1

عنوان سمینار:

طبقه بندی با استفاده از ماشین بردارپشتیبان

classification by support vector machine

ارائه دهنده: زهره منوچهری

۲۵خرداد۱۳۹۴

رئوس مطالب

هوش مصنوعی

- هوش مصنوعی به هوشی که یک ماشین از خود نشان می دهد و یا به دانشی در کامپیوترکه سعی در ایجاد آن دارد گفته می شود.

- هوش مصنوعی علم و مهندسی ایجاد ماشینهایی با هوش ،با به کارگیری از کامپیوتر و الگوگیری از درک هوش انسانی و یا حیوانی و نهایتاً دستیابی به مکانیزم هوش مصنوعی در سطح هوش انسانی می‌باشد.

توانمندی های هوش مصنوعی در پزشکی

۱-شناسایی و تشخیص بیماری

- هنگامی که بیماری فرد پیچیده و نادر است یا تشخیص بیماری بر اساس بی تجربگی بوده سیستم می تواند در تشخیص به پزشک کمک کند

۲-سیستم های تخصصی اطلاعات ازمایشگاه

- این سیستم ها نتایج مربوط به بیماری ها را به طور اتوماتیک تفسیر و نتایج مربوط به ان را اعلام می کنند.

۳- تشخیص و تفسیر تصاویر پزشکی

مثلا سیستم می تواند تصاویر غیر طبیعی MRI یا انژیو گرافی غیر طبیعی را برای متوجه نمودن پزشک علامت دار کند.

۴-هوش مصنوعی به جای پزشک در بخش مراقبت های ویژه

تیم مهندسین یک سیستم هوشمندی را ایجاد کردند تا همانند مغز یک پزشک ، نشانه های حیاتی بیمار را کنترل و سپس ان را ارزیابی و مقدار مناسب دارو های مختلف مورد نیاز را به بیمار می رساند.

svm

یادگیری ماشین

- یادگیری ماشین زمینه نسبتا جدیدی از هوش مصنوعی است که در حال حاضر دوران رشد و تکامل خود را می گذراند. یادگیری ماشین یک زمینه تحقیقاتی بسیار فعال در علوم کامپیوتر است.

- علوم مختلفی با یادگیری ماشین در ارتباط هستند از جمله:

هوش مصنوعی، روانشناسی، فلسفه، تئوری اطلاعات ، آمار و احتمالات، تئوری کنترل و ...

تعریف یادگیری ماشین:

- یادگیری ماشین عبارت است از اینکه چگونه می توان برنامه ای نوشت که از طریق تجربه یادگیری کرده و عملکرد خود را بهتر کند.

در یادگیری ماشین با استفاده از تئوری اطلاعات ،مدلهای ریاضی ساخته می شود که می توانند برای استنتاج استفاده شوند.

دسته بندی یادگیری

- یادگیری با ناظر
- یادگیری بدون ناظر
- یادگیری تقویتی
- یادگیری نیمه نظارتی

یادگیری با ناظر

- یادگیری تحت نظارت، یک روش عمومی در یادگیری ماشین است که در آن به یک سیستم، مجموعه‌ای از جفت‌های ورودی – خروجی ارائه شده و سیستم تلاش می‌کند تا تابعی از ورودی به خروجی را فرا گیرد. یادگیری تحت نظارت نیازمندتعدادی داده ورودی به منظور آموزش سیستم است.

مثالی از یادگیری با ناظر

- یک data base شامل قیمت ۵ خانه و مساحت خانه ها وجود دارد چگونه می توان نرخ خانه ها را بر اساس تابعی از اندازه آنها یاد گرفت؟

مساحت	قیمت
۵۶۰	۳۷
۱۰۱۲	۷۹
۸۹۳	۷۶
۲۱۹۶	۱۳۰
۹۳۶	۸۲

مساحت خانه= x(i) Input feature

قیمت خانه =y(i) Output feature

- عمل یادگیری:

با داشتن یک مجموعه یادگیری میخواهیم تابعی بصورت h:x $\rightarrow$ y یاد بگیریم که

- h(x)بتواند مقدار yرا بخوبی حدس بزند. این تابع فرضیه و یا hypothesisنامیده می شود.

انواع یادگیری

Classification: ماشین یاد می گیرد ورودی ها را به دسته های از پیش تعیین شده نسبت دهد.

Clustering : سیستم یادگیر کشف می کند که کدام ورو دیها با هم در یک دسته قرار می گیرند.

numeric prediction : ماشین یاد می گیرد به جای تعیین دسته بندی یک ورودی مقدار عددی ان را پیش بینی نماید.

رگرسیون و دسته بندی

- رگرسیون:

وقتی که تابع هدف بصورت پیوسته باشد مسئله یادگیری یک مسئله رگرسیون خواهد بود. مثل یادگیری رابطه قیمت و مساحت خانه ها

- دسته بندی (classification):

وقتی که yبتواند تعداد محدودی مقدار گسسته بگیرد مسئله یادگیری یک مسئله دسته بندی خواهد بود. مثل: آیا خانه مورد نظر یک آپارتمان است؟

تعریف: ماشین بردار پشتیبان

- ماشین بردار پشتیبان (support vector machines) یکی از روش های یادگیری با نظارت است که از ان برای طبقه بندی و رگرسیون استفاده می شود.

- این روش از جمله روشهای نسبتا جدیدی است که در سال های اخیر کارایی خوبی نسبت به روش های قدیمی تر برای طبقه بندی از جمله شبکه های عصبی پرسپترون نشان داده است.

تاریخچه

الگوریتم svm اولیه در سال ۱۹۶۳ توسط

Vladimir vapnik ابداع شد و در سال

۱۹۹۵توسط او و همکارش   corenna

cortes برای حالت غیر خطی تعمیم داده

شد.

ایده اصلی

- با فرض اینکه دسته ها بصورت خطی جداپذیر باشند، ابرصفحه هائی با حداکثر حاشیه (maximum margin) را بدست می آورد که دسته ها را جدا کنند.

- در مسایلی که داده ها بصورت خطی جداپذیر نباشند داده ها به فضای با ابعاد بیشتر نگاشت پیدا می کنند تا بتوان آنها را در این فضای جدید بصورت خطی جدا نمود.

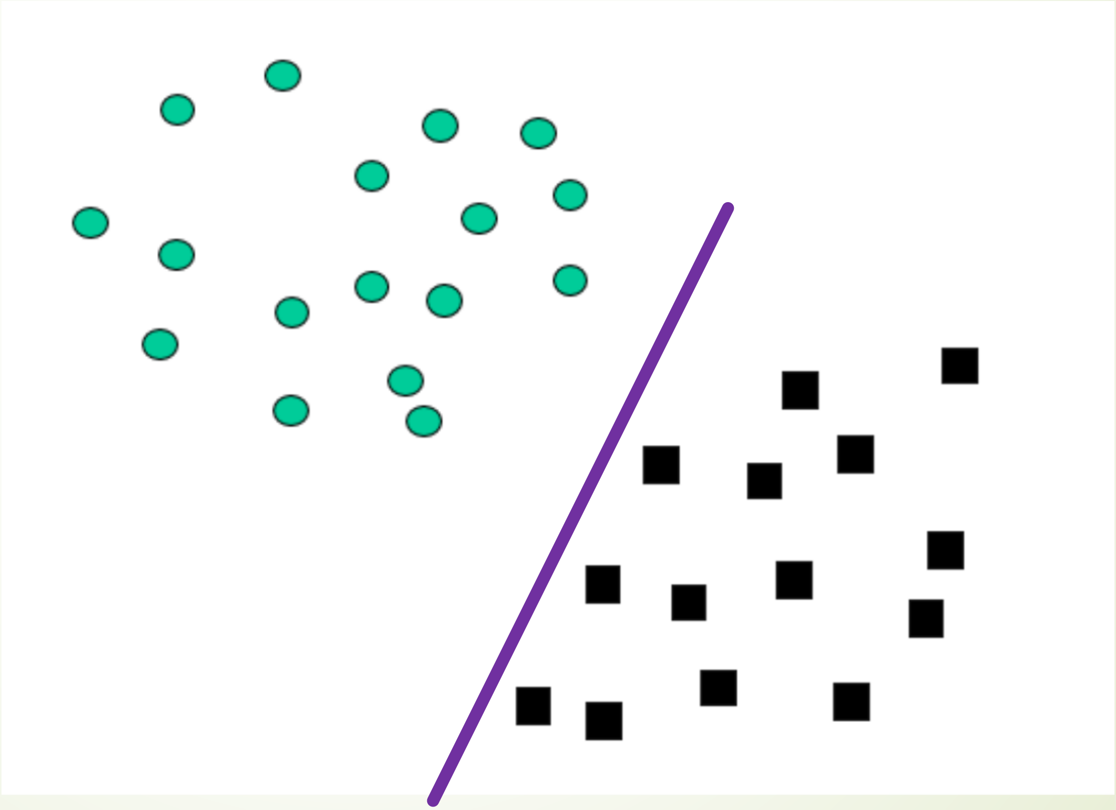
مفهوم ابر صفحه

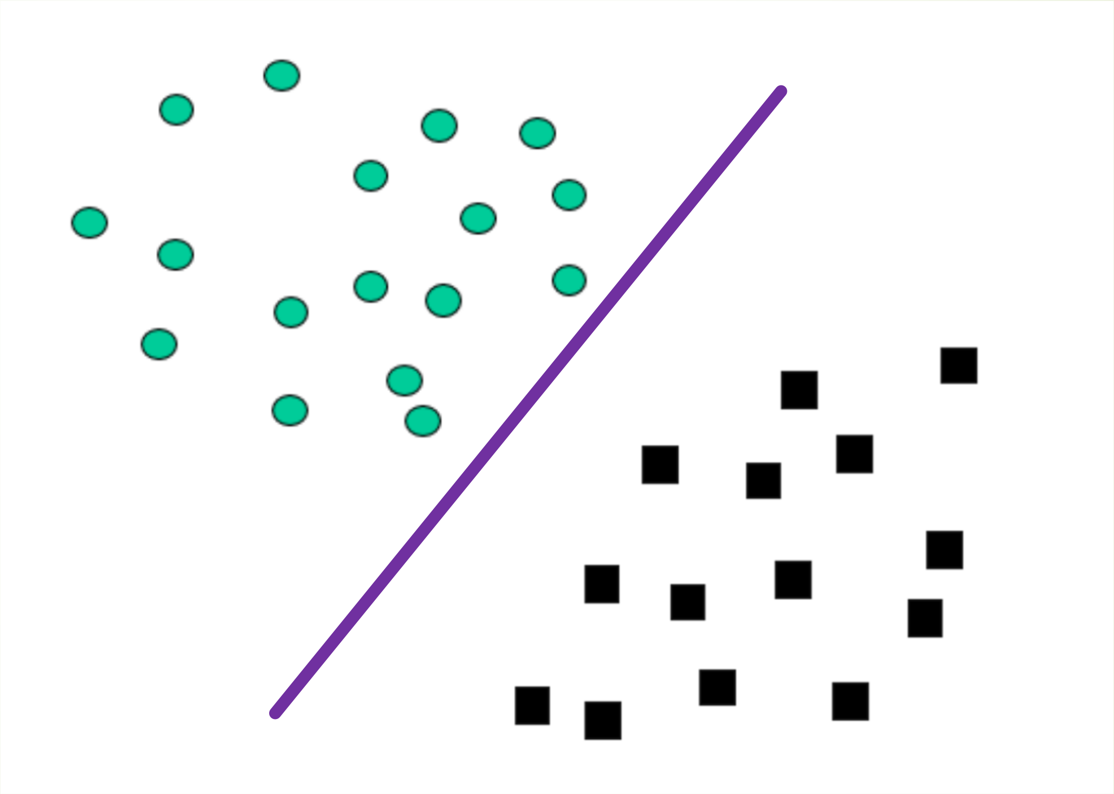
- ابرصفحه[1] یک مفهوم در هندسه است، یک تعمیم از یک صفحه در تعداد متفاوتی از ابعاد. یک ابرصفحه یک زیرفضای k بعدی در یک فضای n بعدی تعریف می کند که k<n . مثلا خط یک ابرصفحه یک بعدی در یک فضا با هر تعداد بعد است. در سه بعد، صفحه یک ابرصفحه دوبعدی است و به همین ترتیب برای فضاهای با ابعاد بالاتر ابرصفحه تعریف می شود.
- hyperplane

