ماشین بردار پشتیبان**
- **ماشین بردار پشتیبان خطی**
  - محاسبه حاشیه
  - عدم وجود نمونه‌ها در حاشیه
- **ماشین بردار پشتیبان خطی (داده‌های جداپذیر)**
  - مسئله بهینه سازی ماشین بردار پشتیبان خطی (داده‌های جداپذیر)
  - حل مسئله بهینه سازی ماشین بردار پشتیبان خطی(داده‌های جداپذیر)
  - بررسی شرایط KKT
  - طرح مسئله دوگان برای حل مسئله بهینه سازی ماشین بردار پشتیبان خطی (داده‌های جداپذیر)
  - مثال ماشین بردار پشتیبان خطی (داده‌های جداپذیر)
- **ماشین بردار پشتیبان خطی (داده‌های جداناپذیر)**
- **ماشین بردار پشتیبان غیرخطی**
  - انواع کرنل (هسته)
  - طرح مسئله دوگان برای حل مسئله بهینه سازی ماشین بردار پشتیبان غیرخطی
  - مثال ماشین بردار پشتیبان غیرخطی
- **ابزارهای پیاده سازی ماشین بردار پشتیبان**

مقدمه

انواع یادگیری ماشین :

- **یادگیری بانظارت :** آموزش مدل با مجموعه داده برچسب‌گذاری شده
- **یادگیری بدون نظارت :** آموزش مدل با مجموعه داده بدون برچسب
- **یادگیری تقویتی :** کسب بهترین نتیجه در تعامل با محیط پویا



ماشین بردار پشتیبان (SVM) :

- یکی از روش‌های یادگیری بانظارت است که برای طبقه‌بندی و رگرسیون مورد استفاده قرار می‌گیرد.

- مبنای کار ماشین بردار پشتیبان، دسته‌بندی داده‌هاست. سعی بر این است که خطی پیدا شود که بیشترین فاصله را از تمام دسته‌های داده‌ها داشته باشد.

تاریخچه [1,2]

ارائه دهنده	اقدامات انجام شده	سال
Veladimir Vapnik	ماشین بردار پشتیبان خطی	1963
Veladimir Vapnik , Corina Cortes	ماشین بردار پشتیبان غیر خطی	1995
Veladimir Vapnik	ماشین بردار پشتیبان برای عمل رگرسیون	1996
Giorgio Valentini	ماشین بردار پشتیبان غیر خطی با هسته(کرنل) چندجمله‌ای و گوسی	2002
- Yiqiang Zhan - Bo-Suk Yang - Shu-Xin Du	- روش آموزش برای افزایش بهره‌وری SVM برای طبقه‌بندی سریع (نتایج تجربی در تصاویر سونوگرافی، عملکرد خوبی نشان داد.) - استفاده از SVM برای تشخیص محصولات معیوب در یک خط تولید انبوه یخچال‌های کوچک - ماشین بردار پشتیبان وزن برای طبقه بندی	2005
- Rung-Ching Chen - Chin-Teng Lin	- روش طبقه‌بندی صفحات وب برای استخراج بردارهای ویژگی از دو روش LSA و WPFS با استفاده از SVM - شبکه عصبی فازی مبتنی بر بردار پشتیبانی SVFNN برای به حداقل رساندن خطای آموزش	2006
Kemal Polat	سیستم تصمیم گیری پزشکی مبتنی بر ماشین بردار پشتیبان LSSVM (تشخیص سرطان سینه)	2007
- Jin-Hyuk Hong - Fabien Lauer	- روش طبقه بندي اثر انگشت نوين با تركيب NBها و OVASVM - ارائه فرمول‌های مختلفی از مسئله بهینه‌سازی به همراه SVMs برای کار طبقه‌بندی	2008

تاریخچه [1,2]

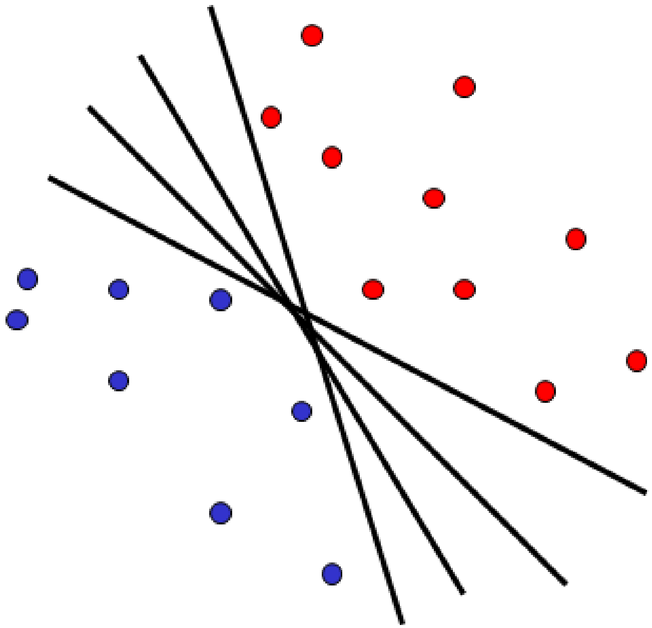
سال	اقدامات انجام شده	ارائه دهنده
2009	- ارائه TSVM و توسعه آن با حداقل مربعات به LSTSVM برای طبقه بندی SVM خطی و غیر خطی	M. Arun Kumar
2010	- ارائه الگوریتم پیشنهادی سیستم طبقه‌بندی ثبت اختراع HGASVM برای اطمینان از استفاده مستمر و منظم اطلاعات ثبت اختراع در فرآیندهای تصمیم گیری یک شرکت - الگوریتم تشخیص ضربان قلب قوی با LS-SVM(انجام خودکار طبقه‌بندی ضربان قلب طبیعی) - MSVM برای طبقه‌بندی الیاف خارجی موجود در پارچه نخی و ارئه سه نوع OAA-DTB MSVM، OAO-VB MSVMو OAO-DAG MSVM	- Chih-Hung Wu - Saibal Dutta - Daoliang Li
2011	- ارائه FSVM برای بررسی مسئله پیش بینی ورشکستگی در سازمان های شرکت ها	Arindam Chaudhuri
2012	- ارائه CDMTSVM (الگوریتم از نوع ماشین بردار پشتیبان دوقلو مبتنی بر حاشیه نزولی TWSVM برای طبقه بندی باینری) و ارائه روشی برای حل مسئله دوگان که به حافظه کمتری نیاز دارد. - ارائه SVM-NN (ادغام روش طبقه بندي KNN با الگوریتم SVM برای تعیین مقدار مناسب پارامتر K)	- Yuan-Hai Shao - Chin Heng Wan
2013	- ارائه مدل ترکیبی جدیدی مبتنی بر الگوریتم کرم شب‌تاب برای پیش‌بینی قیمت سهام بورس - ارائه الگوریتم DE-SVC برای تشخیص الگو - ارائه ماشین بردار پشتیبانی دوقلو با ساختاری جدید S-TWSVM - ارائه DE-SVC برای تشخیص الگوی	- Ahmad Kazem - Youngdae Kim - Zhiquan Qi - Zhen Yang
2014	- الگوریتم FSVM مبتنی بر خوشه‌بندی PIM (PIM-FSVM)برای مسائل طبقه‌بندی - eABC-LSSVM برای پیش بینی قیمت های سری زمانی - توسعه LSPTSVM به NLSPTSVM برای حل مشکلات طبقه بندي غیرخطي	- Zhenning Wu - Zuriani Mustaffa - Shifei Ding

انواع ماشین بردار پشتیبان

- ماشین بردار پشتیبان خطی :
  - داده‌ها جداپذیر باشند؛ دقیقا در دو دسته جدا از هم قرار داشته باشند.
  - داده‌ها جداپذیر نباشند؛ قابل تفکیک به دو دسته جدا از هم نباشند.
- ماشین بردار پشتیبان غیر خطی
- ماشین بردار پشتیبان برای تفکیک کننده‌های چند کلاسه

