شبکه‌های عصبی مصنوعی

جلسه سیزدهم:
ماشین بردار پشتیبان (۱)
(Support Vector Machine = SVM)

محمد فرخی (farrokhi AT iust.ac.ir)

ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان (SVM)

– دو دسته شبکه پیشخورد تا کنون بررسی کردیم:

ماشین بردار پشتیبان (SVM)

- دو دسته شبکه پیشخورد تا کنون بررسی کردیم:

۱- MLP با آموزش پس‌انتشار خطا: از ویژگی‌های این الگوریتم، سادگی آن است ولی الگوریتم به‌آهستگی همگرا می‌شود و بهینه‌بودن آن زیر سوال است.

ماشین بردار پشتیبان (SVM)

– دو دسته شبکه پیشخورد تا کنون بررسی کردیم:

۱– MLP با آموزش پس‌انتشار خطا: از ویژگی‌های این الگوریتم، سادگی آن است ولی الگوریتم به‌آهستگی همگرا می‌شود و بهینه‌بودن آن زیر سوال است.

۲– RBF با آموزش دو مرحله‌ای (K–میانگین و RLS) که زیربهینه است.

ماشین بردار پشتیبان (SVM)

- دو دسته شبکه پیشخورد تا کنون بررسی کردیم:

۱- MLP با آموزش پس‌انتشار خطا: از ویژگی‌های این الگوریتم، سادگی آن است ولی الگوریتم به‌آهستگی همگرا می‌شود و بهینه‌بودن آن زیر سوال است.

۲- RBF با آموزش دو مرحله‌ای (K–میانگین و RLS) که زیربهینه است.

- در این جا، نوع دیگری از شبکه‌های پیشخورد را بررسی می‌کنیم به نام «ماشین بردار پشتیبان» (Support Vector Machine = SVM) که توسط وپنیک (Vapnik) و چرونکیس (Chervonenkis) در سال ۱۹۶۳ ابداع شد.

ماشین بردار پشتیبان (SVM)

– دو دسته شبکه پیشخورد تا کنون بررسی کردیم:

۱– MLP با آموزش پس‌انتشار خطا: از ویژگی‌های این الگوریتم، سادگی آن است ولی الگوریتم به‌آهستگی همگرا می‌شود و بهینه‌بودن آن زیر سوال است.

۲– RBF با آموزش دو مرحله‌ای (K–میانگین و RLS) که زیربهینه است.

– در این جا، نوع دیگری از شبکه‌های پیشخورد را بررسی می‌کنیم به نام «ماشین بردار پشتیبان» (Support Vector Machine = SVM) که توسط وپنیک (Vapnik) و چروننکیس (Chervonenkis) در سال ۱۹۶۳ ابداع شد.

ماشین بردار پشتیبان (SVM)

- دو دسته شبکه پیشخورد تا کنون بررسی کردیم:

۱- MLP با آموزش پس‌انتشار خطا: از ویژگی‌های این الگوریتم، سادگی آن است ولی الگوریتم به‌آهستگی همگرا می‌شود و بهینه‌بودن آن زیر سوال است.

۲- RBF با آموزش دو مرحله‌ای (K–میانگین و RLS) که زیربهینه است.

- در این جا، نوع دیگری از شبکه‌های پیشخورد را بررسی می‌کنیم به نام «ماشین بردار پشتیبان» (Support Vector Machine = SVM) که توسط وپنیک (Vapnik) و چرونکیس (Chervonenkis) در سال ۱۹۶۳ ابداع شد.

- SVM شبکه‌ای است با آموزش باینری و ویژگی‌های جالب.

ماشین بردار پشتیبان (SVM)

– دو دسته شبکه پیشخورد تا کنون بررسی کردیم:

۱– MLP با آموزش پس‌انتشار خطا: از ویژگی‌های این الگوریتم، سادگی آن است ولی الگوریتم به‌آهستگی همگرا می‌شود و بهینه‌بودن آن زیر سوال است.

۲– RBF با آموزش دو مرحله‌ای (K–میانگین و RLS) که زیربهینه است.

– در این جا، نوع دیگری از شبکه‌های پیشخورد را بررسی می‌کنیم به نام «ماشین بردار پشتیبان» (Support Vector Machine = SVM) که توسط وپنیک (Vapnik) و چرونکیس (Chervonenkis) در سال ۱۹۶۳ ابداع شد.