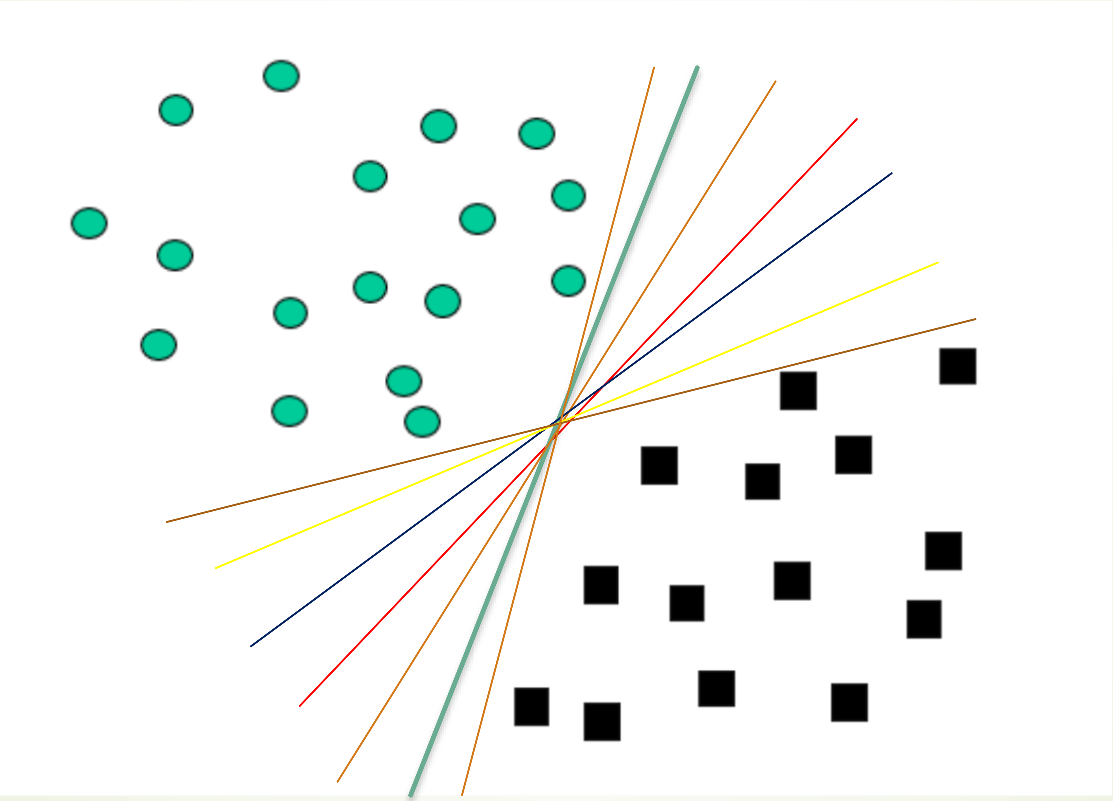
عمل جداسازی خطی

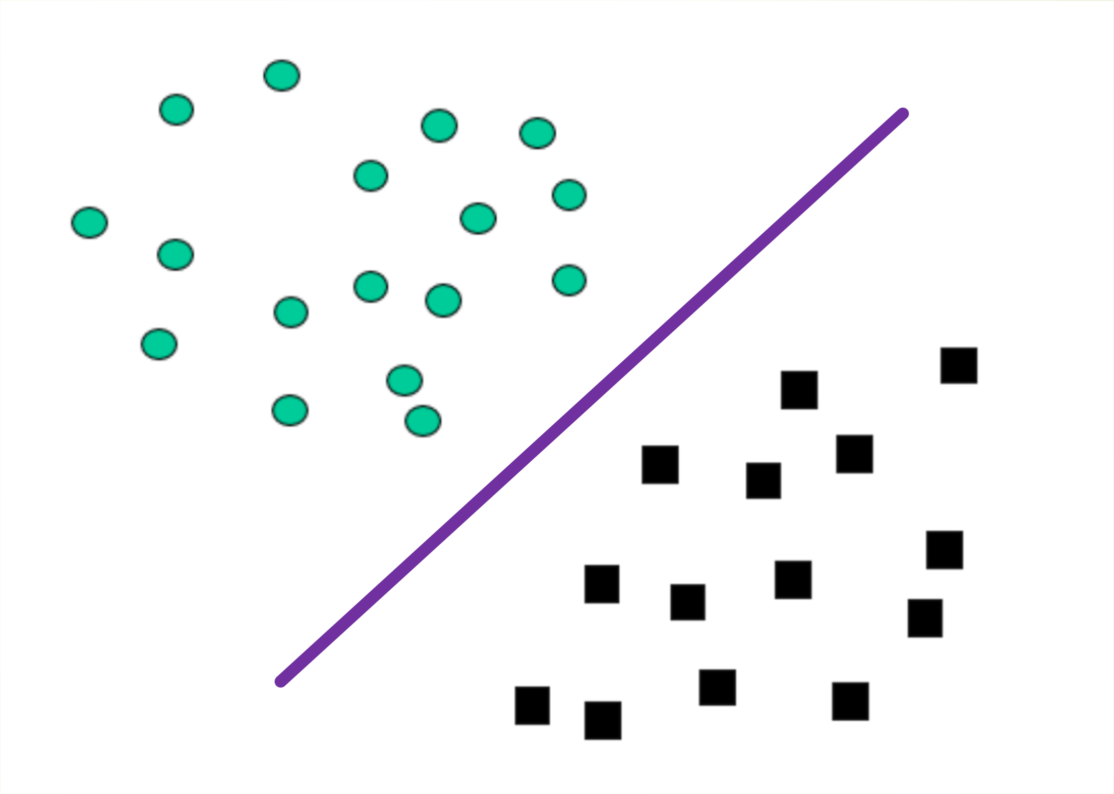
- در SVM یک داده به صورت یک بردار P بعدی(یا یک لیست از P عدد) دیده می شود و ما می خواهیم بدانیم می توان چنین نقاطی را با یک ابرصفحه P-1 بعدی جدا کرد. این عمل جداسازی خطی نامیده می شود. ابرصفحه های بسیاری وجود دارند که می توانند داده ها را جدا کنند. الگوریتم های مختلفی از جمله پرسپترون می توانند این جداسازی را انجام دهند.سوال اینجاست که چه ابرصفحه ای را انتخاب کنیم؟

Which of the linear separators is optimal?

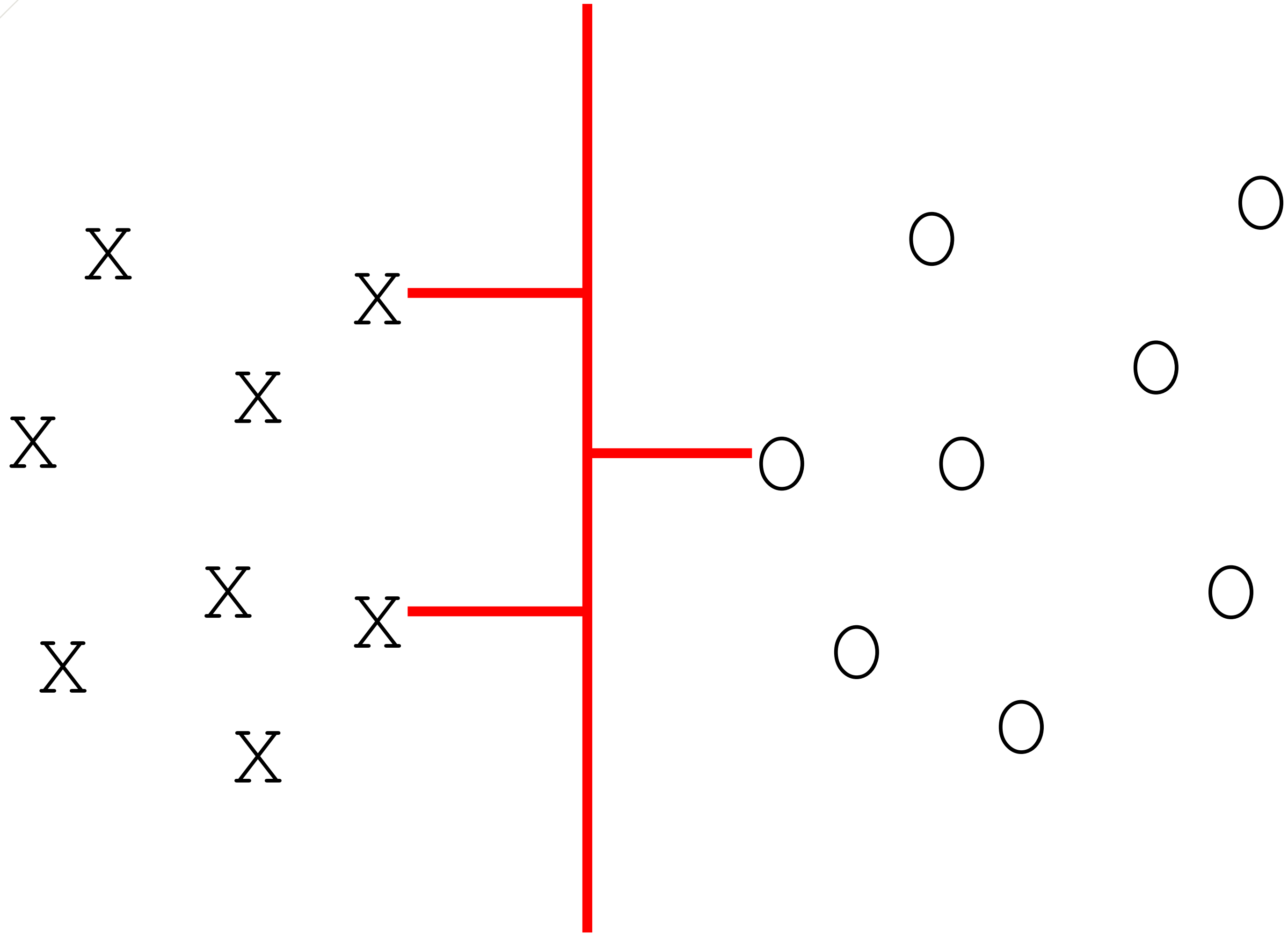






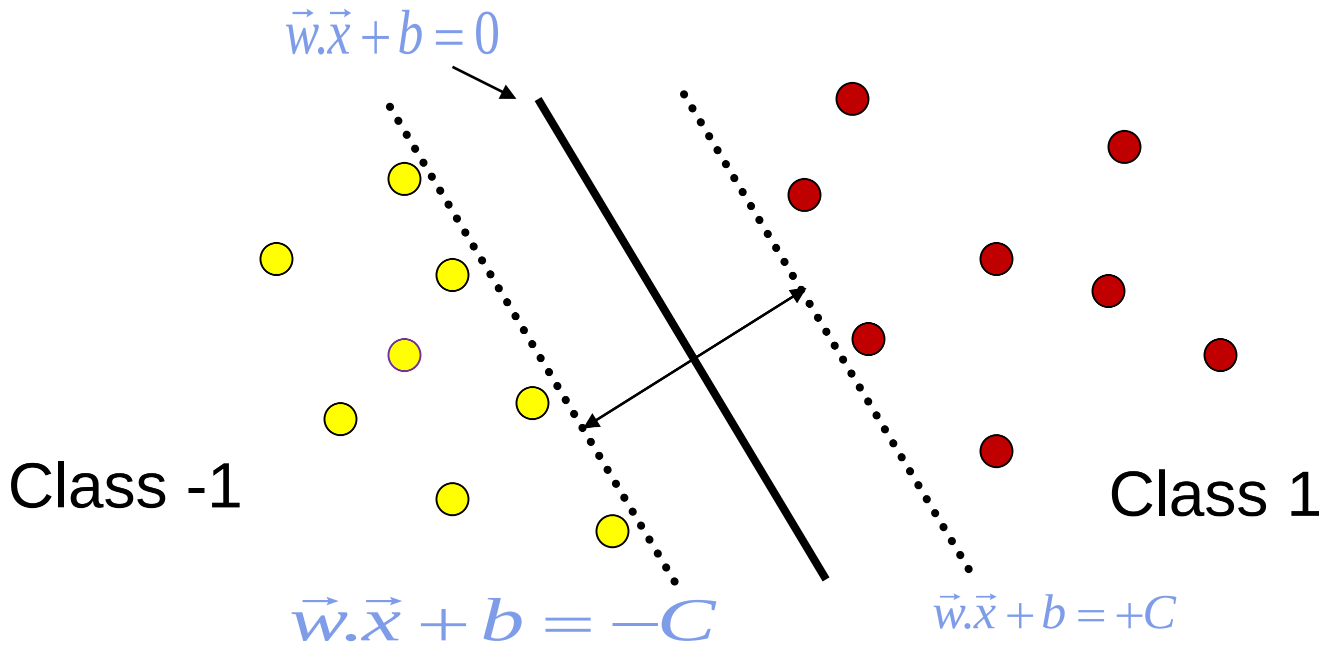


Maximizing the Margin



ایده SVM برای جدا سازی دسته ها

- دو صفحه مرزی موازی با صفحه دسته بندی رسم کرده و آندو را آنقدر از هم دور میکنیم که به داده ها برخورد کنند.
- صفحه دسته بندی که بیشترین فاصله را از صفحات مرزی داشته باشد، بهترین جدا کننده خواهد بود.