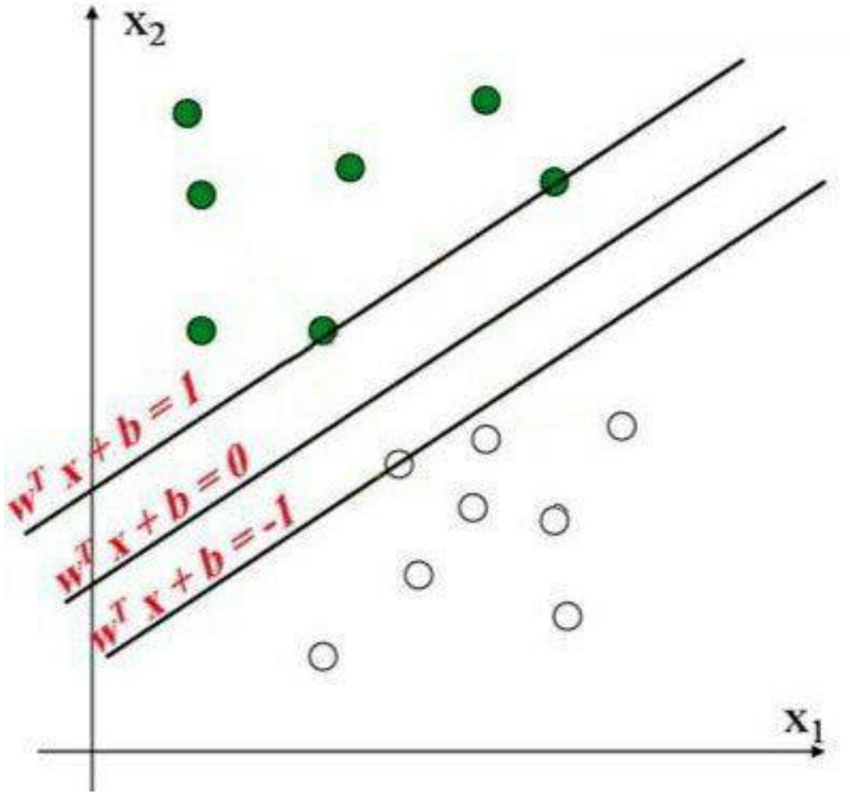
ماشین بردار پشتیبان خطی

- هدف SVM دسته بندی خطی داده هاست و خطی (ابرصفحه‌ای) انتخاب می شود که بیشترین فاصله از تمام دسته ها را داشته باشد(حاشیه اطمینان بیشتری داشته باشد).



- معادلات حاصل از ماشین بردار پشتیبان خطی قابل تعمیم به حالت جداناپذیر و غیرخطی می‌باشد، لذا این حالت اساس تعریف سایر حالات است.

ماشین بردار پشتیبان خطی



- انتخاب مجموعه داده برچسب‌گذاری شده
- انتخاب دو بردار پشتیبان(بردار مرزی) که داده‌ها را تفکیک می‌کند و هیچ داده‌ای بین آن‌ها نیست.
- بیشینه سازی فاصله بین این دو بردار پشتیبان (حاشیه)
- ابرصفحه وسط حاشیه، بردار جداکننده بهینه خواهد بود.

ماشین بردار پشتیبان خطی

- مجموعه داده‌ی آموزشی

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^p,\ y_i \in \{-1,1\}\}_{i=1}^{n}$$

- لم: معادله خط جداکننده در فضای دو بعدی

$$\mathrm{w}_1\mathrm{x}_1 + \mathrm{w}_2\mathrm{x}_2 + \mathrm{b} = 0$$

- لم: معادله ابرصفحه جداکننده در فضا با ابعاد بالا

$$\sum_i w_i x_i + b = 0$$

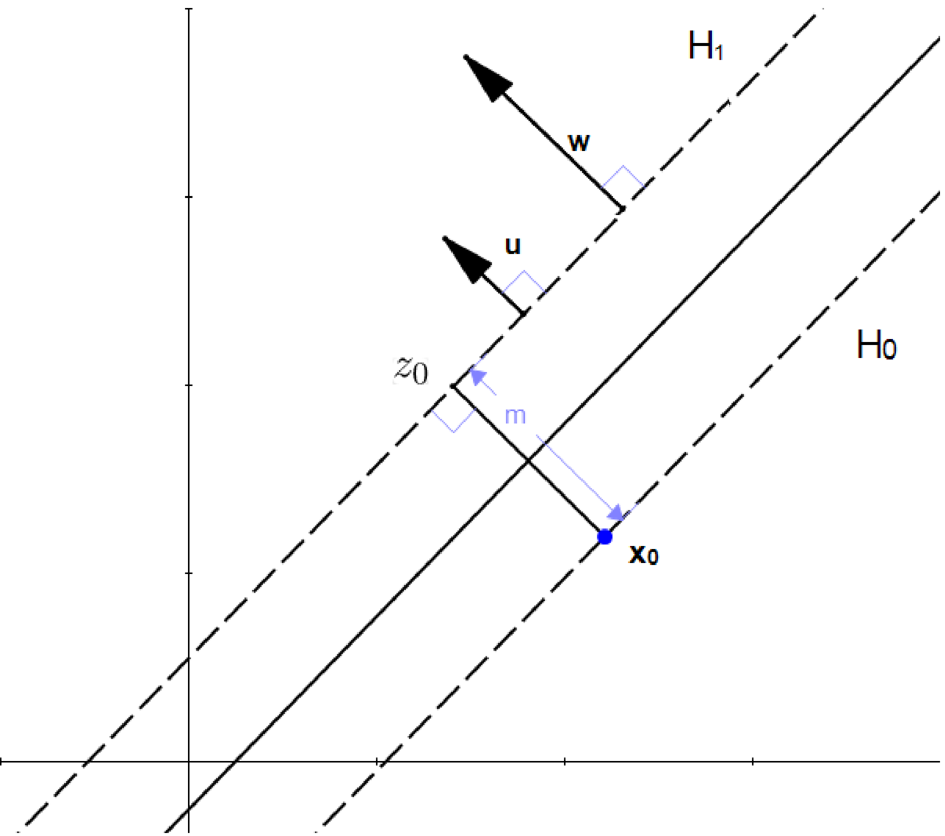
$$\mathrm{u} = \vec{\mathrm{w}}.\vec{\mathrm{x}} + \mathrm{b} \quad \Rightarrow \quad \vec{\mathrm{w}}.\vec{\mathrm{x}} + \mathrm{b} = 0$$

- انتخاب دو بردار پشتیبان (بردارهای مرزی)

$$\begin{cases} \vec{\mathrm{w}}.\vec{\mathrm{x}} + \mathrm{b} = +1 \\ \vec{\mathrm{w}}.\vec{\mathrm{x}} + \mathrm{b} = -1 \end{cases}$$

حاشیه

- انتخاب نقطه x_0 روی یک بردار پشتیبان و عمود بر بردار پشتیبان دیگر
- یافتن بردار هم اندازه با m که بر H_1 عمود است :



➡ $u = \frac{w}{\|w\|}$

➡ $\|u\| = 1$

➡ $k = mu$

➡ $\|k\| = m$

➡ $k = mu = m\frac{w}{\|w\|}$

حاشیه

- به دست آوردن نقطه z_0:

$$z_0 = x_0 + k$$

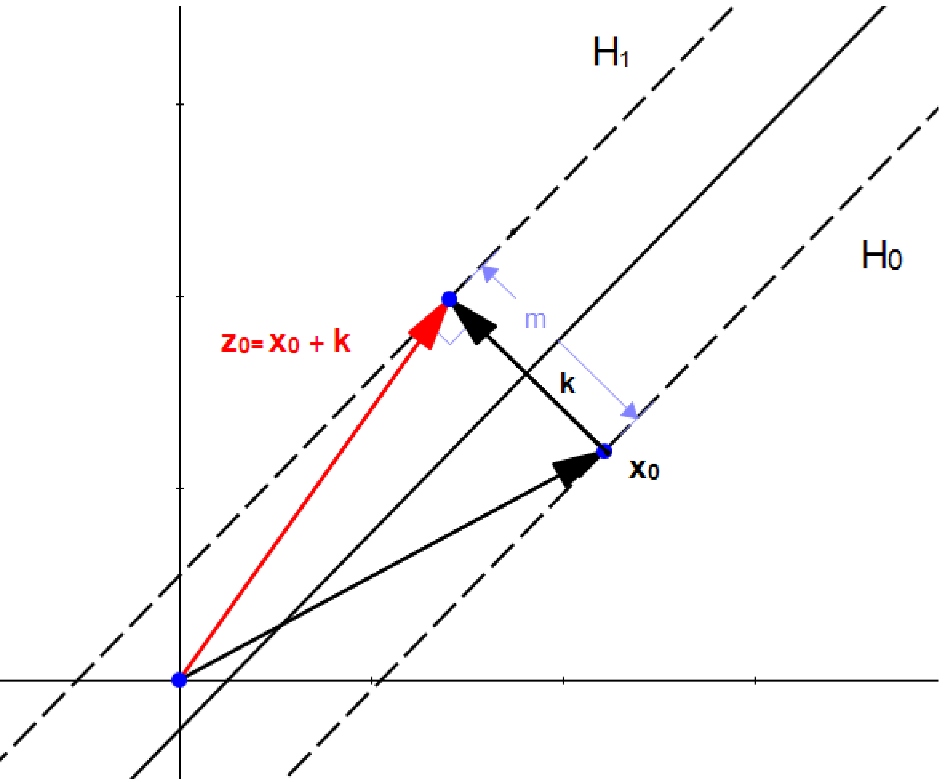
- نقطه z_0 روی بردار پشتیبان H_1 قرار دارد :