– SVM شبکه‌ای است با آموزش باینری و ویژگی‌های جالب.

– اصول عملکرد این شبکه به‌مفهوم کلاسه‌بندی کردن الگوها به‌قرار زیر است:

ماشین بردار پشتیبان (SVM)

– دو دسته شبکه پیشخورد تا کنون بررسی کردیم:

۱– MLP با آموزش پس‌انتشار خطا: از ویژگی‌های این الگوریتم، سادگی آن است ولی الگوریتم به‌آهستگی همگرا می‌شود و بهینه‌بودن آن زیر سوال است.

۲– RBF با آموزش دو مرحله‌ای (K–میانگین و RLS) که زیربهینه است.

– در این جا، نوع دیگری از شبکه‌های پیشخورد را بررسی می‌کنیم به نام «ماشین بردار پشتیبان» (Support Vector Machine = SVM) که توسط وپنیک (Vapnik) و چرونکیس (Chervonenkis) در سال ۱۹۶۳ ابداع شد.

– SVM شبکه‌ای است با آموزش باینری و ویژگی‌های جالب.

– اصول عملکرد این شبکه به‌مفهوم کلاسه‌بندی کردن الگوها به‌قرار زیر است:

با دردست داشتن نمونه‌های آموزش، ماشین بردار پشتیبان ابرصفحه‌ای (hyperplane) به‌عنوان سطح تصمیم‌گیری تشکیل می‌دهد به‌طوری که حاشیه جداسازی (margin of separation) بین نمونه‌های مثبت و منفی بیشینه شود.

ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان (SVM)

– نکته مهم در مساله آموزش بردار پشتیبان، در «کرنل ضرب داخلی» بین بردار پشتیبان و بردار x از داده‌های ورودی است.

ماشین بردار پشتیبان (SVM)

– نکته مهم در مساله آموزش بردار پشتیبان، در «کرنل ضرب داخلی» بین بردار پشتیبان و بردار x از داده‌های ورودی است.

– مهم‌تر از آن این‌که بردارهای پشتیبان شامل زیرمجموعه‌ای از نقاط داده است که توسط الگوریتم آموزش استخراج می‌شود.

ماشین بردار پشتیبان (SVM)

- نکته مهم در مساله آموزش بردار پشتیبان، در «کرنل ضرب داخلی» بین بردار پشتیبان و بردار x از داده‌های ورودی است.

- مهم‌تر از آن این‌که بردارهای پشتیبان شامل زیرمجموعه‌ای از نقاط داده است که توسط الگوریتم آموزش استخراج می‌شود.

- در واقع، نام کرنل (هسته) نیز از همین جا گرفته شده است.

ماشین بردار پشتیبان (SVM)

- نکته مهم در مساله آموزش بردار پشتیبان، در «کرنل ضرب داخلی» بین بردار پشتیبان و بردار x از داده‌های ورودی است.

- مهم‌تر از آن این‌که بردارهای پشتیبان شامل زیرمجموعه‌ای از نقاط داده است که توسط الگوریتم آموزش استخراج می‌شود.

- در واقع، نام کرنل (هسته) نیز از همین جا گرفته شده است.

- برخلاف روش کرنلِ زیربهینه در شبکه RBF، اصول طراحی SVM بهینه است.

ماشین بردار پشتیبان (SVM)

– نکته مهم در مساله آموزش بردار پشتیبان، در «کرنل ضرب داخلی» بین بردار پشتیبان و بردار $\mathbf{x}$ از داده‌های ورودی است.

– مهم‌تر از آن این‌که بردارهای پشتیبان شامل زیرمجموعه‌ای از نقاط داده است که توسط الگوریتم آموزش استخراج می‌شود.

– در واقع، نام کرنل (هسته) نیز از همین جا گرفته شده است.

– برخلاف روش کرنلِ زیربهینه در شبکه RBF، اصول طراحی SVM بهینه است.

– بهینه‌بودن آن ریشه در بهینه‌سازی محدب دارد.

ماشین بردار پشتیبان (SVM)

- نکته مهم در مساله آموزش بردار پشتیبان، در «کرنل ضرب داخلی» بین بردار پشتیبان و بردار x از داده‌های ورودی است.