معادله خط در حالت دو بعدی

$$w_1X_1 + w_2X_2 + b = 0$$

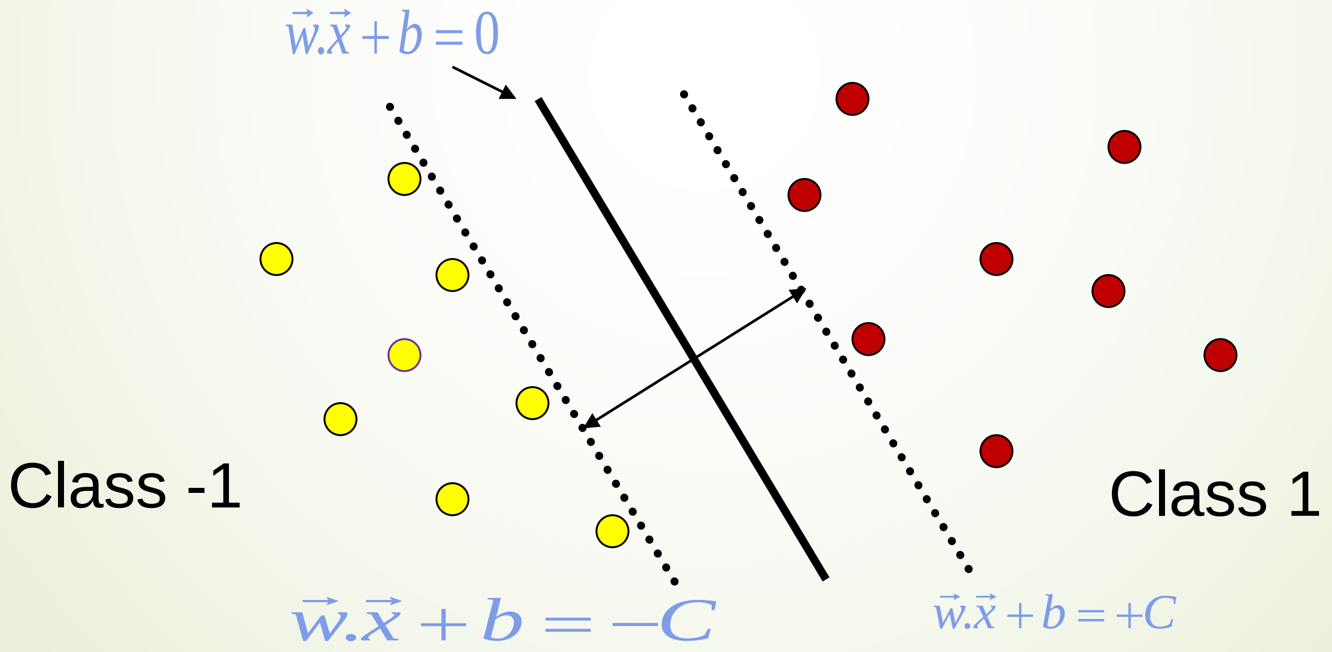
$$\sum_{i=0}^{n} wi.xi + b = 0$$

معادله خط در حالت n بعدی

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

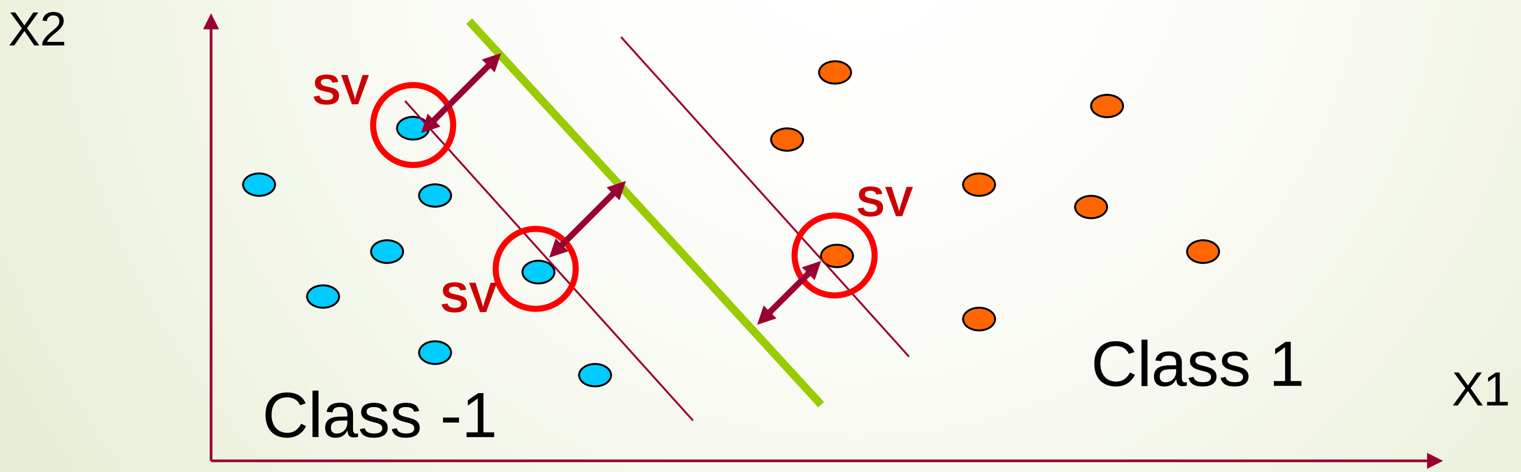
چرا حداکثر حاشیه؟

- بر طبق قضیه ای در تئوری یادگیری اگر مثالهای آموزشی بدرستی دسته بندی شده باشند، از بین جداسازهای خطی، آن جداسازی که حاشیه داده های آموزشی را حداکثر میکند خطای تعمیم را حداقل خواهد کرد.



بردار پشتیبان

- نزدیکترین داده های آموزشی به ابر صفحه های جدا کننده بردار پشتیبان نامیده می شوند.



فرموله کردن مسئله:

- تعدادی داده ی آموزشی داریم، یک مجموعه از نقاط به فرم:

- $D = \{(x_i, y_i) | x_i \in \mathbb{R}^p, y_i \in \{1, -1\}\} \quad i = 1, \dots, n$

که y_i می تواند 1 یا *1*- باشد، که مشخص می کند نقطه x_i متعلق به چه کلاسی است. هر x_i یک بردار p بعدی است. می خواهیم جداکننده ماکزیمم حاشیه ای پیدا کنیم که نقاطی که y_i آنها برابر *1* است از نقاطی که y_i آنها *1*- است جدا کند. هر ابرصفحه می تواند به صورت یک مجموعه از نقاط x که در رابطه زیرصدق می کنند نوشته شود:

$$\langle w, x \rangle + b = 0$$

$$w_1x_1 + w_2x_2 \dots + w_nx_n + b = 0$$

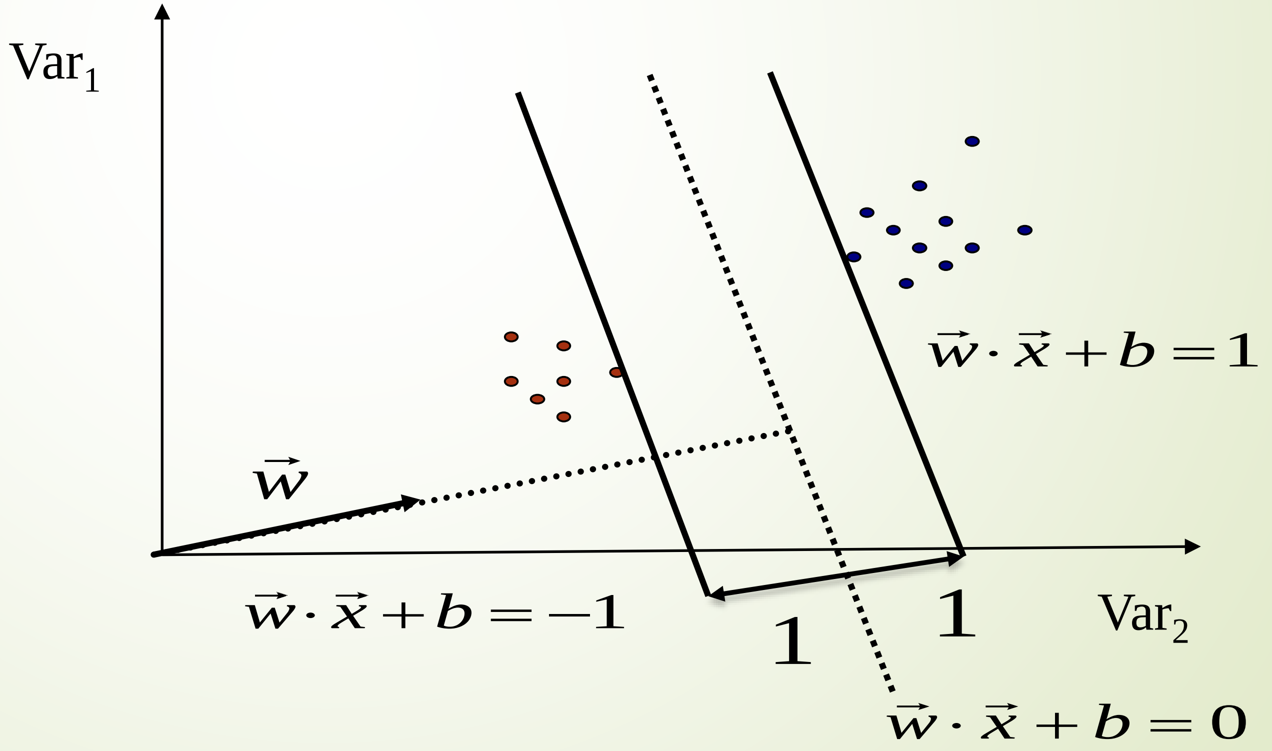
- میخواهیم مقادیر W, bرابگونه ای پیدا کنیم که:
- نمونه های آموزشی را بدقت دسته بندی کند .
- با این فرض که داده ها بصورت خطی جدا پذیر باشند فاصله بین ابر صفحه های موازی یعنی پهنای حاشیه را حداکثر نماید.

معادلات ابر صفحه ها

- این ابرصفحه های موازی می توانند با معادله های زیر توصیف شوند:
- $W.X + b \geq 1 \quad \forall\, x_i \; with \; y_i = 1$

و

- $W.X + b \leq 1 \quad \forall\, x_i \; with \; y_i = -1$



- با استفاده از روابط هندسی پهنای حاشیه برابر است با $\frac{2}{\|w\|}$. بنابراین می خواهیم $\|w\|$را مینیمم کنیم. همچنین می خواهیم از قرارگرفتن نقاط در حاشیه جلوگیری کنیم.

محدودیت زیر را نیز اضافه می کنیم:

برای هر i :

- $W.X + b \geq 1 \quad \forall\, x_i \; with \; y_i = 1$

$$y_i(w.x_i + b) \geq 1$$

- $W.X + b \leq 1 \quad \forall\, x_i \; with \; y_i = -1$

- حل مسئله بهینه سازی $\frac{2}{\|w\|}$ مشکل است، زیرا به $\|w\|$ بستگی دارد. خوشبختانه می توان معادله را با توجه به رابطه ی $\frac{1}{2}\|w\|^2 = \|w\|$ بدون تغییر راه حل، جایگزین کرد (w و b های یکسانی بدست خواهد آمد).

هدف

- Minimise $\|w\|^2$
- Subject to : $y_i (\langle w,x_i\rangle + b) \geq ۱$ for all i
  - Note that $\|\mathbf{w}\|^2 = \mathbf{w}^T\mathbf{w}$

این یک مسئله quadratic programming با محدودیت هائی بصورت نامعادلات خطی است. روشهای شناخته شده ای برای چنین مسئله هائی بوجود آمده اند.

ضرایب لاگرانژ

ضرایب لاگرانژ یک استراتژی برای پیدا کردن مینیمم یا ماکزیمم یک تابع با توجه به محدودیت ها است.

$$\begin{cases} \frac{\partial}{\partial \mathbf{x}}\left(f(\mathbf{x}) + \alpha g(\mathbf{x})\right)\Big|_{\mathbf{x}=\mathbf{x}_0} = 0 \\ g(\mathbf{x}) = 0 \end{cases}$$

$$\begin{cases} \frac{\partial}{\partial \mathbf{x}}\left(f(\mathbf{x}) + \sum_{i=1}^{n} \alpha_i g_i(\mathbf{x})\right)\Big|_{\mathbf{x}=\mathbf{x}_0} = 0 \\ g_i(\mathbf{x}) = 0 \qquad \text{for } i = 1, \ldots, m \end{cases}$$

- در مواردیکه $g_i(x) \le 0$ ضریب لاگرانژ α_i باید مثبت شود.

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad g_i(\mathbf{x}) \le 0 \quad \text{for } i = 1, \ldots, m$$

$$\begin{cases} \frac{\partial}{\partial \mathbf{x}}\left(f(\mathbf{x}) + \sum_i \alpha_i g_i(\mathbf{x})\right)\Big|_{\mathbf{x}=jx_0} = 0 \\ g_i(\mathbf{x}) \le 0 \qquad \text{for } i = 1, \ldots, m \end{cases}$$

راه حل معادله

- باید تابع لاگرانژ زیر را می نیمم کنیم:

$$L(\mathbf{w},b,\alpha)=\frac{1}{2}\|\mathbf{w}\|^2-\sum_{i=1}^{N}\alpha_i[y_i(\mathbf{w}\mathbf{x}_i+b)-1]$$

$$\text{wrt.constraint } \alpha_i \geq 0,\ i=1,\ldots N$$

و بنابراین از این تابع نسبت به w وb مشتق می گیریم:

$$\frac{\partial L(\mathbf{w},b,\alpha)}{\partial \mathbf{w}}=\mathbf{w}-\sum_{i=1}^{N}\alpha_i y_i \mathbf{x}_i=0$$

$$\frac{\partial L(\mathbf{w},b,\alpha)}{\partial b}=\sum_{i=1}^{N}\alpha_i y_i=0$$

$$KKT\ cond:\alpha_i[y_i(\mathbf{w}\mathbf{x}_i+b)-1]=0$$

The Dual Problem

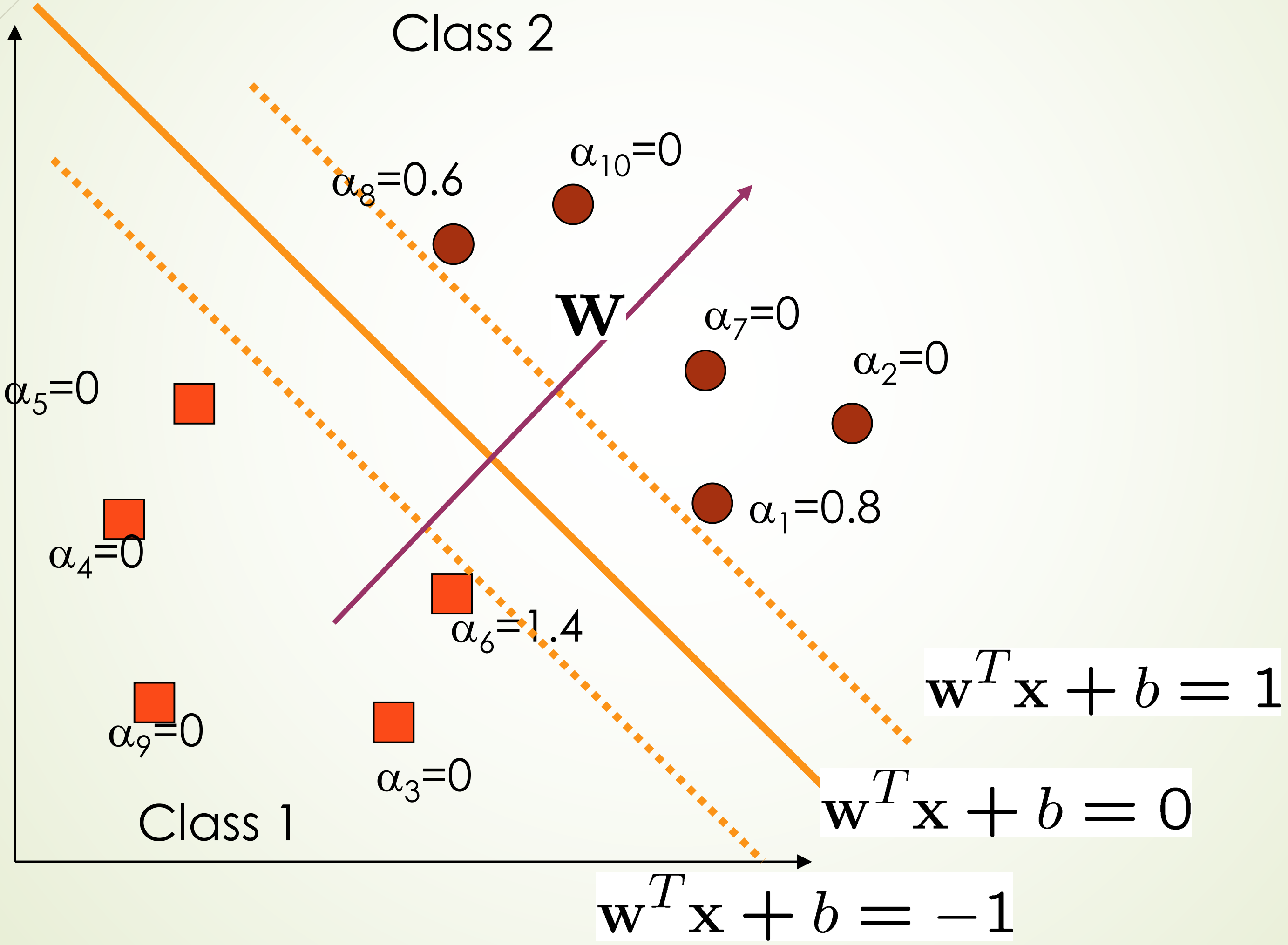
اگر ما $\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$ و $\sum_{i=1}^{n} \alpha_i y_i = 0$ را در لاگرانژ قرار دهیم خواهیم داشت:

$$
\begin{aligned}
\mathcal{L} &= \frac{1}{2} \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i^T \sum_{j=1}^{n} \alpha_j y_j \mathbf{x}_j + \sum_{i=1}^{n} \alpha_i \left(1 - y_i \left(\sum_{j=1}^{n} \alpha_j y_j \mathbf{x}_j^T \mathbf{x}_i + b \right) \right) \\
&= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \alpha_i y_i \sum_{j=1}^{n} \alpha_j y_j \mathbf{x}_j^T \mathbf{x}_i - b \sum_{i=1}^{n} \alpha_i y_i \\
&= -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^{n} \alpha_i
\end{aligned}
$$

که این تنها تابعی از α_i خواهد بود.