$$\left.\begin{aligned} w.z_0 + b = 1 \\ z_0 = x_0 + k \end{aligned}\right\} \Rightarrow w.(x_0 + k) + b = 1$$

$$\Rightarrow w.(x_0 + m\frac{w}{\|w\|}) + b = 1$$

$$\Rightarrow w.x_0 + m\frac{w.w}{\|w\|} + b = 1$$

$$\Rightarrow w.x_0 + m\frac{\|w\|^2}{\|w\|} + b = 1$$



حاشیه

$$\Rightarrow \quad w.x_0 + m\|w\| + b = 1$$

$$\Rightarrow \quad w.x_0 + b = 1 - m\|w\|$$

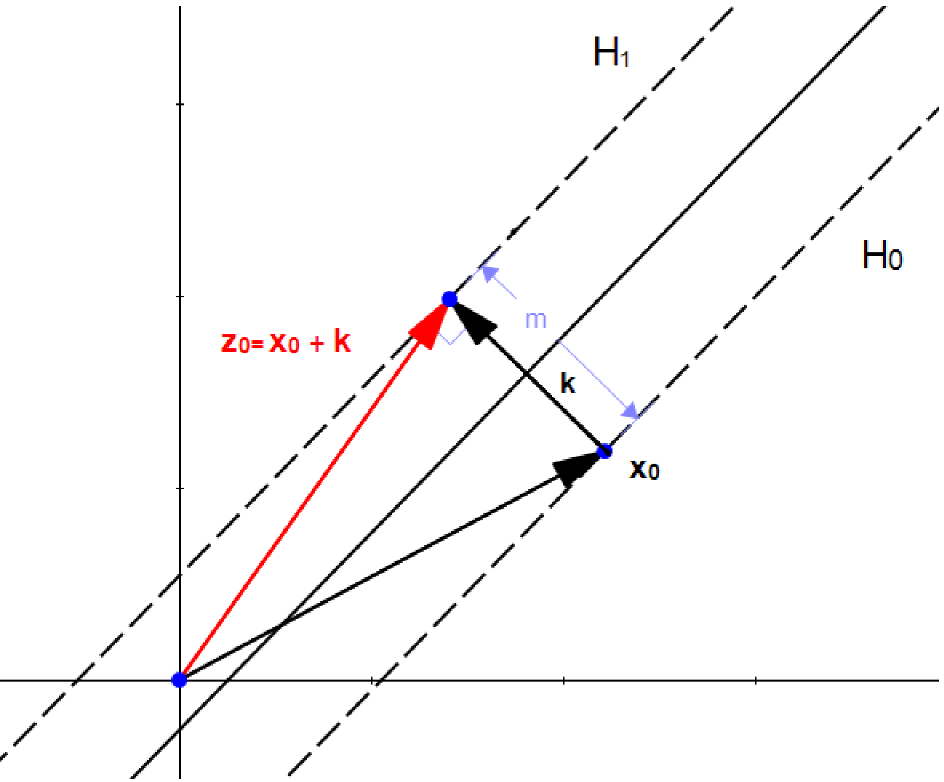
- نقطه **x_0** روی H_0 قرار دارد :

$$w.x_0 + b = -1$$

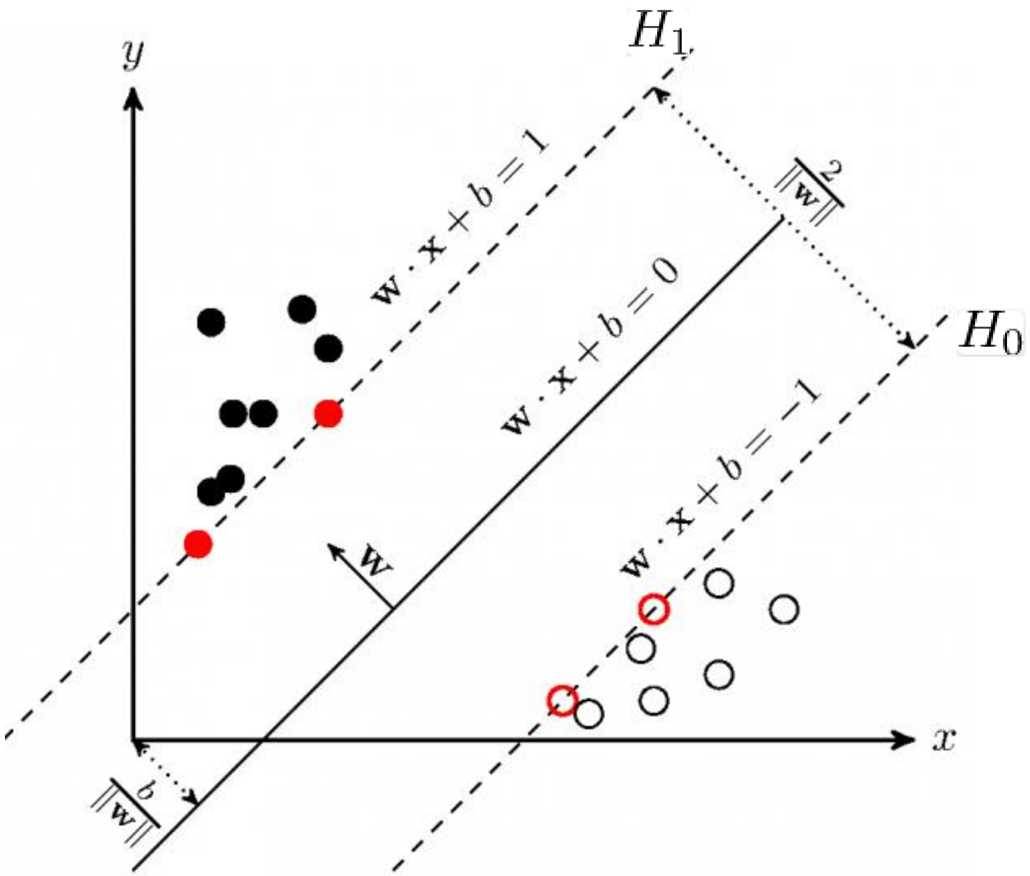
$$\Rightarrow \quad -1 = 1 - m\|w\|$$

$$\Rightarrow \quad m\|w\| = 2$$

$$\Rightarrow \quad m = \frac{2}{\|w\|}$$



عدم وجود نمونه‌ها در حاشیه



- فضا به دو دسته نمونه با ویژگی زیر تقسیم شده:

$$\begin{cases} \vec{w}.\vec{x}+b \geq +1 & \text{for } y_i = 1 \\ \vec{w}.\vec{x}+b \leq -1 & \text{for } y_i = -1 \end{cases}$$

- ترکیب این دو رابطه:

$$y_i(\vec{w}.\vec{x}+b) - 1 \geq 0 \qquad \forall i$$

SVM خطی (داده‌ها جداپذیر)

- هدف SVM پیدا کردن بردار جداکننده بهینه است.
- در فضایی که نمونه‌های آموزشی جداپذیر خطی هستند، بردار جداکننده بهینه بیشترین فاصله را با نمونه‌های هر دسته دارد.
- افزایش فاصله بردار جداکننده بهینه با نمونه‌ها برابر با افزایش حاشیه بین دو بردار پشتیبان مرزی است، به گونه‌ای که از ورود نقاط به حاشیه جلوگیری شود.
- افزایش حاشیه با کمینه کردن $\|w\|$ در رابطه $\frac{2}{\|w\|}$ ایجاد می‌شود.

مسئله بهینه سازی

- مسئله بهینه سازی مقید :

1 $$\min \quad \frac{1}{2}\|w\|^2$$
$$\text{s.t.} \quad y_i(\vec{w}.\vec{x}+b)-1 \geq 0 \qquad \forall i$$

- روش‌های حل مسئله درجه دوم:
  - استفاده از جدول سیمپلکس
  - روش KKT
  - روش‌های تقریب خطی و جداسازی
  - جهت گرادیان (Gradient direction)
  - سریع‌ترین شیب کاهنده (Steepest descent)
  - گرادیان مزدوج (Conjugate gradient)
  - نیوتون (Newton) و شبه نیوتون (Quas-Newton)
  - لاگرانژ (Lagrange)

- تبدیل مسئله بهینه سازی مقید به مسئله بهینه سازی نامقید با روش لاگرانژ :

2 $$\max_{\alpha_i} \ \min_{w,b} \ L_p = \frac{1}{2}\|w\|^2 - \sum_i \alpha_i(y_i(\vec{w}.\vec{x}+b)-1) \quad , \quad \alpha_i \geq 0$$

حل مسئله بهینه سازی

x^* بردار کمینه سراسری تابع خواهد بود اگر یکی از موارد زیر برقرار باشد :

x^* کوچکترین کمینه در بین تمام کمینه‌های محلی باشد.