- مهم‌تر از آن این‌که بردارهای پشتیبان شامل زیرمجموعه‌ای از نقاط داده است که توسط الگوریتم آموزش استخراج می‌شود.

- در واقع، نام کرنل (هسته) نیز از همین جا گرفته شده است.

- برخلاف روش کرنلِ زیربهینه در شبکه RBF، اصول طراحی SVM بهینه است.

- بهینه‌بودن آن ریشه در بهینه‌سازی محدب دارد.

- اگرچه این ویژگی مفید، به‌قیمت افزایش پیچیدگی محاسبات تمام می‌شود.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- داده‌های موردنظر $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- داده‌های موردنظر $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$

- فرض‌می‌شود که الگوهای زیرمجموعه $d_i = +1$ از الگوهای زیرمجموعه $d_i = -1$ جداپذیر خطی‌اند.

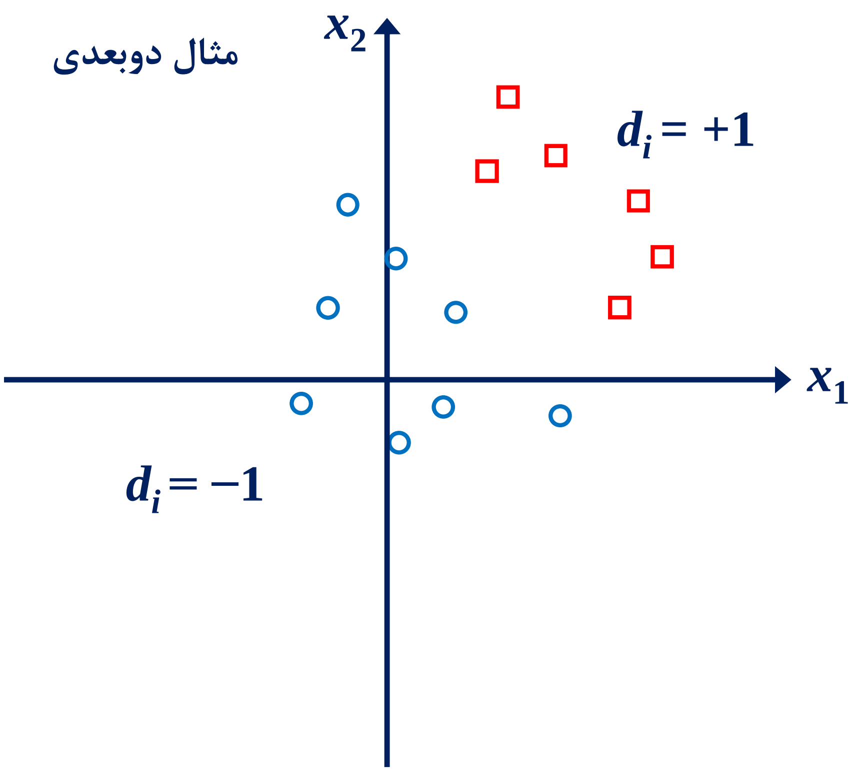
ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- داده‌های موردنظر $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$

- فرض‌می‌شود که الگوهای زیرمجموعه $d_i = +1$ از الگوهای زیرمجموعه $d_i = -1$ جداپذیر خطی‌اند.