به این دلیل dual Problem نامیده می شود چون اگر w را پیدا کنیم همه α_i بدست می ایند و بلعکس.

ضرایب α_i متغیر های دو گان (dual variables) نامیده می شوند .
هر کدام از نقاط اموزشی ($\mathbf{x}_i$) متغیر دو گان مربوط به خود را دارند.



The Dual Problem

مسئله اصلی primal problem نامیده می شود.

بنابراین dual Problem به صورت زیر است که باید maximum شود!!!

$$\text{max.}\quad W(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$$\text{subject to } \alpha_i \geq 0, \qquad \sum_{i=1}^{n} \alpha_i y_i = 0$$

که این یک مسئله Quadratic Programing است.

بعد از اینکه α_i بدست امد w نیز از $\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$ بدست می اید. b نیز با جایگذاری در معادله اولیه بدست می اید.

دسته بندی داده های جدید

- پس از آنکه مقادیر (a*, b*)با حل معادلات quadratic بر اساس داده های ورودی بدست آمد، می توان SVMرا برای دسته بندی نمونه های جدید بکار برد.
اگر xیک نمونه جدید باشد، دسته بندی آن بصورت زیر مشخص می شود:

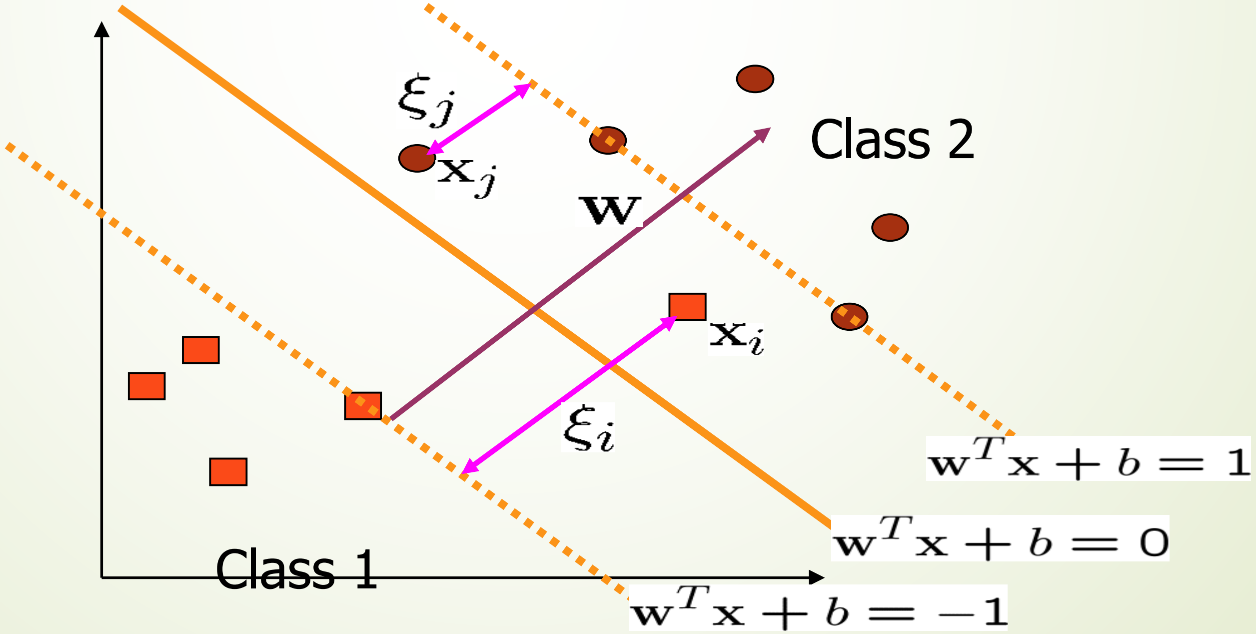
- sign[f(x, a*, b*)], where

$$f(\mathbf{x},\alpha^*,b^*)=\mathbf{w}^*\mathbf{x}+b^*=\sum_{i=1}^{N}\alpha_i^* y_i \mathbf{x}_i\mathbf{x}+b^*=\sum_{i\in SV}\alpha_i^* y_i \mathbf{x}_i\mathbf{x}+b^*$$

Data enters only
in the form of
dot products!

متغیر slack

- اگر داده ی آموزشی به صورت خطی قابل جداسازی نبود متغیر کمکی را می توان اضافه کرد تا اجازه ی جداسازی اشتباه را برای نمونه های نویزی بدهد. متغیرهای کمکی اجازه می دهند بعضی از نمونه ها درون حاشیه قرار بگیرند اما آنها را جریمه می کند.این متغیرها انحراف یک نقطه داده از جداکننده ایده آل آن را اندازه می گیرند.
- این کار با معرفی متغیر ξ_i انجام می شود که متغیر slack نامیده می شو د که تقریبی از

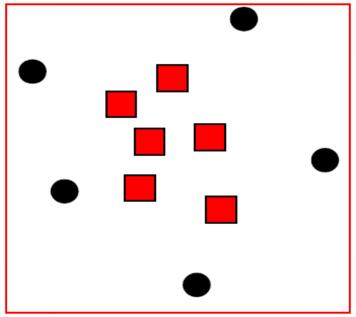


متغیر slack

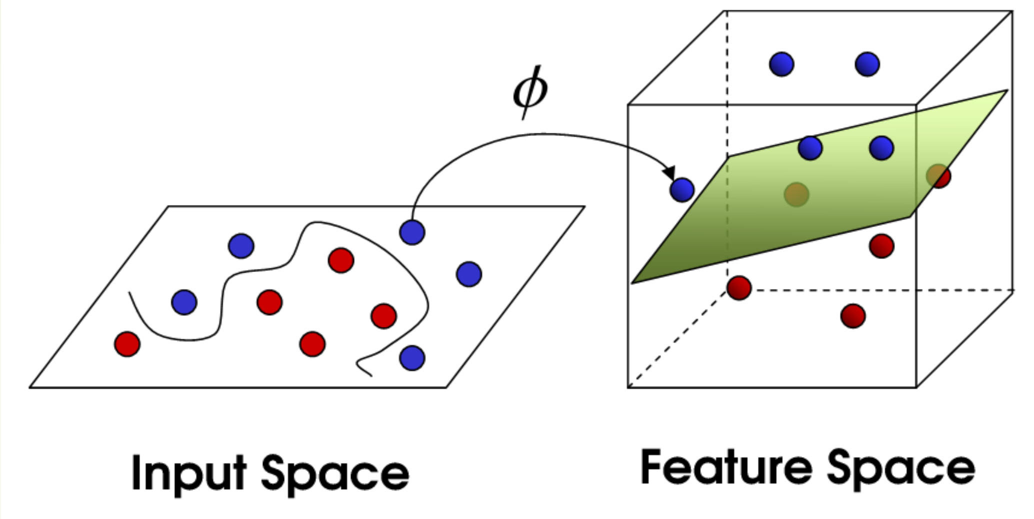
- فرمول جدید شامل متغیرهای کمکی به صورت زیر درمی آید:
- $\phi(\overrightarrow{w},\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$
- $y_i(w.x_i + b) \geq 1 - \xi_i$
- عبارت $C\sum_{i=1}^{N}\xi_i$ حداکثر تعداد خطا را تعیین می کند. هرچه c بزرگ تر باشد حساسیت بیشتری نسبت به خطاهایی که اجازه می دهیم اتفاق بیافتد، در نظر گرفته می شود.به c پارامتر تنظیم کننده گفته می شود و باید توسط کاربر انتخاب شود.
- برای مسئله فوق نیز تکنینک های برنامه نویسی درجه دو استفاده شده مقادیر α به دست می آید.

svm های غیر خطی:فضای ویژگی(Feature Space)

- مجموعه داده هایی که به طور خطی قابل جداسازی هستند با مقداری نویز خوب کار می کنند. اما اگر مجموعه داده ما خیلی پیچیده بود چه؟
- شکل زیر را در نظر بگیرید:



- این مجموعه داده را نمی توان به سادگی با یک خط دسته بندی کرد. در این موارد از نگاشت به فضایی با ابعاد بالاتر استفاده می شود.



چرا انتقال می دهیم؟

• عملگرهای خطی در فضای ویژگی معادل عملگرهای غیرخطی در فضای ورودی هستند(در نتیجه با نگاشت تغییری در مسئله ایجاد نمی کنیم).

• جداسازی با انتقال مناسب راحت تر است.

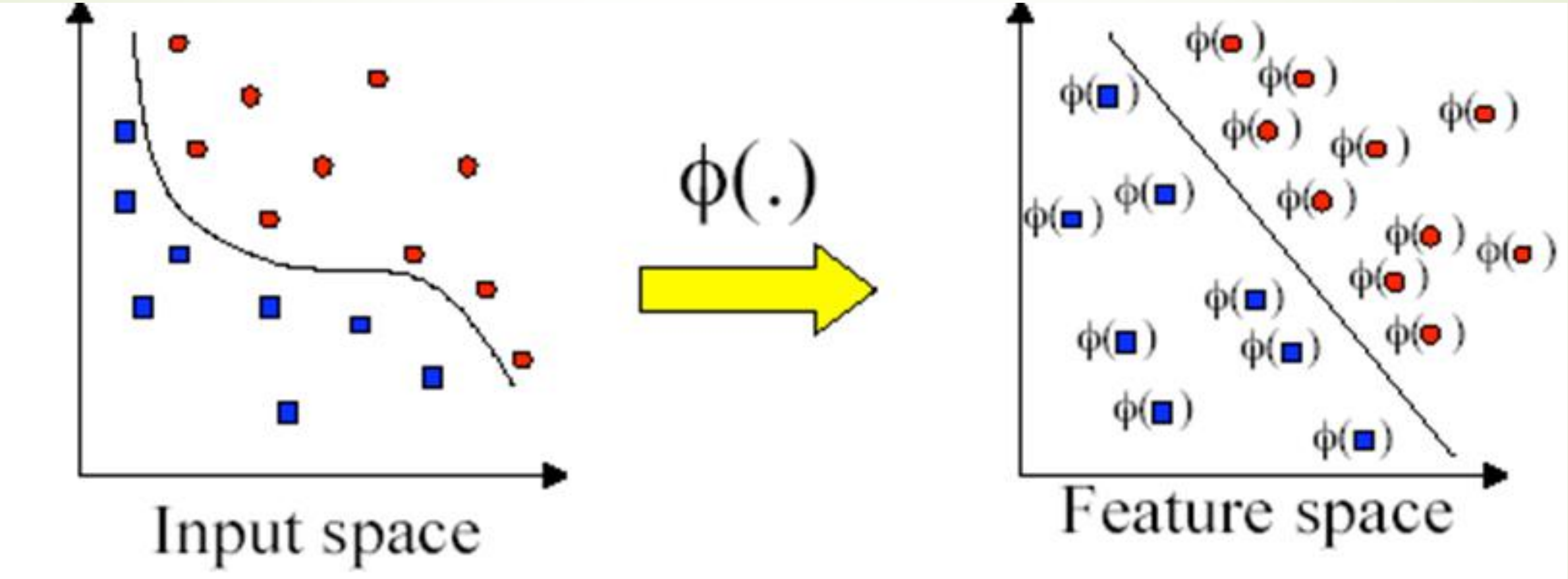
- برای انتقال ابتدا تابع $\varphi(x)$ را برای نگاشت به فضای دیگر،پیدا می کنیم. سپس فرمول SVM به صورت زیر تغییر می یابد:
- $\min \frac{1}{2} \|\mathrm{w}\|^2 + \mathrm{C} \sum_{\mathrm{i}=1}^{\mathrm{N}} \xi_{\mathrm{i}}$
- با توجه به محدودیت زیر:
- $y_{\mathrm{i}}(\mathrm{w.x_i} + \mathrm{b}) \geq 1 - \xi_i , \forall x_i \quad , \xi_i \geq 0$
- فضای نگاشت اصلی همواره می تواند به فضای ویژگی با ابعاد بالاتر دیگری که مجموعه آموزشی قابل جداسازی باشد،نگاشت یابد. انجام محاسبات در فضای ویژگی چون ابعاد بیشتری دارد می تواند پرهزینه باشد. در حالت کلی این فضا بی نهایت است. برای غلبه بر این مشکل از حقه کرنل استفاده می شود.
- Kernel trick

تابع کرنل

- تابع کرنل ،یک جداکننده خطی متکی بر ضرب داخلی بردارهاست که به صورت
- $k(x_i,x_j) = {x_i}^T x_j$ می باشد. اگر نقاط با استفاده از انتقال $\varphi: x \to \varphi(x)$ به فضای ویژگی (فضای با ابعاد بالاتر) انتقال یابند، ضرب داخلی آنها به صورت
- $k(x_i,x_j) = \varphi(x_i)^T . \varphi(x_j)$

تبدیل خواهد شد. بنابراین ما از محاسبه ی $\varphi(x)$ خودداری کرده و به جای آن از تابع کرنل استفاده می کنیم.