قضیه: فرض کنید $f: R^n \rightarrow R$ یک تابع محدب باشد و x^* یک کمینه محلی باشد، آنگاه x^* یک کمینه سراسری خواهدبود.

تابع محدب: تابع $f: R^n \rightarrow R$ محدب است اگر $\forall x_1, x_2 \in R^n$ وجود داشته باشد که:

$$f[\lambda x_1 + (1-\lambda)x_2] \leq \lambda f(x_1) + (1-\lambda)f(x_2) \qquad , \lambda \in (0,1)$$

- قضیه: نقطه بحرانی x جواب کمینه محلی تابع $f: R^n \rightarrow R$ خواهد بود اگر:

$\nabla f(x) = 0$

ماتریس $\nabla^2 f(x)$ معین مثبت باشد.

$$\text{Hessian} = \nabla^2 f(x) = \begin{pmatrix} \frac{\delta^2 f}{\delta x_1^2} & \cdots & \frac{\delta^2 f}{\delta x_1 \delta x_n} \\ \vdots & \ddots & \vdots \\ \frac{\delta^2 f}{\delta x_n \delta x_1} & \cdots & \frac{\delta^2 f}{\delta x_n^2} \end{pmatrix}$$

حل مسئله بهینه سازی

- ماتریس معین مثبت :

$$z^T((\nabla^2 f(x))z > 0, \forall z \in R^n \quad , \quad z \neq 0 \Longleftarrow x^T A x > 0 \quad \forall\ x \in R^n \quad , \quad x \neq 0$$

دترمینان زیرماتریس‌های اصلی آن مثبت باشد.

در متلب با chol (A)

زیرماتریس اصلی(پیشرو):

فرض کنید A یک ماتریس n ✖ n باشد، ماتریس حاصل از حذف n-k سطر و ستون آخر را زیرماتریس اصلی مرتبه k گویند.

$$\Rightarrow A = \begin{bmatrix} a_{۱۱} & a_{۱۲} & a_{۱۳} \\ a_{۲۱} & a_{۲۲} & a_{۲۳} \\ a_{۳۱} & a_{۳۲} & a_{۳۳} \end{bmatrix}$$

$$\Rightarrow A_{۱} = [a_{۱۱}] \ , \ A_{۲} = \begin{bmatrix} a_{۱۱} & a_{۱۲} \\ a_{۲۱} & a_{۲۲} \end{bmatrix} \ , \ A_{۳} = \begin{bmatrix} a_{۱۱} & a_{۱۲} & a_{۱۳} \\ a_{۲۱} & a_{۲۲} & a_{۲۳} \\ a_{۳۱} & a_{۳۲} & a_{۳۳} \end{bmatrix}$$

- مسئله‌ی به دست آمده برای SVM یک مسئله‌ی محدب است.

شرایط KKT :

- شرایط KKT برای مسئله محدب، شرط لازم و کافی برای داشتن پاسخ بهینه برای مسائل مقید می‌باشد.

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0; \quad i \in \{1, \dots, m\} \\ & h_j(x) = 0; \quad j \in \{1, \dots, \ell\} \end{aligned}$$

1. $\nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{j=1}^{\ell} \lambda_j \nabla h_j(x^*) = \mathbf{0}$
2. $\mu_i \ge 0, \text{ for } i = 1, \dots, m$
3. $g_i(x^*) \le 0, \text{ for } i = 1, \dots, m$
4. $h_j(x^*) = 0, \text{ for } j = 1, \dots, \ell$
5. $\mu_i g_i(x^*) = 0, \text{ for } \ i = 1, \dots, m.$

حل مسئله بهینه سازی SVM

$$\min \quad \frac{1}{2}\|w\|^2$$
$$\text{s.t.} \quad y_i(\vec{w}.\vec{x}+b)-1 \geq 0 \quad \forall i$$

$$\Rightarrow \quad \max_{\alpha_i} \min_{w,b} \quad LP = \frac{1}{2}W^TW - \sum_i \alpha_i[Y_i(W^TX_i+b)-1]$$
$$\alpha_i \geq 0$$

- شرایط KKT :

1. $$\frac{\partial LP}{\partial W} = 0 \rightarrow W - \sum_i \alpha_i Y_i X_i = 0 \rightarrow \boxed{W = \sum_i \alpha_i Y_i X_i}$$
2. $$\frac{\partial LP}{\partial b} = 0 \rightarrow \sum_i \alpha_i Y_i = 0$$
3. $$y_i(\vec{w}.\vec{x}+b)-1 \geq 0 \ , \ i = 1, \dots, l$$
4. $$\alpha_i[Y_i(W^TX_i+b)-1]=0 \quad \forall i \quad , \quad \alpha_i \geq 0 \quad \forall i$$

حل مسئله بهینه سازی SVM

- طبق شرط KKT :

$$\alpha_i[Y_i(W^TX_i + b) - 1]=0$$

$$IF\ \alpha_i > 0 \Rightarrow Y_i(W^TX_i + b) - 1 = 0$$
$$Y_i(W^TX_i + b) = 1$$

- ضرب دو طرف در Y_i:

$$Y_i^2(W^TX_i + b) = Y_i \Rightarrow W^TX_i + b = Y_i \Rightarrow \boxed{b = Y_i - W^TX_i}$$

- مقدار b برحسب w و مقدار w برحسب α_i به دست می‌آیند، از این رو باید مقدار بهینه برای α_i محاسبه شود.
- با به دست آمدن w و b معادله‌ی ابرصفحه جداکننده به دست می‌آید. $\vec{w}.\vec{x} + b = 0$

مسئله دوگان

- جایگذاری w در تابع هدف :

$$LD = \frac{1}{2}W^TW \quad - \sum_i \alpha_i[Y_i(W^TX_i + b) - 1]$$

$$= \frac{1}{2}\left(\sum_j \alpha_j Y_j X_j\right)^T \left(\sum_i \alpha_i Y_i X_i\right) - \sum_i \alpha_i \left[Y_i\left(\left(\sum_j \alpha_j Y_j X_j\right)^T X_i + b\right) - 1\right]$$

- معادله‌ی دوگان برای به دست آوردن α_i بهینه به صورت بیشینه سازی رابطه زیر است :

$$\max \; LD = -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j \underbrace{Y_iY_jX_i{}^TX_j}_{Q} + \sum \alpha_i$$

$$\text{s.t.} \quad \sum_i \alpha_i Y_i = 0$$

$$\alpha_i > 0 \quad \forall_i$$

$$\alpha = \begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_n\end{bmatrix}$$

$$\Rightarrow$$

$$\min \quad W(\alpha) = -\alpha^T + \frac{1}{2}\alpha^TQ\alpha$$

$$\text{s.t.} \quad \alpha^Ty = 0$$

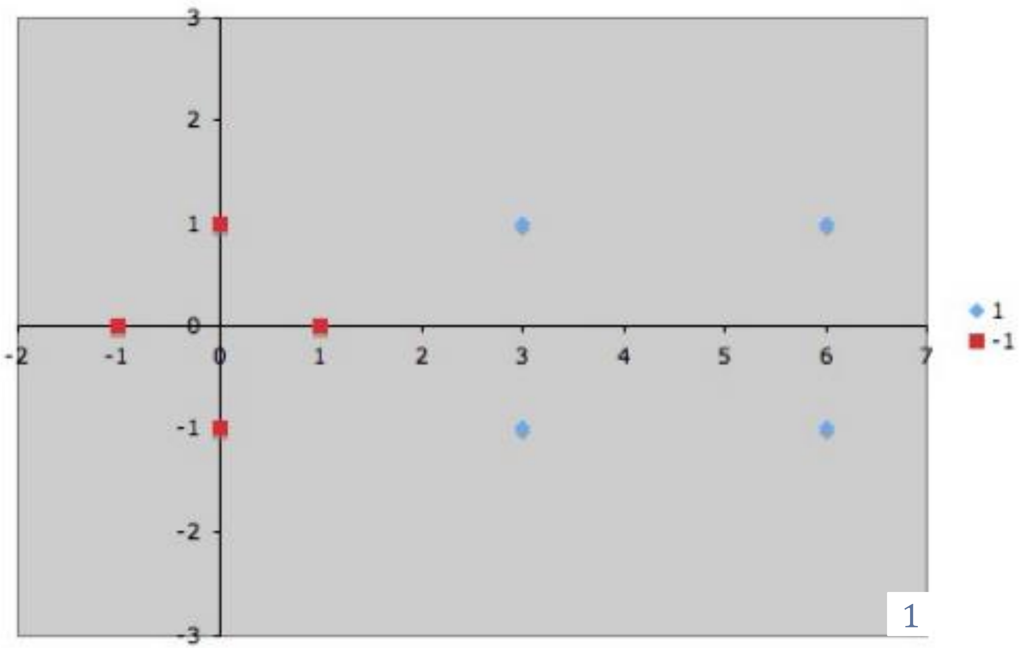
$$0 < \alpha$$

مثال SVM خطی(داده‌ها جداپذیر)