مثال دوبعدی



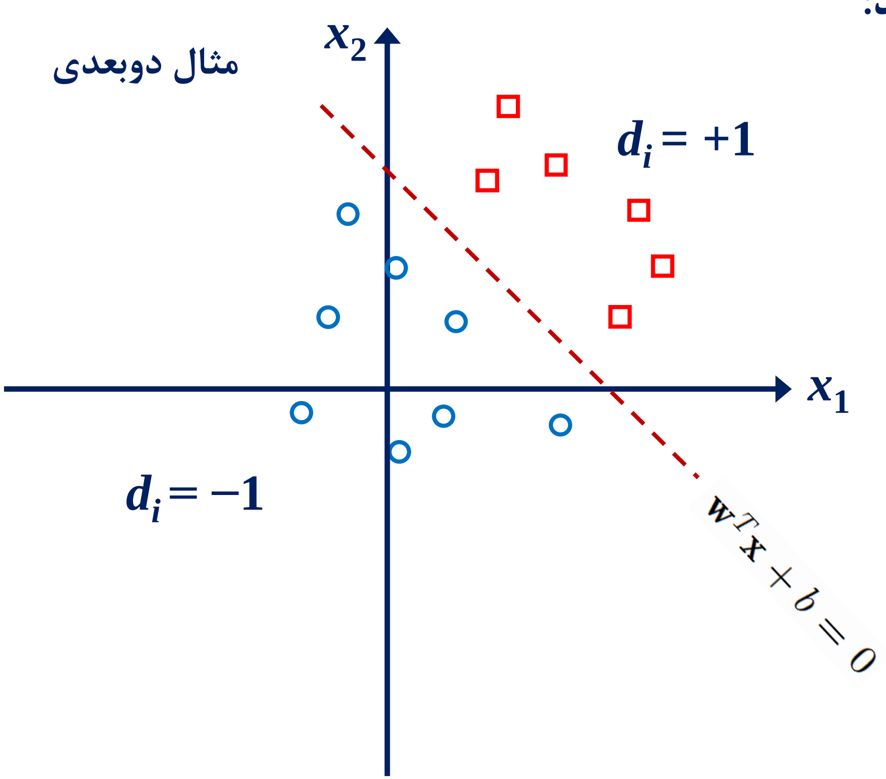
ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– داده‌های موردنظر $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$

– فرض‌می‌شود که الگوهای زیرمجموعه $d_i = +1$ از الگوهای زیرمجموعه $d_i = -1$ جداپذیر خطی‌اند.

– سطح تصمیم‌گیری (ابرصفحه) زیر را درنظر بگیرید:



ماشین بردار پشتیبان (SVM)

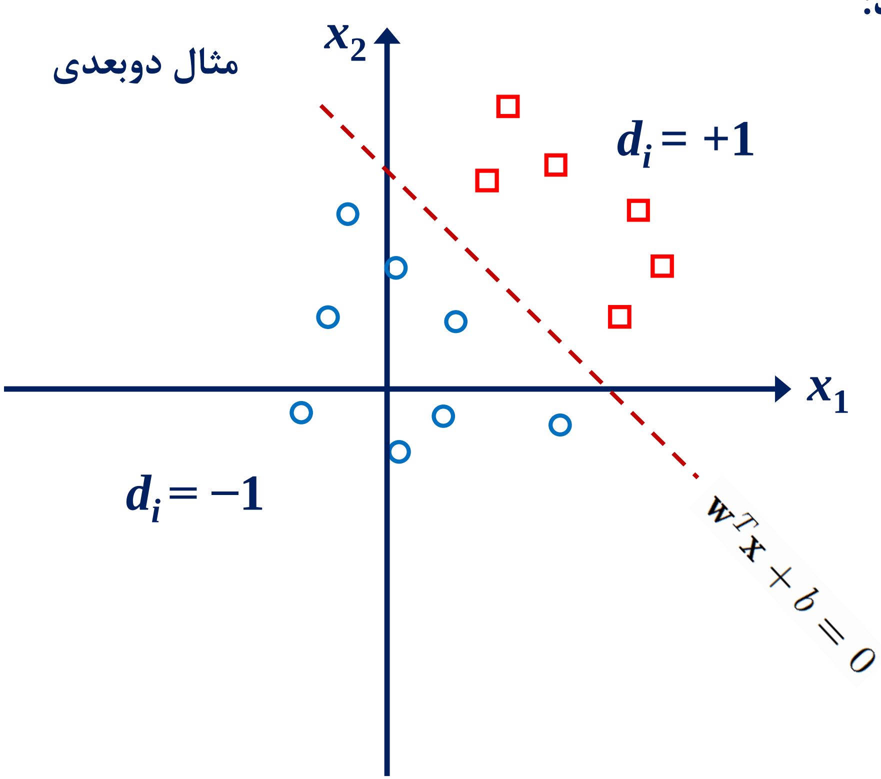
۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- داده‌های موردنظر $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$

- فرض‌می‌شود که الگوهای زیرمجموعه $d_i = +1$ از الگوهای زیرمجموعه $d_i = -1$ جداپذیر خطی‌اند.