- انتخاب کرنل های غیرخطی به ما اجازه ی ساخت جداکنندهای خطی در فضای ویژگی را می دهد در صورتی که آنها در فضای اصلی غیرخطی هستند.



متداول ترین توابع کرنل

- Linear kernel: → $$K(\mathbf{x}_i,\mathbf{x}_j)=\mathbf{x}_i^T\mathbf{x}_j$$
- Polynomial kernel: → $$K(\mathbf{x}_i,\mathbf{x}_j)=(1+\mathbf{x}_i^T\mathbf{x}_j)^p$$
- Gaussian (Radial-Basis Function (RBF)) kernel: → $$K(\mathbf{x}_i,\mathbf{x}_j)=\exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{2\sigma^2}\right)$$
- Sigmoid: → $$K(\mathbf{x}_i,\mathbf{x}_j)=\tanh(\beta_0\mathbf{x}_i^T\mathbf{x}_j+\beta_1)$$

- با استفاده از توابع کرنل مسئله به فرم زیر تبدیل خواهد شد:

- $max.\ Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j k(\vec{x}_i, \vec{x}_j)$

- $s.t: \sum_{i=1}^{N} \alpha_i y_i = 0 \quad ,0 \leq \alpha_i \leq C \ for\ i = 1, \ldots, N$

کاربرد های svm

- در اقتصاد : پیش بینی رکود
- تشخیص صورت :دسته بندی و کاهش ویژگی برای تشخیص سریع صورت
- انتخاب ژن در دسته بندی سرطان
- شناسایی گفتار
- شناسایی الگو
- ...

مزایا و معایب svm

- آموزش نسبتا ساده است
- برای داده های با ابعاد بالا تقریبا خوب جواب می دهد.
- مصالحه بین پیچیدگی دسته بندی کننده و میزان خطا به طور واضح کنترل می شود.
- به یک تابع کرنل خوب و انتخاب پارامتر C نیاز دارد.

خلاصه روش اجرا

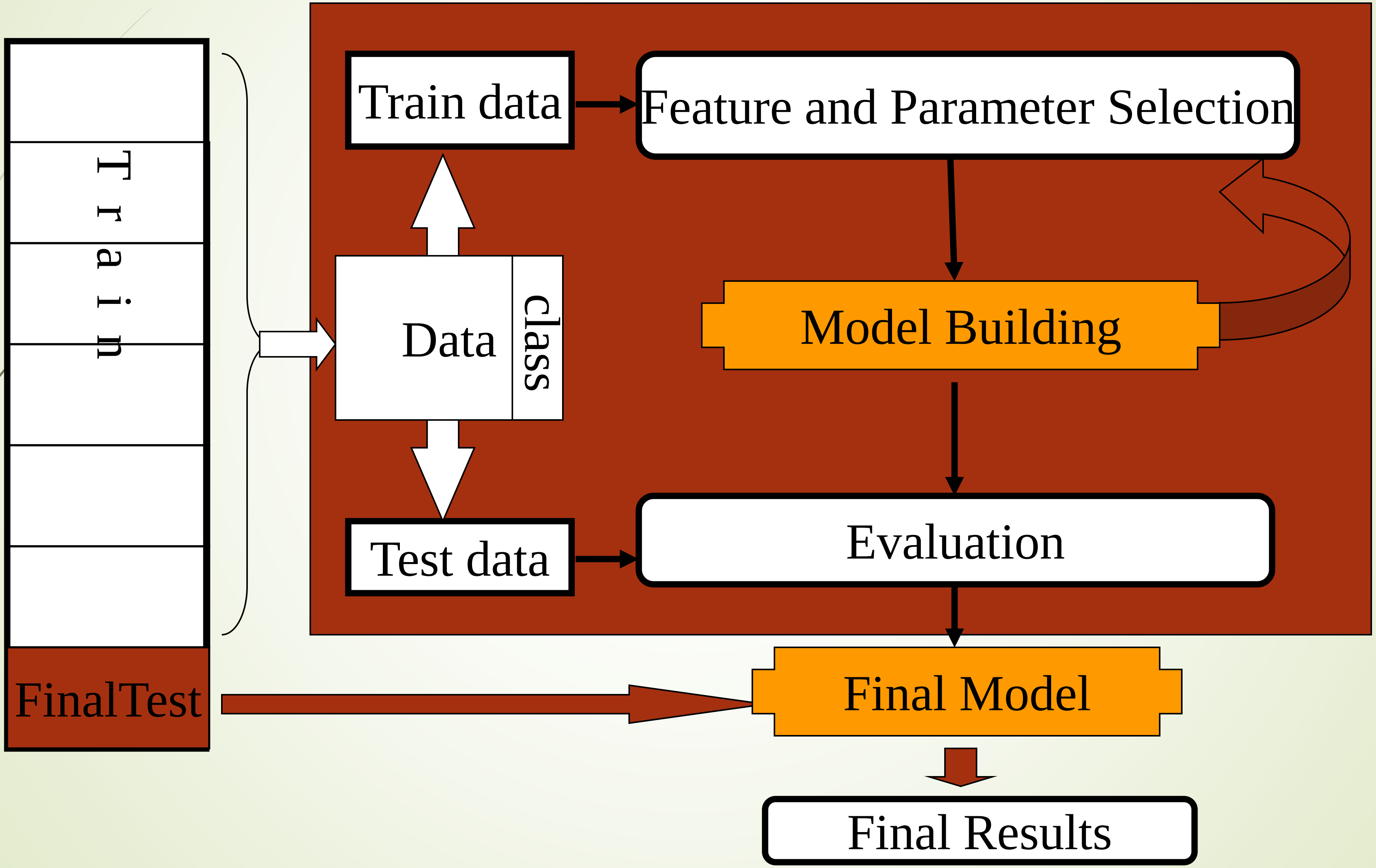
I. ماتریس الگو را آماده می کنیم.

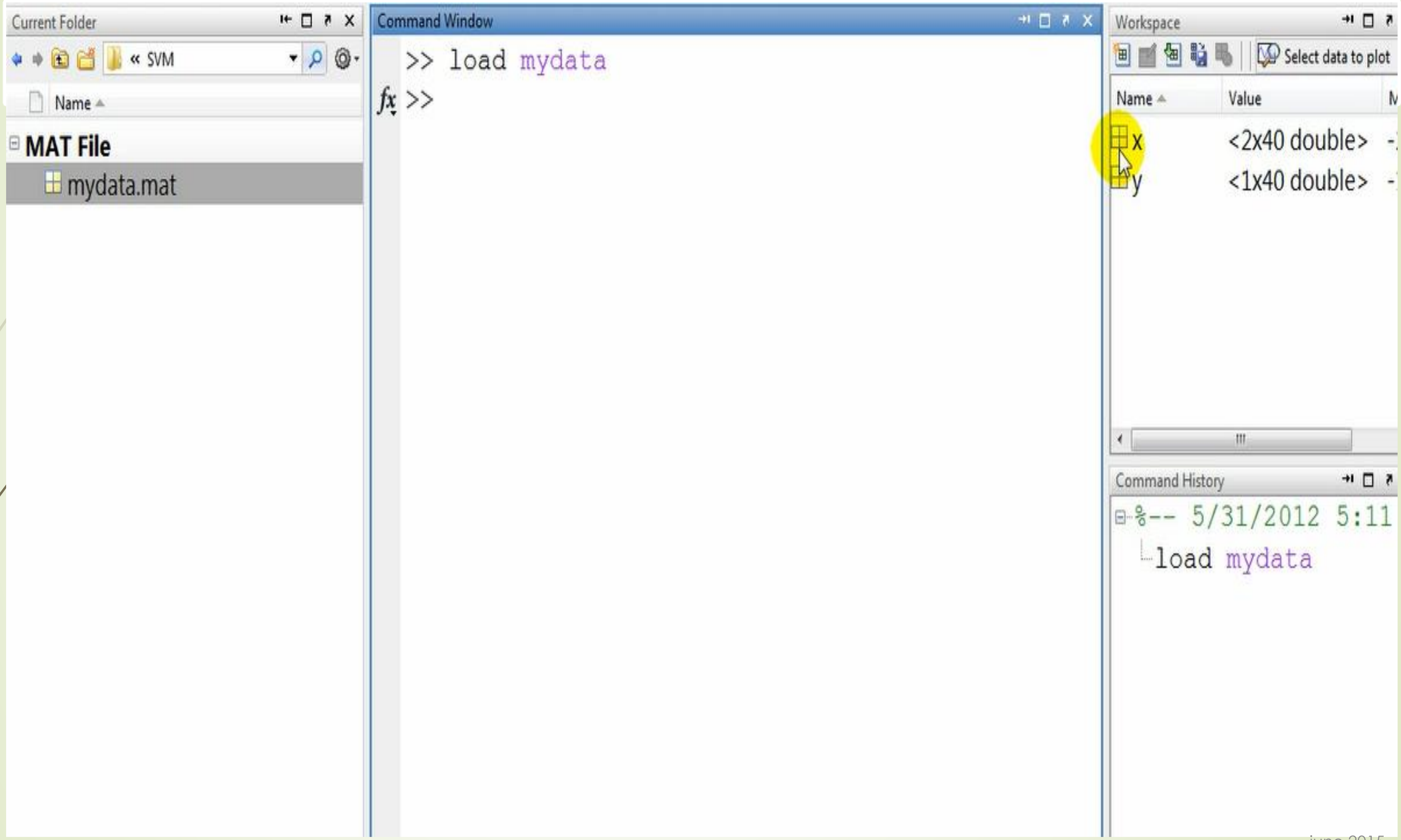
II. تابع کرنلی را برای استفاده انتخاب می کنیم.

III. پارامتر تابع کرنل و مقدار C را انتخاب می کنیم.

IV. برای محاسبه ی مقادیر α_i الگوریتم آموزشی را با استفاده از حل کننده های QP اجرا می کنیم.

V. داده های جدید با استفاده از مقادیر α_i و بردارهای پشتیبان می توانند دسته بندی شوند


Train
FinalTest
Train data
Feature and Parameter Selection
Data
class
Model Building
Test data
Evaluation
Final Model
Final Results


Current Folder
« SVM
Name
MAT File
mydata.mat
Command Window
>> load mydata
fx >>
Workspace
Select data to plot
Name
Value
x
<2x40 double>
y
<1x40 double>
Command History
%-- 5/31/2012 5:11
load mydata

Command Window

```
>> [x' y']

ans =

   -2.2588    0.7172    1.0000
    0.8622    1.6302    1.0000
    0.3188    0.4889    1.0000
   -1.3077    1.0347    1.0000
   -0.4336    0.7269    1.0000
    0.3426   -0.3034    1.0000
    3.5784    0.2939    1.0000
    2.7694   -0.7873    1.0000
   -1.3499    0.8884    1.0000
    3.0349   -1.1471    1.0000
    0.7254   -1.0689    1.0000
   -0.0631   -0.8095    1.0000
    0.7147   -2.9443    1.0000
   -0.2050    1.4384    1.0000
   -0.1241    0.3252    1.0000
    4.0214    2.0391   -1.0000
    6.6564    3.8023   -1.0000
    2.0569    2.7922   -1.0000
    3.2649    6.9080   -1.0000
    2.8833    4.8252   -1.0000
    0.2663    5.3790   -1.0000
    2.3792    2.9418   -1.0000
    0.4619    3.5314   -1.0000
    4.1885    3.7275   -1.0000
    1.7441    5.0984   -1.0000
    3.1416    3.7221   -1.0000
    2.2299    4.7015   -1.0000
    3.4292    1.9482   -1.0000
    2.1510    3.6462   -1.0000
    3.6929    3.1764   -1.0000
```

```
clc;
clear;
close all;

%% Load Data

load mydata;

n=numel(y);

ClassA=find(y==1)
ClassB=find(y==-1);

figure;
plot(x(1,ClassA),x(2,ClassA),'ro');
hold on;
plot(x(1,ClassB),x(2,ClassB),'bs');
legend('Class A','Class B');