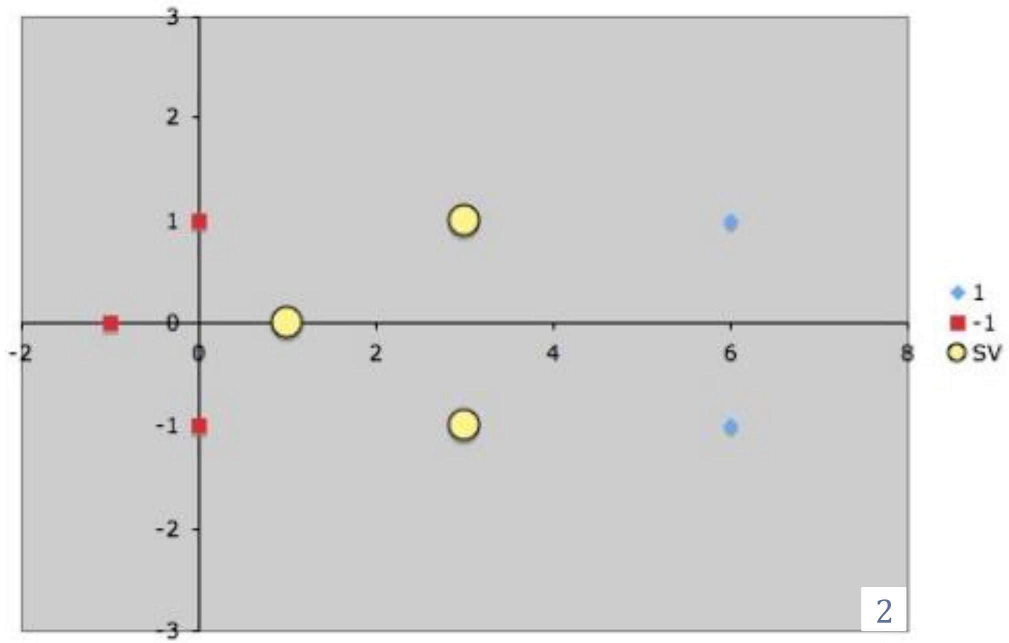
- نقاط با برچسب مثبت: $\left\{\begin{pmatrix}3\\1\end{pmatrix},\begin{pmatrix}3\\-1\end{pmatrix},\begin{pmatrix}6\\1\end{pmatrix},\begin{pmatrix}6\\-1\end{pmatrix}\right\}$
- نقاط با برچسب منفی: $\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix},\begin{pmatrix}0\\-1\end{pmatrix},\begin{pmatrix}-1\\0\end{pmatrix}\right\}$

نقاط بردار پشتیبان: $\left\{X_1=\begin{pmatrix}1\\0\end{pmatrix},X_2=\begin{pmatrix}3\\1\end{pmatrix}\ X_3=\begin{pmatrix}3\\-1\end{pmatrix}\right\}$

برچسب این نقاط: $Y_1=-1\ ,\ Y_2=1\ ,\ Y_3=1$



1



2

مثال SVM خطی(داده‌ها جداپذیر)

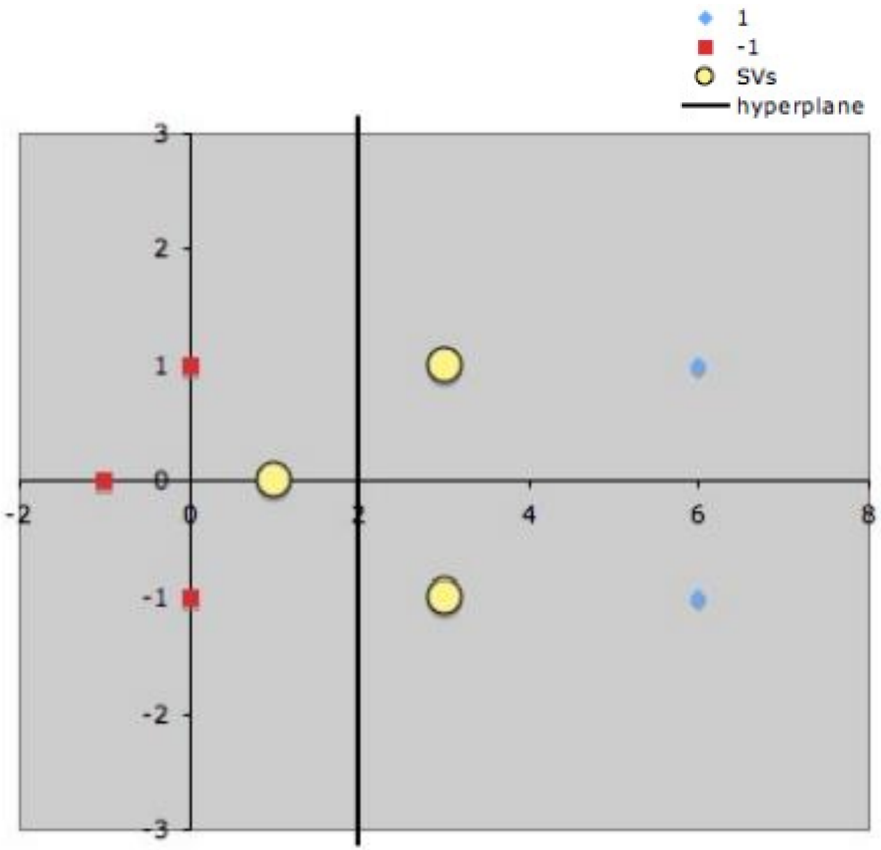
$$W=\sum_i \alpha_i Y_i X_i \quad , \quad \sum_i \alpha_i Y_i = 0$$

$$\begin{cases} WX_1+b=-1 \\ WX_2+b=1 \\ WX_3+b=1 \\ \alpha_1Y_1+\alpha_2Y_2+\alpha_3Y_3=0 \end{cases} \Longrightarrow \begin{cases} (\alpha_1Y_1X_1+\alpha_2Y_2X_2+\alpha_3Y_3X_3)\ X_1+b=-1 \\ (\alpha_1Y_1X_1+\alpha_2Y_2X_2+\alpha_3Y_3X_3)\ X_2+b=1 \\ (\alpha_1Y_1X_1+\alpha_2Y_2X_2+\alpha_3Y_3X_3)\ X_3+b=1 \\ \alpha_1Y_1+\alpha_2Y_2+\alpha_3Y_3=0 \end{cases}$$

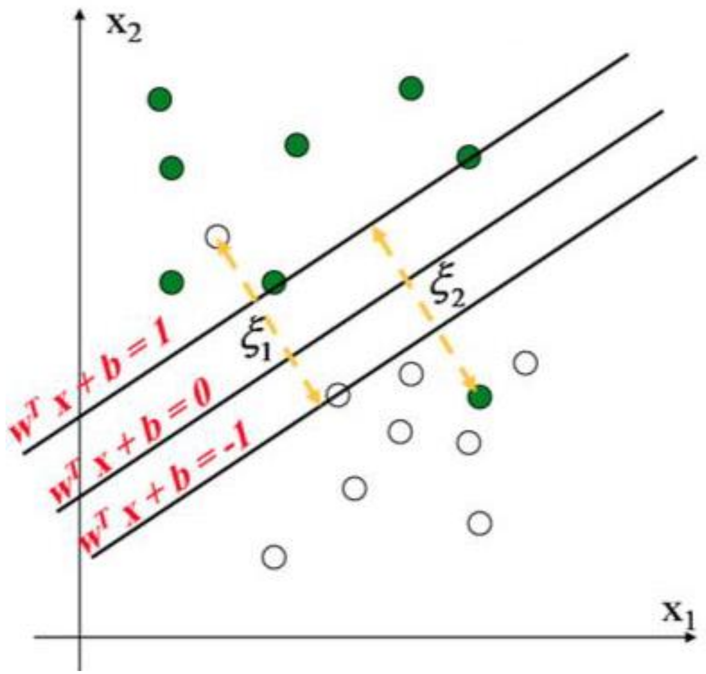
$$\begin{cases} -\alpha_1+3\alpha_2+3\alpha_3+b=-1 \\ -3\alpha_1+10\alpha_2+8\alpha_3+b=1 \\ -3\alpha_1+8\alpha_2+10\alpha_3+b=1 \\ \alpha_1Y_1+\alpha_2Y_2+\alpha_3Y_3=0 \end{cases} \Longrightarrow \begin{cases} \alpha_1=0.5 \\ \alpha_2=0.25 \\ \alpha_3=0.25 \\ b=-2 \end{cases}$$

$$W=\sum_i \alpha_i Y_i X_i = -0.5\begin{pmatrix}1\\0\end{pmatrix}+0.25\begin{pmatrix}3\\1\end{pmatrix}+0.25\begin{pmatrix}3\\-1\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}$$

$$WX+b=0 \Longrightarrow \boxed{\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}X_1\\X_2\end{pmatrix}-2}=0 \Longrightarrow X_1=2$$



SVM خطی (داده‌ها جداناپذیر)



- داده‌های موجود به سادگی قابل تفکیک به دو دسته نیستند و با نویز همراه‌اند.