- سطح تصمیم‌گیری (ابرصفحه) زیر را درنظر بگیرید:

مثال دوبعدی



بنابراین

$$\mathbf{w}^T\mathbf{x} + b \geq 0 \quad \text{for} \quad d_i = +1$$

$$\mathbf{w}^T\mathbf{x} + b < 0 \quad \text{for} \quad d_i = -1$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– داده‌های موردنظر $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$

– فرض‌می‌شود که الگوهای زیرمجموعه $d_i = +1$ از الگوهای زیرمجموعه $d_i = -1$ جداپذیر خطی‌اند.

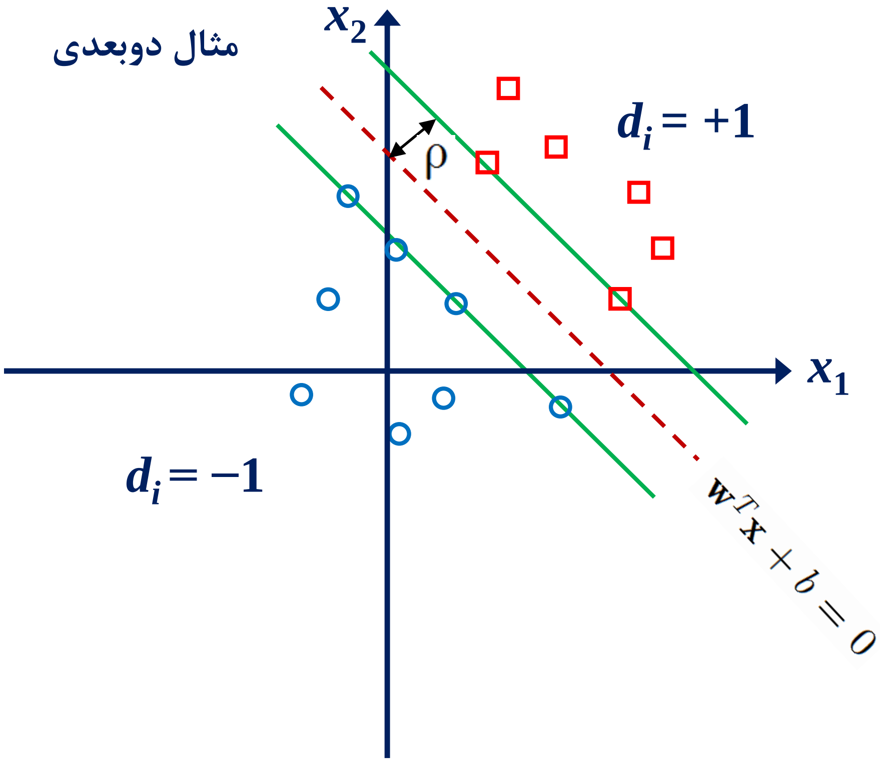
– سطح تصمیم‌گیری (ابرصفحه) زیر را درنظر بگیرید:

بنابراین

$$\mathbf{w}^T\mathbf{x} + b \geq 0 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x} + b < 0 \quad \text{for} \quad d_i = -1$$

– برای $\mathbf{w}$ و b معلوم، جداسازی بین این ابرصفحه و نزدیک‌ترین داده را «حاشیه جداسازی» نامیده و با ρ نشان می‌دهیم.

مثال دوبعدی



ماشین بردار پشتیبان (SVM)