```

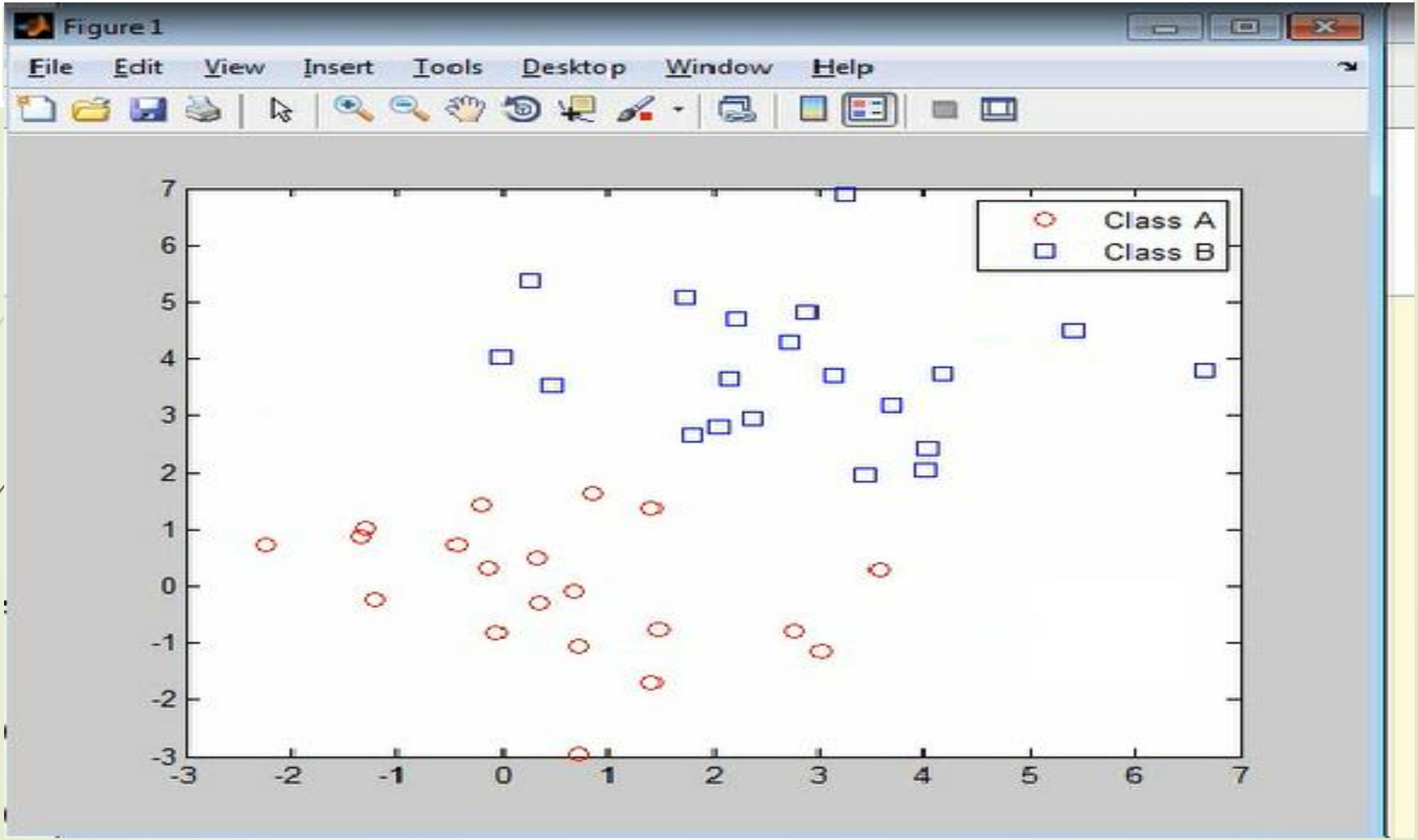
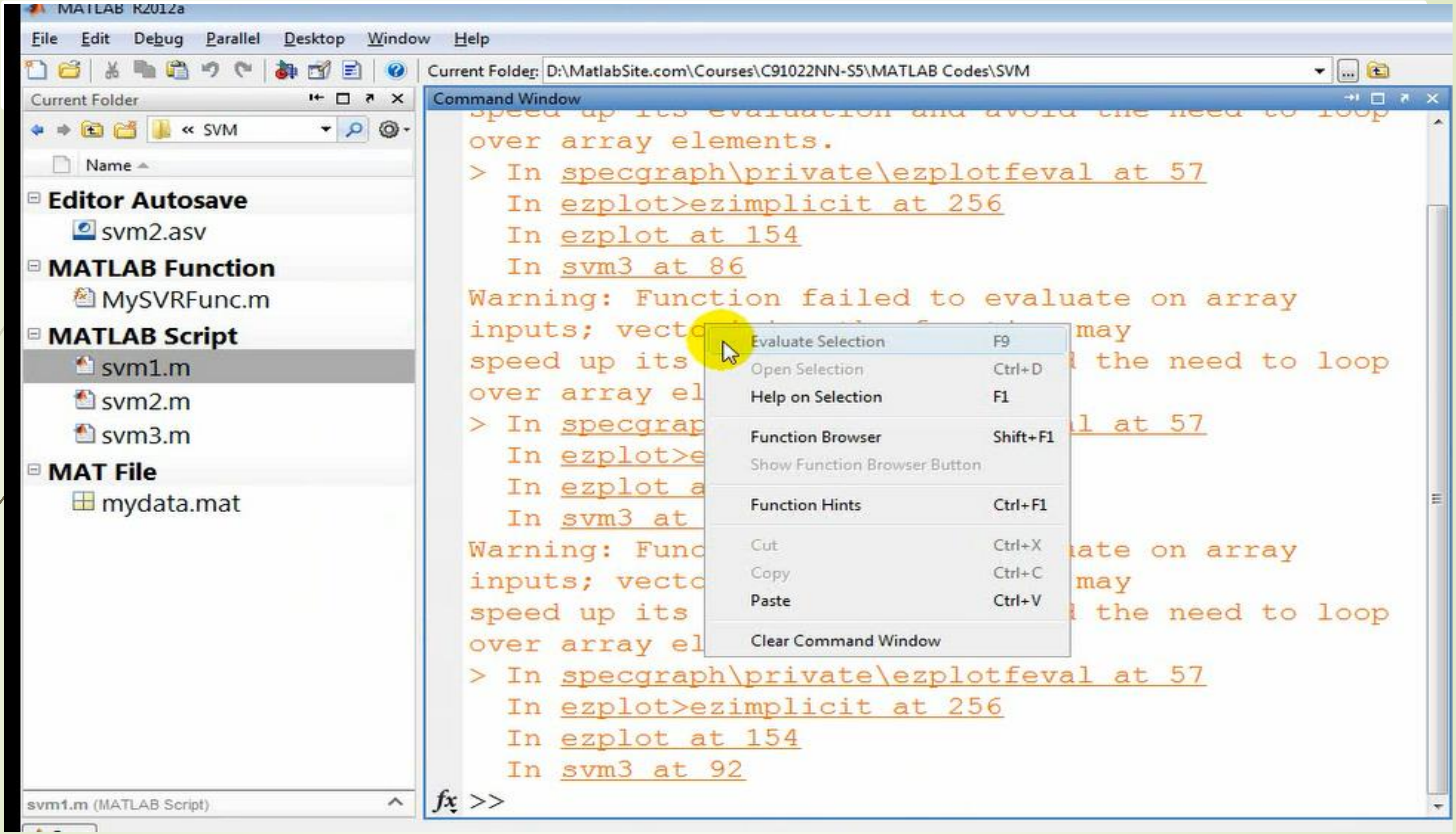
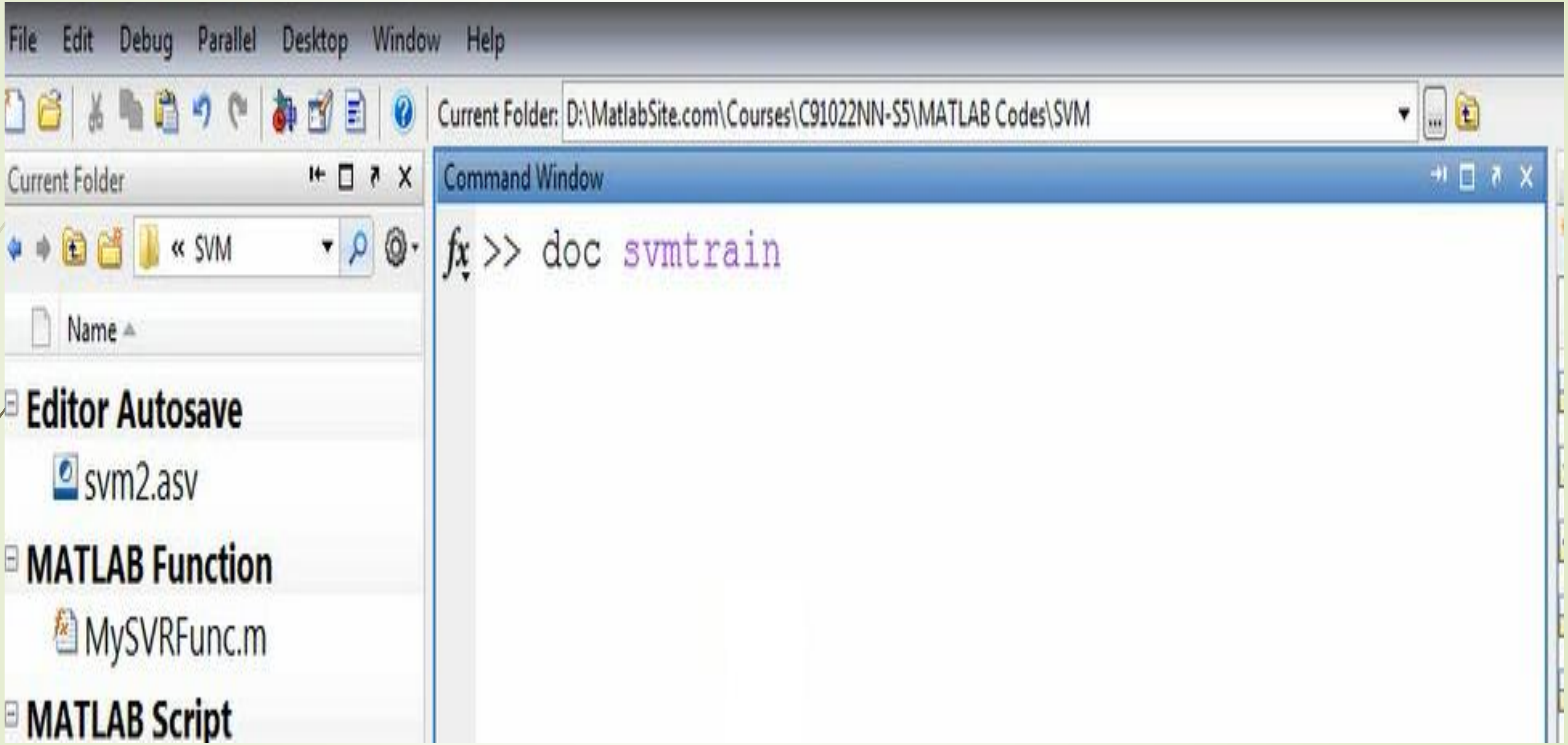
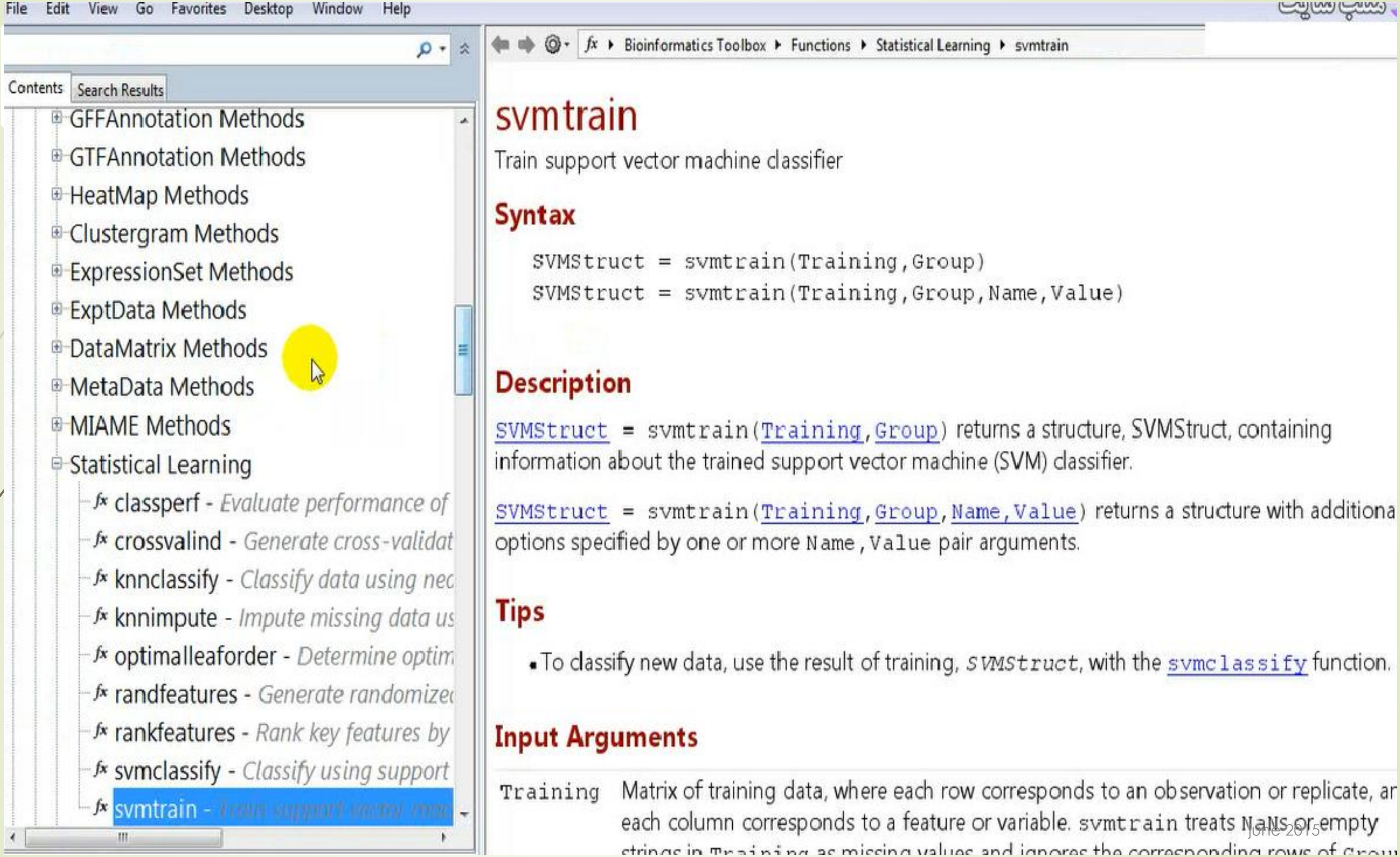


Figure 1
File Edit View Insert Tools Desktop Window Help
Class A
Class B
7
6
5
4
3
2
1
0
-1
-2
-3
-3 -2 -1 0 1 2 3 4 5 6 7


MATLAB R2012a
File Edit Debug Parallel Desktop Window Help
Current Folder: D:\MatlabSite.com\Courses\C91022NN-S5\MATLAB Codes\SVM
Current Folder
SVM
Name
Editor Autosave
svm2.asv
MATLAB Function
MySVRFunc.m
MATLAB Script
svm1.m
svm2.m
svm3.m
MAT File
mydata.mat
svm1.m (MATLAB Script)
Command Window
over array elements.
> In specgraph\private\ezplotfeval at 57
In ezplot>ezimplicit at 256
In ezplot at 154
In svm3 at 86
Warning: Function failed to evaluate on array
inputs; vecto
may
speed up its
the need to loop
over array el
Evaluate Selection F9
Open Selection Ctrl+D
Help on Selection F1
Function Browser Shift+F1
Show Function Browser Button
Function Hints Ctrl+F1
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Clear Command Window
> In specgraph\private\ezplotfeval at 57
In ezplot>ezimplicit at 256
In ezplot at 154
In svm3 at 92
fx >>