- گسترش ایده مبحث قبل با تعریف متغیر مثبت کاهشی (کمبود) در قیود مسئله.

$$\begin{cases} W^T X_i + b + \xi_i \geq 1 \\ W^T X_i + b - \xi_i \leq -1 \end{cases}$$

$$\left.\begin{array}{lll} IF\ Y_i = 1 & W^T X_i + b \geq 1 - \xi_i & , \xi_i \geq 0 \\ IF\ Y_i = -1 & W^T X_i + b \leq -1 + \xi_i & , \xi_i \geq 0 \end{array}\right\} \quad Y_i(W^T X_i + b) \geq 1 - \xi_i$$

SVM خطی (داده‌ها جداناپذیر)

- $\sum_i \xi_i$ یک کران بالا برای تعداد خطاهای آموزش می‌باشد.
- تابع هدف به صورت روبرو خواهد بود :
- مقدار C توسط کاربر انتخاب می‌شود.

1 $$\min \frac{1}{2}W^T W + C\sum_i \xi_i$$

$$\text{s.t.} \quad Y_i(W^T X_i + b) \geq 1 - \xi_i \;\; \forall_i$$

$$\xi_i \geq 0$$

- تبدیل مسئله بهینه سازی مقید به مسئله بهینه سازی نامقید با روش لاگرانژ :

2 $$\max_{\alpha_i, \mu_i} \min_{w, b, \xi_i} \quad LP = \frac{1}{2}W^T W + C\sum_i \xi_i - \sum_i \alpha_i\left[Y_i(W^T X_i + b) - 1 + \xi_i\right] - \sum_i \mu_i \xi_i$$

$$\mu_i \geq 0 \quad , \quad \alpha_i \geq 0 \quad , \quad \xi_i \geq 0 \;\; \forall i$$

SVM خطی (داده‌ها جداناپذیر)

- اعمال شرایط KKT :

$$\frac{\partial L_p}{\partial w_v} = w_v - \sum_i \alpha_i y_i x_{iv} = 0 \quad , \quad v = 1, \ldots, d \quad \Rightarrow \quad \boxed{w = \sum_i \alpha_i y_i x_i}$$

$$\frac{\partial L_p}{\partial b} = -\sum_i \alpha_i y_i = 0 \quad , \quad \frac{\partial L_p}{\partial \xi_i} = C - \alpha_i - \mu_i = 0$$

$$\alpha_i \left[y_i (w^T x_i + b) - 1 + \xi_i \right] = 0 \quad \Rightarrow \quad \text{if } \alpha_i > 0 \; ; \quad y_i (w^T x_i + b) - 1 + \xi_i = 0$$

$$\Rightarrow \quad {y_i}^2 (w^T x_i + b) = y_i - \xi_i y_i \quad \Rightarrow \quad \boxed{b = y_i - \xi_i y_i - w^T x_i}$$

- با حل مسئله دوگان روبرو ومحاسبه مقدار بهینه برای α_i، مقادیر بهینه برای w و b به دست می‌آید و با جایگذاری در معادله‌ی بردار جداکننده‌ی زیر، ابرصفحه جداکننده مشخص می‌شود.

$$\vec{w}.\vec{x} + b = 0$$

$$\max \quad LD = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j Y_i Y_j {X_i}^T X_j + \sum \alpha_i$$

$$\text{s.t.} \quad 0 < \alpha_i \le C, \qquad \sum_i \alpha_i y_i = 0$$

SVM خطی (داده‌ها جداناپذیر)

نقاط با برچسب مثبت: $\left\{\begin{pmatrix}0\\-1\end{pmatrix},\begin{pmatrix}3\\1\end{pmatrix}\begin{pmatrix}3\\-1\end{pmatrix}\begin{pmatrix}6\\-1\end{pmatrix}\right\}$

نقاط با برچسب منفی: $\left\{\begin{pmatrix}-1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}6\\1\end{pmatrix}\right\}$

$c=1 \quad \Rightarrow \quad W=\begin{pmatrix}0.18\\-1.27\end{pmatrix} \quad , \quad b=-0.81$

$$\begin{pmatrix}0.18 & -1.27\end{pmatrix}\begin{pmatrix}X_1\\X_2\end{pmatrix}-0.81=0$$

$$WX+b=0 \quad \Rightarrow \quad 0.18X_1-1.27X_2=0.81$$

$c=0.5 \quad \Rightarrow \quad W=\begin{pmatrix}0.148\\-0.999\end{pmatrix} \quad , \quad b=-0.63$

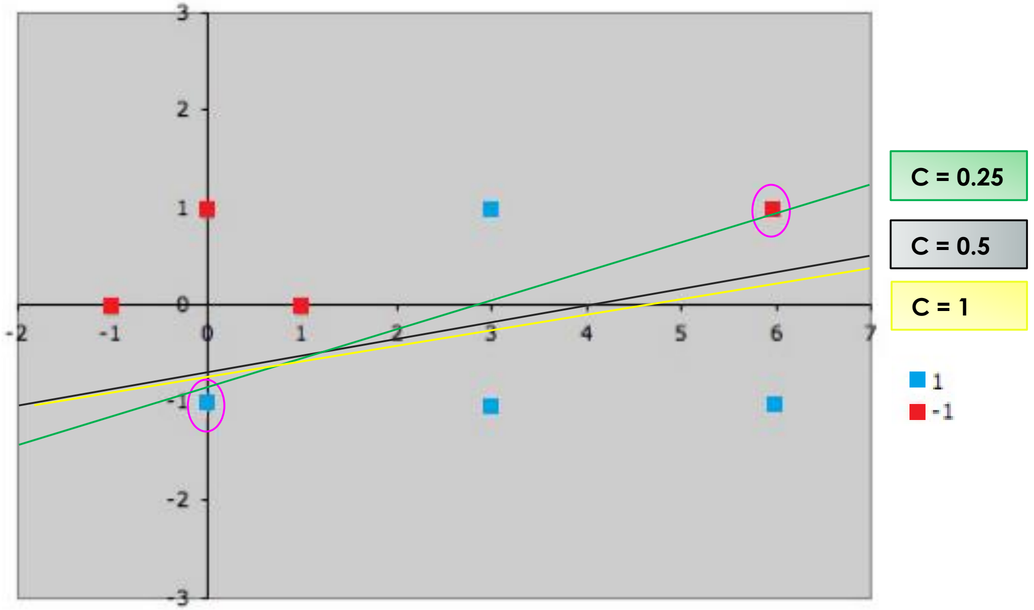
$$\begin{pmatrix}0.148 & -0.999\end{pmatrix}\begin{pmatrix}X_1\\X_2\end{pmatrix}-0.63=0$$

$$WX+b=0 \quad \Rightarrow \quad 0.148X_1-0.999X_2=0.63$$

$c=0.25 \quad \Rightarrow \quad W=\begin{pmatrix}0.15\\-0.55\end{pmatrix} \quad , \quad b=-0.45$

$$\begin{pmatrix}0.15 & -0.55\end{pmatrix}\begin{pmatrix}X_1\\X_2\end{pmatrix}-0.45=0$$

$$WX+b=0 \quad \Rightarrow \quad 0.15X_1-0.55X_2=0.45$$



SVM خطی (داده‌ها جداناپذیر)

افزایش مجموعه داده‌ی آموزشی

$$\begin{cases} c = 1 \\ c = 0.5 \\ c = 0.25 \end{cases} \Rightarrow \quad W = \begin{pmatrix} 0.33 \\ 0 \end{pmatrix} \quad , \quad b = -1$$