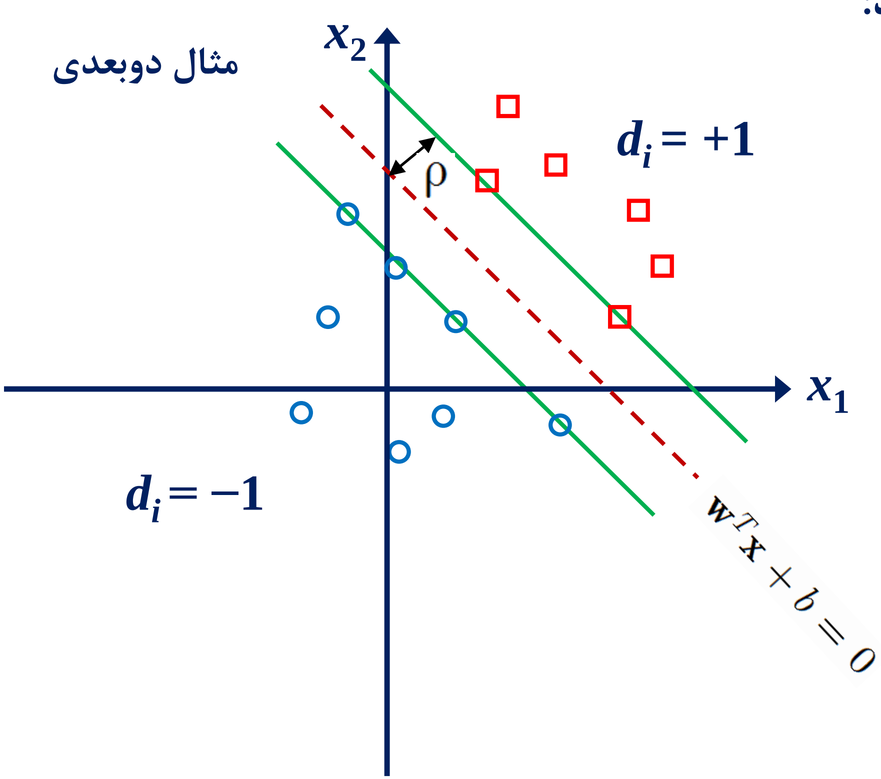
۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– داده‌های موردنظر $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$

– فرض‌می‌شود که الگوهای زیرمجموعه $d_i = +1$ از الگوهای زیرمجموعه $d_i = -1$ جداپذیر خطی‌اند.

– سطح تصمیم‌گیری (ابرصفحه) زیر را درنظر بگیرید:

مثال دوبعدی



بنابراین

$$\mathbf{w}^T\mathbf{x} + b \geq 0 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x} + b < 0 \quad \text{for} \quad d_i = -1$$

– برای $\mathbf{w}$ و b معلوم، جداسازی بین این ابرصفحه و نزدیک‌ترین داده را «حاشیه جداسازی» نامیده و با ρ نشان می‌دهیم.

– هدف SVM یافتن ابرصفحه‌ای است که حاشیه جداسازی (ρ) را بیشینه کند.

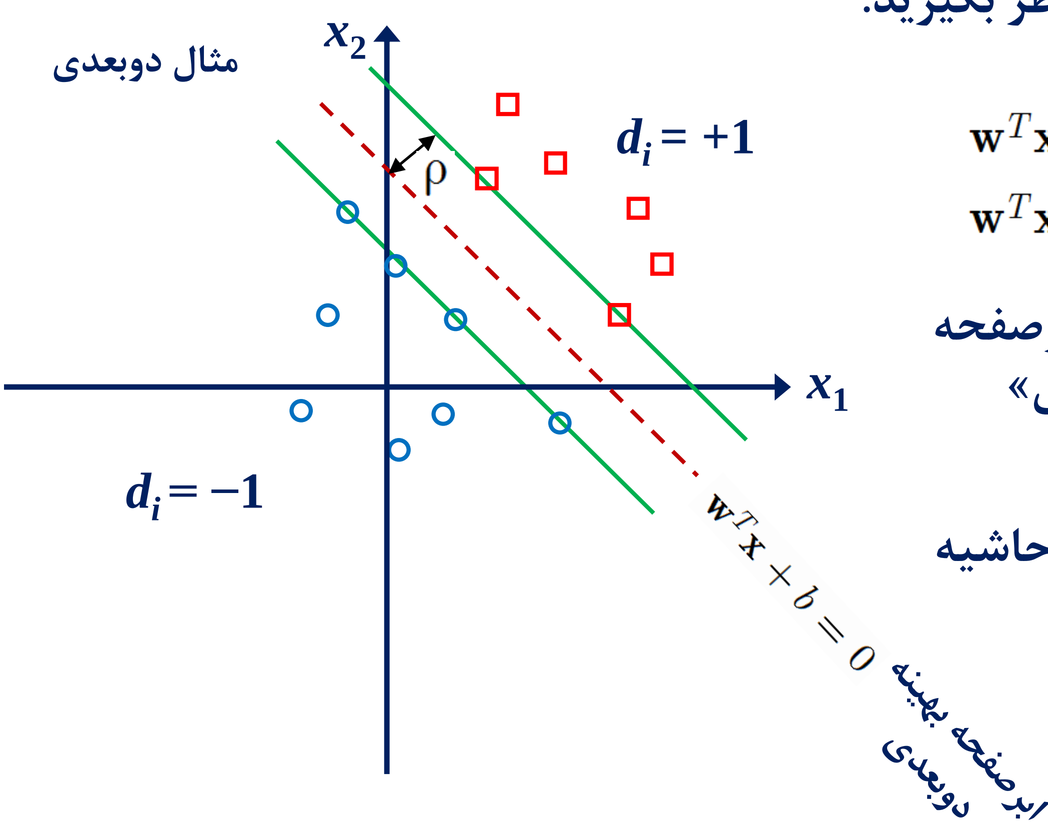
ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– داده‌های موردنظر $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$