چطور می توان از توابع موجود در تولباکس bio informatic برای پیاده سازی svm استفاده کرد؟




File Edit View Go Favorites Desktop Window Help
Contents
Search Results
GFFAnnotation Methods
GTFAnnotation Methods
HeatMap Methods
Clustergram Methods
ExpressionSet Methods
ExptData Methods
DataMatrix Methods
MetaData Methods
MIAME Methods
Statistical Learning
classperf - Evaluate performance of
crossvalind - Generate cross-validat
knnclassify - Classify data using nea
knnimpute - Impute missing data us
optimalleaforder - Determine optim
randfeatures - Generate randomize
rankfeatures - Rank key features by
svmclassify - Classify using support
svmtrain -
Bioinformatics Toolbox ▸ Functions ▸ Statistical Learning ▸ svmtrain
svmtrain
Train support vector machine classifier
Syntax
SVMStruct = svmtrain(Training,Group)
SVMStruct = svmtrain(Training,Group,Name,Value)
Description
SVMStruct = svmtrain(Training,Group) returns a structure, SVMStruct, containing information about the trained support vector machine (SVM) classifier.
SVMStruct = svmtrain(Training,Group,Name,Value) returns a structure with additiona
options specified by one or more Name,Value pair arguments.
Tips
• To classify new data, use the result of training, SVMStruct, with the svmclassify function.
Input Arguments
Training
Matrix of training data, where each row corresponds to an observation or replicate, ar
each column corresponds to a feature or variable. svmtrain treats NaNs or empty

svm

MATLAB R2012a

File Edit Debug Parallel Desktop Window Help

Current Folder: D:\MatlabSite.com\Courses\C91022NN-S5\MATLAB Codes\SVM

Command Window

```
>> TrainInputs=x';
>> TrainTargets=y';
>> svmstruct=svmtrain(TrainInputs,TrainTargets)

svmstruct = 

          SupportVectors: [44x2 double]
                   Alpha: [44x1 double]
                    Bias: -0.3244
          KernelFunction: @linear_kernel
      KernelFunctionArgs: {}
              GroupNames: [50x1 double]
    SupportVectorIndices: [44x1 double]
               ScaleData: [1x1 struct]
           FigureHandles: []

fx >>
```

svm3.m | MySVRFunc.m | svm4.m

```
ClassA=find(y==1);
ClassB=find(y==-1);

TrainInputs=x';
TrainTargets=y';

n=numel(TrainTargets);

%% Design SVM

C=10;

svmstruct=svmtrain(TrainInputs,TrainTargets,...
    'boxconstraint',C,...
    'kernel_function','rbf',...
    'rbf_sigma',0.5,...
    'showplot',true);
```

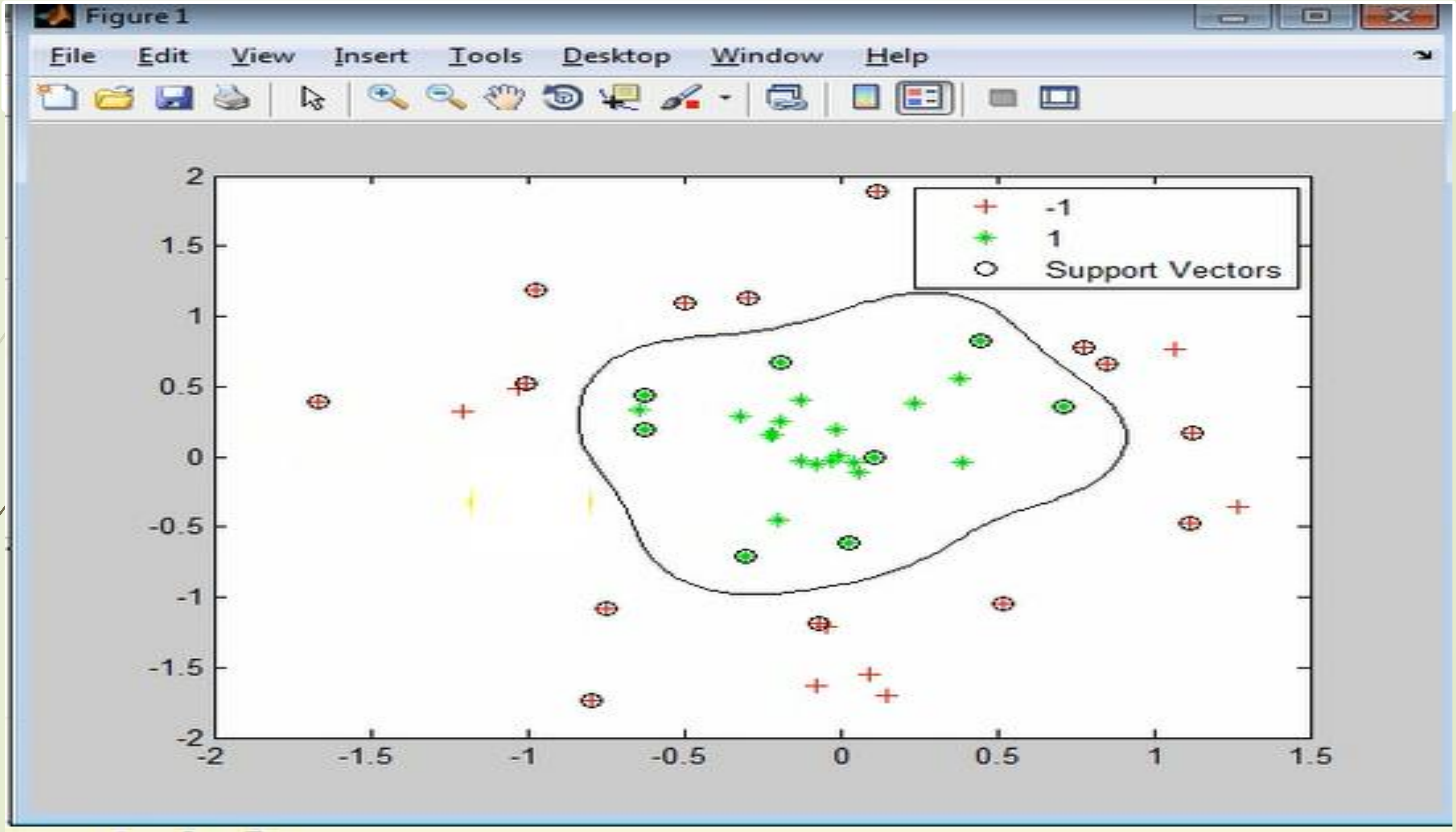


Figure 1
File Edit View Insert Tools Desktop Window Help
-1
1
Support Vectors
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2
-1.5
-1
-0.5
0
0.5
1
1.5

```matlab
TrainInputs=x';
TrainTargets=y';

n=numel(TrainTargets);

%% Design SVM

C=10;

svmstruct=svmtrain(TrainInputs,TrainTargets,...
    'boxconstraint',C,...
    'kernel_function','linear',...
    'rbf_sigma',0.5,...
    'showplot',true);
```

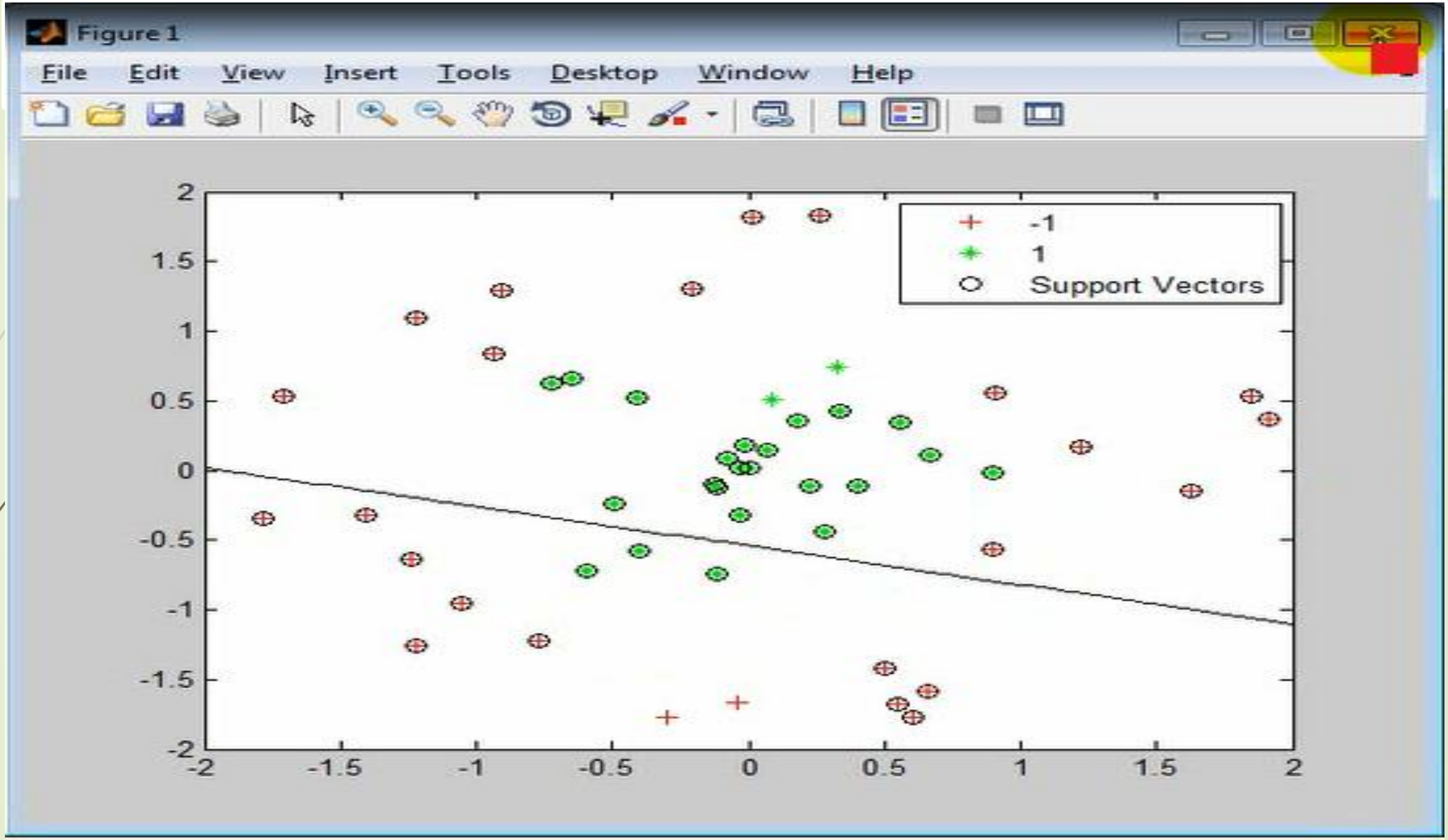
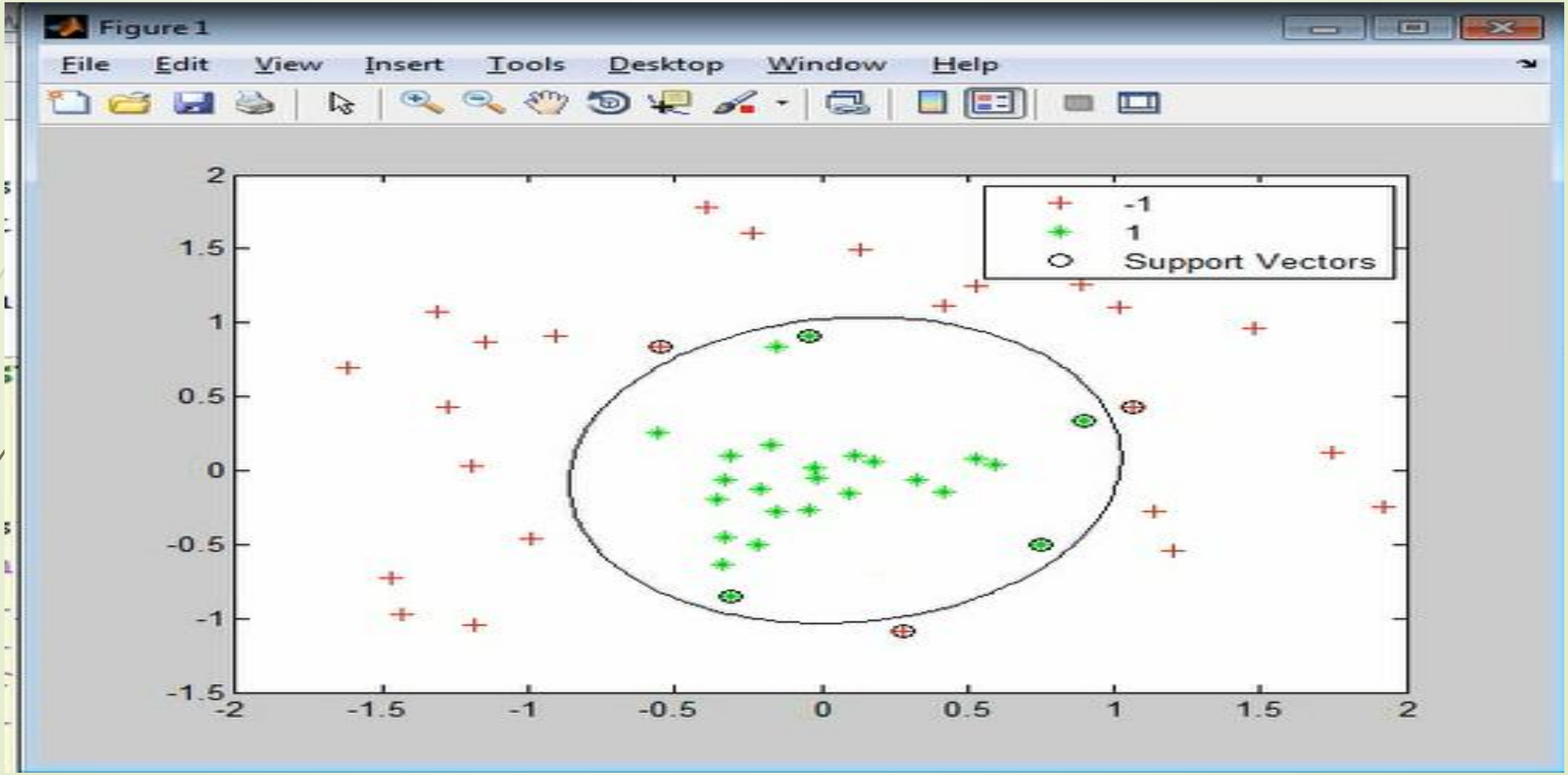
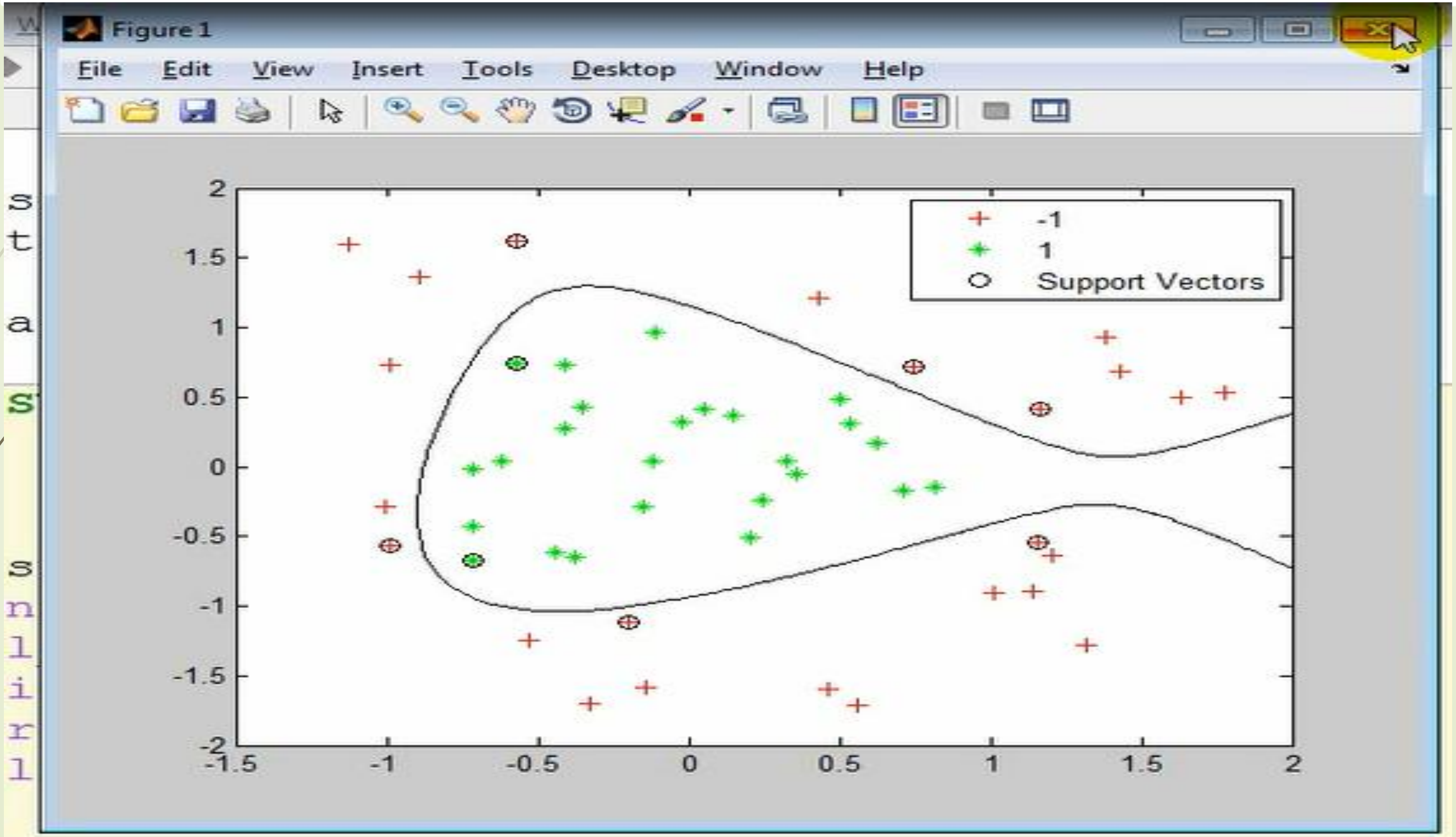