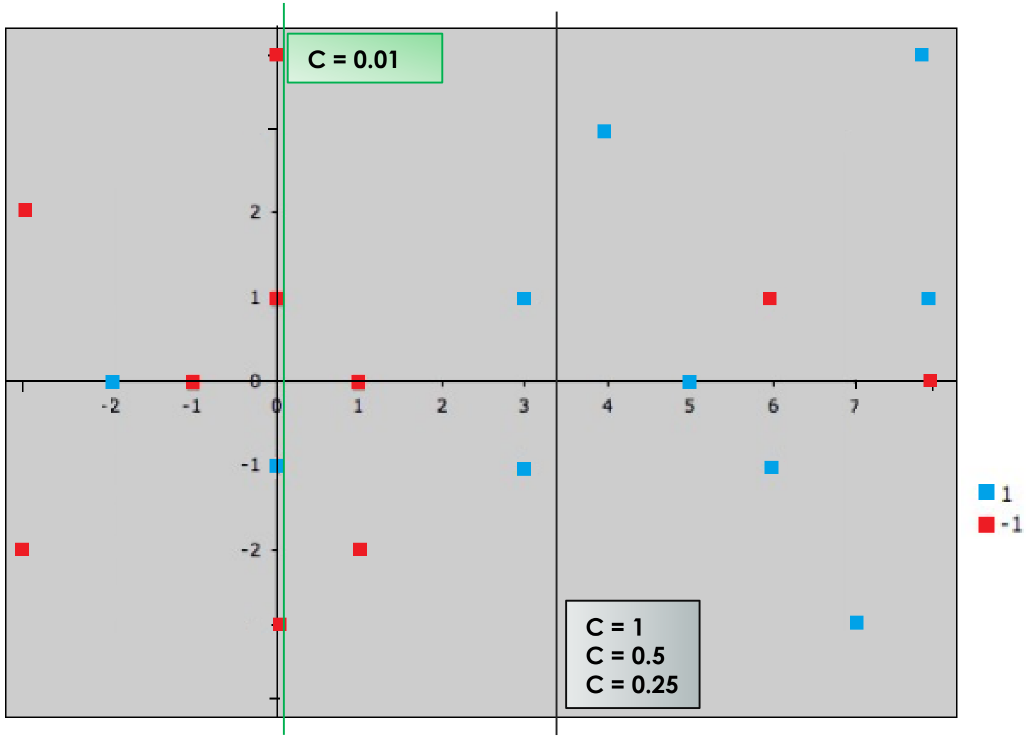
$$\begin{pmatrix} 0.33 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} - 1 = 0$$

$$WX + b = 0 \quad \Rightarrow \quad 0.33X_1 = 1 \quad \Rightarrow \quad X_1 = 3.03$$

$$c = 0.01 \quad \Rightarrow \quad W = \begin{pmatrix} 0.18 \\ 0 \end{pmatrix} \quad , \quad b = -0.45$$

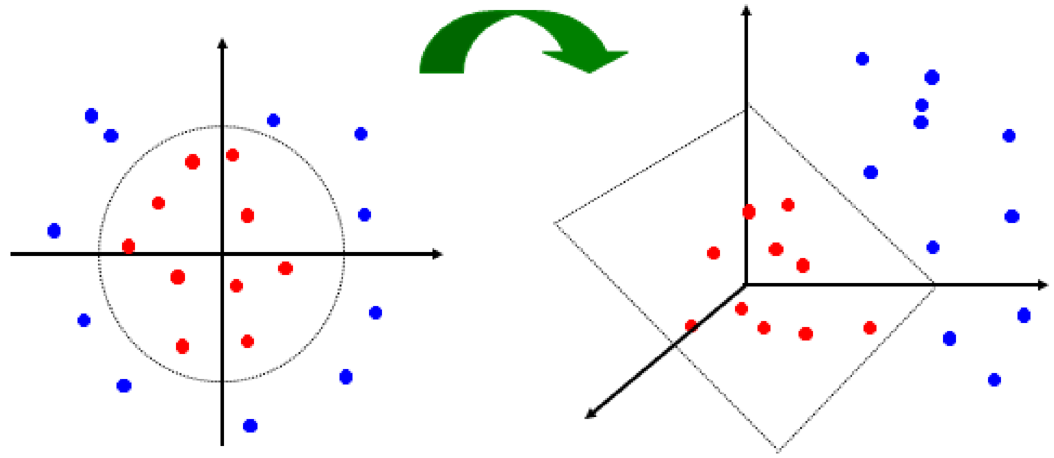
$$\begin{pmatrix} 0.18 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} - 0.45 = 0$$

$$WX + b = 0 \quad \Rightarrow \quad 0.18X_1 = 0.45 \quad \Rightarrow \quad X_1 = 0.025$$



ماشین بردار پشتیبان غیرخطی

- ماشین بردار پشتیبان با اضافه کردن حقه‌ی کرنل برای حالت غیر خطی تعمیم داده شد.
- هدف SVM غیرخطی دسته بندی داده هایی است که به صورت خطی جداپذیر نیستند.
- داده ها به فضاهای بالاتر (فضای هیلبرت) انتقال می‌یابند که به صورت خطی جداپذیر باشند و با یک ابرصفحه، دسته‌بندی شوند.



فضای هیلبرت H:
- انتقال جبر بردارها از فضای دو یا سه بعدی به ابعاد متناهی یاحتی نامتناهی
- فضای برداری دارای ساختار ضرب داخلی

$$\Phi: R^d \longrightarrow \mathcal{H}$$

$$K(x_i, x_j) = \Phi(x_i).\Phi(x_j)$$

ماشین بردار پشتیبان غیرخطی

- برخی از توابع هسته (Kernel) پرکاربرد:
  - خطی
  - چندجمله‌ای(Polynomial)
  - (Gaussian radial base) rbf
  - سیگموئید(Sigmoid)

1. $K(x_i, x_j) = x_i . x_j$
2. $K(x_i, x_j) = (x_i . x_j + 1)^p$
3. $K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$
4. $K(x_i, x_j) = \tanh(\beta x_i . x_j + \delta)$

- داده‌ها از فضایی که در آن جداپذیر خطی نبودند به فضایی برده می‌شوند که جداپذیر خطی باشند.

$$\min \quad \frac{1}{2}\|w\|^2$$

$$\text{s.t.} \quad y_i(w.\Phi(x)+b) - 1 \geq 0 \qquad \forall i$$

ماشین بردار پشتیبان غیرخطی

- تبدیل مسئله مقید به مسئله نامقید با استفاده از روش لاگرانژ

1 $$\max_{\alpha_i} \min_{w,b} L_p = \frac{1}{2}\|w\|^2 - \sum_i \alpha_i(y_i(\overline{w}.\Phi(x) + b) - 1) \quad , \quad \alpha_i \geq 0$$

- با اعمال شرایط KKT روابط زیر به دست می‌آید:

$$\boxed{W = \sum \alpha_i Y_i \Phi(X_i)} \quad , \quad \sum_i \alpha_i Y_i = 0 \quad , \quad \boxed{b = Y_i - W^T \Phi(X_i)}$$

- دوگان مسئله بهینه سازی :

2 $$\max \quad L_D = \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \underbrace{K(x_i, x_j)}_{\Phi(x_i).\Phi(x_j)}$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq C,$$