Figure 1
File Edit View Insert Tools Desktop Window Help
-1
1
Support Vectors
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2

کرنل: چند جمله ای درجه ۲



کرنل : چند جمله ای درجه ۳



کرنل MLP

svm1.m
svm2.m
svm3.m
MySVRFunc.m
svm4.m

```matlab
TrainInputs=x';
TrainTargets=y';

n=numel(TrainTargets);

%% Design SVM

C=10;

svmstruct=svmtrain(TrainInputs,TrainTargets,...
    'boxconstraint',C,...
    'kernel_function','mlp',...
    'rbf_sigma',0.5,...
    'polyorder',2,...
    'mlp_params',[1 -2],...
    'showplot',true);
```

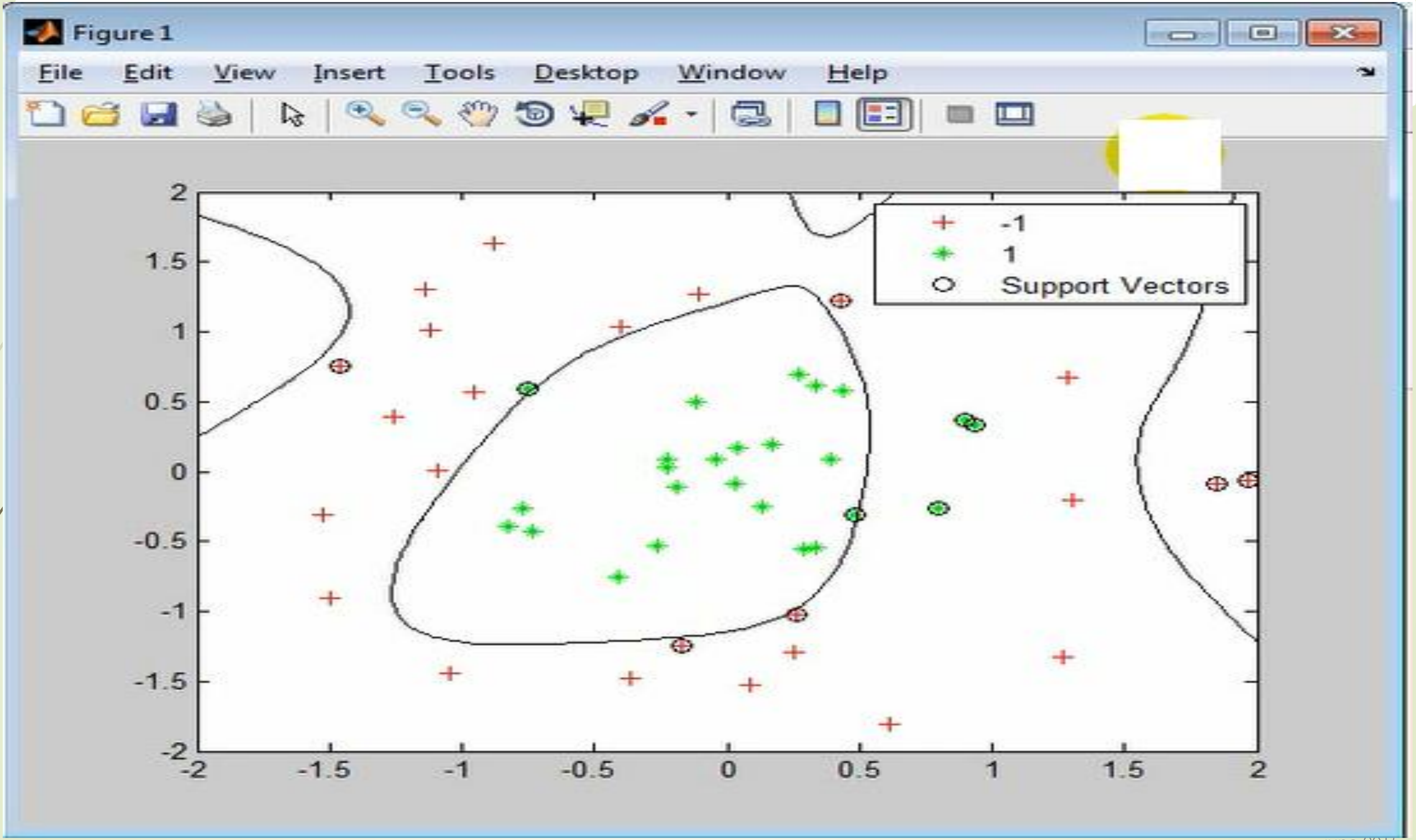


Figure 1
File Edit View Insert Tools Desktop Window Help
-1
1
Support Vectors
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2 -1.5 -1 -0.5 0 0.5 1 1.5 2

پیش بینی عود مجدد سرطان سینه به کمک سه تکنیک داده کاوي

- تحقیقی که توسط دکتر عباس طلوعی و همکارانش انجام شده، در خصوص بیماران مبتلا به سرطان سینه که حداقل هرکدام به مدت دو سال تحت پیگیري بوده اند می باشد.
- اطلاعات این بیماران در مرکز تحقیقات سرطان جهاد دانشگاهی بدست امده است.
- به منظور توسعه مدلهاي پیش بینی جهت پیش بینی عود سرطان سینه، از درختان

تصمیم گیري (C.5) ، ماشین بردار پشتیبان (SVM: Support Vector Machines) و

تکنیکهاي شبکه هاي عصبی مصنوعی (ANNs: Artificial Neural Networks) با بهره

گیري از پایگاه داده مذکور استفاده شده است.

مواد و روش اجرا

- پایگاه داده قبل از پردازش شامل ۱۱۸۹ نمونه و ۲۶ ویژگی اولیه بود که پس از پردازش اولیه، تمیزکردن داده ها و حذف نمونه هاي محتوي متغيرها بامقاديرمفقود
- شده، ۵۴۷ نمونه با اطلاعات و متغيرهاي كامل و ۲۲ویژگی نهایی باقی ماند. که به دو گروه منتهی به عود یا عدم عود ختم گردیدند.

از اين تعداد ۱۱۷ بیمار دچار عودبيماري و ۴۳۰ بیمار نيز فاقد عود مجدد بيماري بودند.

- جدول ۱ شامل متغیرهای پیش بینی برای مدل سازی عود مجدد سرطان سینه می باشد

۱	عود مجدد (Local Recurrence)	۰- عود بیماری ۱- عدم عود
۲	سن تشخیص بیماری (Age Diagnosis)	0 = "<35" 1 = "35-44" 2 = "45-55" 3 = ">55"
۳	سن قاعدگی (Age Menarche)	۰- شروع قاعدگی بیشتر از ۱۲ سال ۱- شروع قاعدگی ۱۲ ≥ (خطرساز)
۴	سن یائسگی (Age Menopause)	۰- سن شروع یائسگی کمتر از ۵۰ سال ۱- سن شروع یائسگی بیشتر از یا مساوی ۵۰ سال (خطرساز) ۲- فرد هنوز یائسه نشده
۵	سابقه نازایی (Infertility)	۰- بیمار سابقه نازایی ندارد (No) ۱- بیمار سابقه نازایی دارد (Yes)
۶	سابقه فامیلی ابتلا به سرطان پستان (Family History)	۰- سابقه ای وجود ندارد (No) ۱- سابقه بیماری در خانواده وجود دارد (Yes)
۷	سابقه فامیلی ابتلا به سایر انواع سرطان (Other Cancer)	0 = "no" 1 = "mail breast cancer" 2 = "prostate cancer" 3 = "colon cancer" 4 = "ovarian cancer" 5 = "utrus cancer"
۸	محل قرارگیری توده سرطانی (Location)	1 = "uoq" 2 = "uiq" 3 = "loq" 4 = "liq" 5 = "central(nipple areole)" 6 = "axilla" 12 = "upper half" 13 = "latral half" 14 = "uoq and liq" 24 = "medial half" 30 = "three quadrant" 34 = "lower half" 50 = "diffuse"

- داده ها با نسبت ۷۰ به ۳۰ به ترتیب براي داده هاي آموزشی و آزمایشی پارتیشن بندي شدند. دقت مدل، درصد تعداد دفعاتی است که نمونه هاي آزمایشی با موفقیت دسته بندي می شوند. اگر دقت مدل قابل قبول باشد می توان مدل را براي دسته بندي داده هایی که دسته آنها مشخص نیستند، به کار برد. براي انجام آموزش و آزمایش داده ها به صورت تصادفی به ۱۰ قسمت و در۱۰ آزمایش مختلف تقسیم می شوند.سپس این داده ها به دو کلاس عود مجدد و عدم عودمجدد تقسیم می شوند.

جدول شماره دو، میزان دقت، حساسیت و ویژگی رابرای سه روش مختلف دسته بندي نشان می دهد.

جدول ۲: مقایسه نتایج حاصل از سه مدل داده کاوی

	دقت	حساسیت	ویژگی
درخت تصمیم (C5.0)	۰/۹۳۶	۰/۹۵۸	۰/۹۰۷
شبکه های عصبی مصنوعی (MLP)	۰/۹۴۷	۰/۹۵۶	۰/۹۲۸
ماشین بردار پشتیبان (SVM)	۰/۹۵۷	۰/۹۷۱	۰/۹۴۵

Resources

- http://www.kernel-machines.org/
- http://www.support-vector.net/
- http://www.support-vector.net/icml-tutorial.pdf
- http://www.kernel-machines.org/papers/tutorial-nips.ps.gz
- http://www.clopinet.com/isabelle/Projects/SVM/applist.html

Resources

- [1] b.E. Boser et al. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on computational learning theory 5 144-152, Pittsburgh, 1992.
- [2] l. Bottou et al. Comparison of classifier methods: a case study in handwritten digit recognition. Proceedings of the 12th IAPR international conference on pattern recognition, vol. 2, pp. 77-82.
- [3] v. Vapnik. The nature of statistical learning theory. 2nd edition, Springer, 1999.

Thank you!