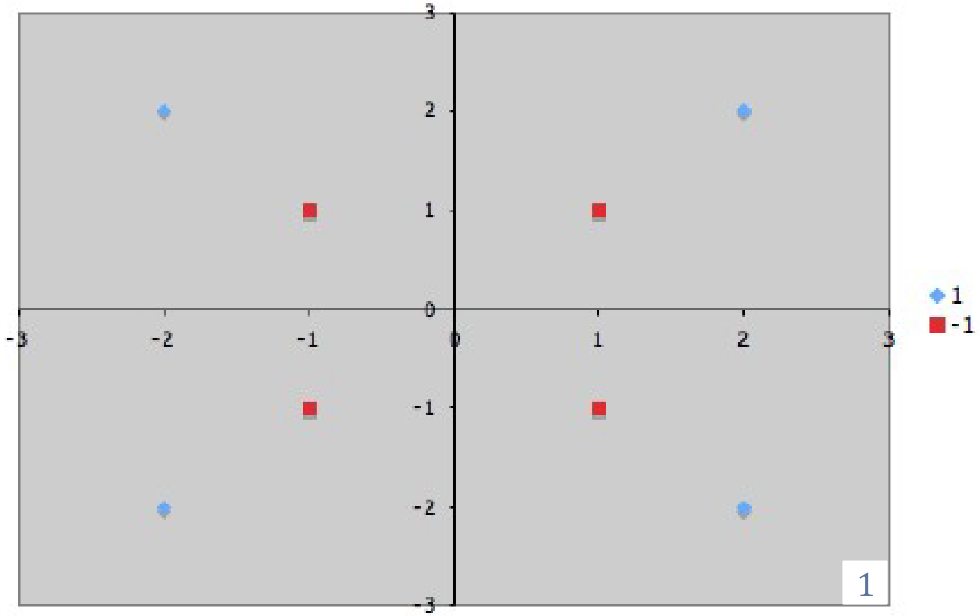
$$\sum_i \alpha_i y_i = 0$$

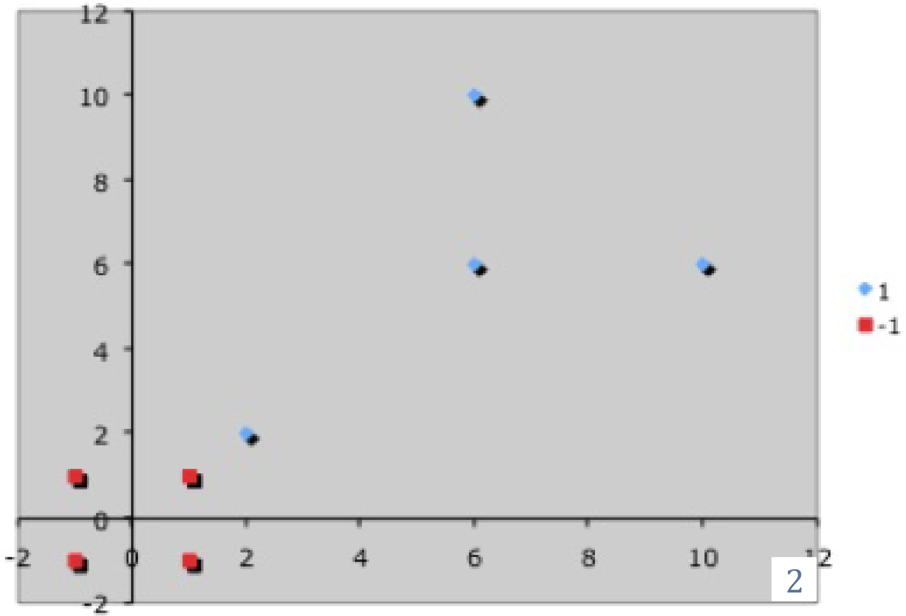
مثال SVM غیرخطی

برچسب مثبت: $\left\{\begin{pmatrix}2\\2\end{pmatrix},\begin{pmatrix}2\\-2\end{pmatrix},\begin{pmatrix}-2\\-2\end{pmatrix},\begin{pmatrix}-2\\2\end{pmatrix}\right\}$ → $\left\{\begin{pmatrix}2\\2\end{pmatrix},\begin{pmatrix}6\\10\end{pmatrix},\begin{pmatrix}6\\6\end{pmatrix},\begin{pmatrix}10\\6\end{pmatrix}\right\}$

برچسب منفی: $\left\{\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix},\begin{pmatrix}-1\\-1\end{pmatrix},\begin{pmatrix}-1\\1\end{pmatrix}\right\}$ → $\left\{\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix},\begin{pmatrix}-1\\-1\end{pmatrix},\begin{pmatrix}-1\\1\end{pmatrix}\right\}$

$$\Phi_1\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{cases}\begin{pmatrix}4-x_2+|x_1-x_2|\\4-x_1+|x_1-x_2|\end{pmatrix} & \text{if } \sqrt{x_1^2+x_2^2}>2\\ \begin{pmatrix}x_1\\x_2\end{pmatrix} & \text{otherwise}\end{cases}$$





مثال SVM غیرخطی

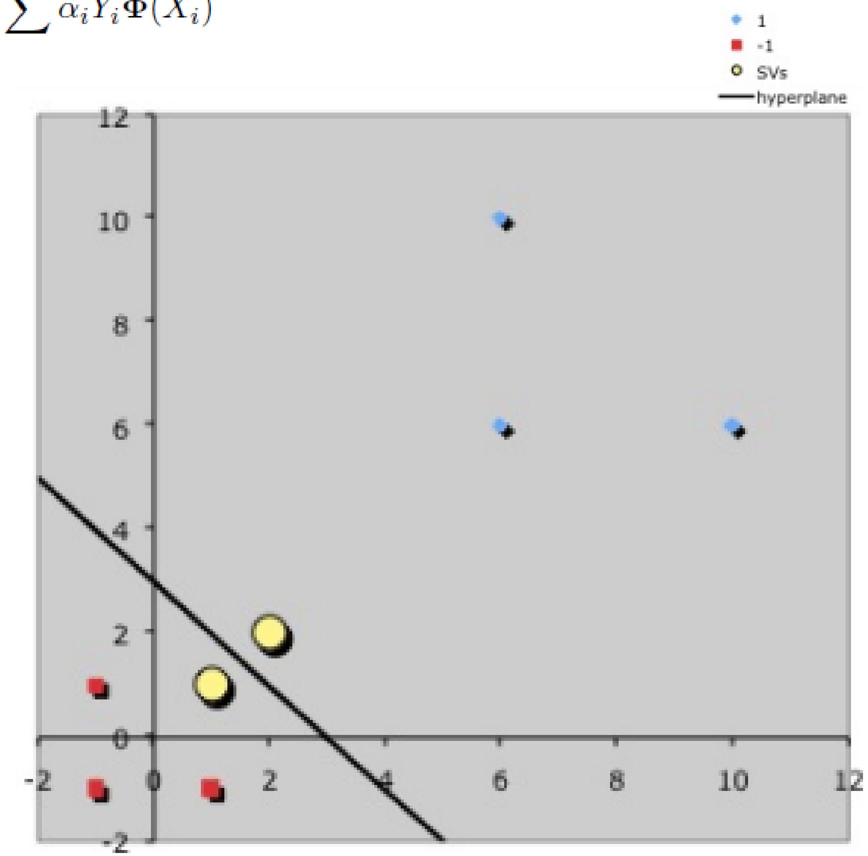
$$\left\{X_1 = \begin{pmatrix}1\\1\end{pmatrix}, X_2 = \begin{pmatrix}2\\2\end{pmatrix}\right\} \quad , \quad \left\{\Phi(X_1) = \begin{pmatrix}1\\1\end{pmatrix}, \Phi(X_2) = \begin{pmatrix}2\\2\end{pmatrix}\right\} \quad , \quad \sum_i \alpha_i Y_i = 0 \quad , \quad W = \sum \alpha_i Y_i \Phi(X_i)$$

$$\begin{cases} WX_1 + b = -1 \\ WX_2 + b = 1 \\ \alpha_1 Y_1 + \alpha_2 Y_2 = 0 \end{cases} \Rightarrow \begin{cases} [\alpha_1 Y_1 \Phi(X_1) + \alpha_2 Y_2 \Phi(X_2)]\Phi(X_1) + b = -1 \\ [\alpha_1 Y_1 \Phi(X_1) + \alpha_2 Y_2 \Phi(X_2)]\Phi(X_2) + b = 1 \\ \alpha_1 Y_1 + \alpha_2 Y_2 = 0 \end{cases}$$

$$\begin{cases} -2\alpha_1 + 4\alpha_2 + b = -1 \\ -4\alpha_1 + 8\alpha_2 + b = 1 \\ \alpha_1 = \alpha_2 \end{cases} \Rightarrow \begin{cases} \alpha_1 = 1 \\ \alpha_2 = 1 \\ b = -3 \end{cases}$$

$$W = \sum \alpha_i Y_i \Phi(X_i) = 1\begin{pmatrix}1\\1\end{pmatrix} + 1\begin{pmatrix}2\\2\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}$$

$$WX + b = 0 \Rightarrow \boxed{\begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}X_1\\X_2\end{pmatrix} - 3} = 0 \Rightarrow X_1 + X_2 = 3$$



ابزارهای پیاده‌سازی SVM

- متلب (MATLAB)

مجموعه‌ای از توابع پیاده سازی شده svm - struct – matlab

استفاده از بسته fitcsvm برای طبقه‌بندی

- ++C و جاوا

استفاده از بسته LIBSVM برای طبقه‌بندی ، رگرسیون و دسته‌بندی چند کلاسه

- پایتون (Python)

کتابخانه scikit-learn بسته svm.svc برای طبقه‌بندی

- R

استفاده از بسته e1071

- لینگو (Lingo)

حل مسائل بهینه‌سازی ماشین بردار پشتیبان با لینگو امکان‌پذیر است.

مراجع

1. Nayak, J., et al. (2015). "A comprehensive survey on support vector machine in data mining tasks: applications & challenges." International Journal of Database Theory and Application **8**(1): 169-186.
2. Wang, H., et al. (2017). "Research Survey on Support Vector Machine." People's Repub. China: 95-103.
3. Byun, H. and S.-W. Lee (2002). Applications of support vector machines for pattern recognition: A survey. International Workshop on Support Vector Machines, Springer.
4. فتاحی, م. مروری بر ماشین های بردار پشتیبان, دانشگاه رازی.
5. Suykens, J. A. and J. Vandewalle (1999). "Least squares support vector machine classifiers." Neural processing letters **9**(3): 293-300.
6. KOWALCZYK, A. (2018). "SVMs - An overview of Support Vector Machines." from https://www.svm-tutorial.com/2017/02/svms-overview-support-vector-machines/.
7. KOWALCZYK, A. (2018). "SVM - Understanding the math - the optimal hyperplane." from https://www.svm-tutorial.com/2015/06/svm-understanding-math-part-3/.
8. KOWALCZYK, A. (2018). "SVM - Understanding the math - Unconstrained minimization." from https://www.svm-tutorial.com/2016/09/unconstrained-minimization/.
9. (1397). "آشنایی با ماشین بردار پشتیبان – SVM مرور" https://www.outlier.ir/2017/06/07/svm-tutorial-overview/.

سپاس
از حسن توجه
شما عزیزان