– فرض‌می‌شود که الگوهای زیرمجموعه $d_i = +1$ از الگوهای زیرمجموعه $d_i = -1$ جداپذیر خطی‌اند.

– سطح تصمیم‌گیری (ابرصفحه) زیر را درنظر بگیرید:



بنابراین

$$\mathbf{w}^T\mathbf{x} + b \geq 0 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x} + b < 0 \quad \text{for} \quad d_i = -1$$

– برای $\mathbf{w}$ و b معلوم، جداسازی بین این ابرصفحه و نزدیک‌ترین داده را «حاشیه جداسازی» نامیده و با ρ نشان می‌دهیم.

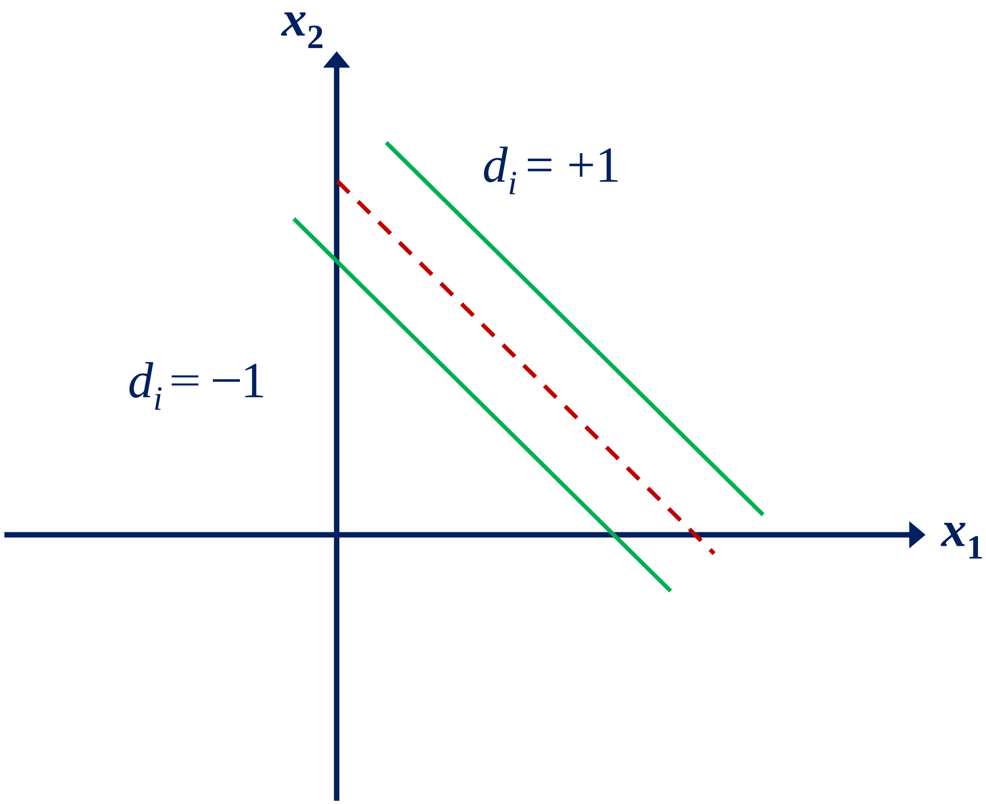
– هدف SVM یافتن ابرصفحه‌ای است که حاشیه جداسازی (ρ) را بیشینه کند.

– چنین سطح تصمیم‌گیری را ابرصفحه بهینه می‌نامند.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- این ضرب داخلی در تشکیل ابرصفحه را از دیدگاه متفاوتی بررسی می‌کنیم.

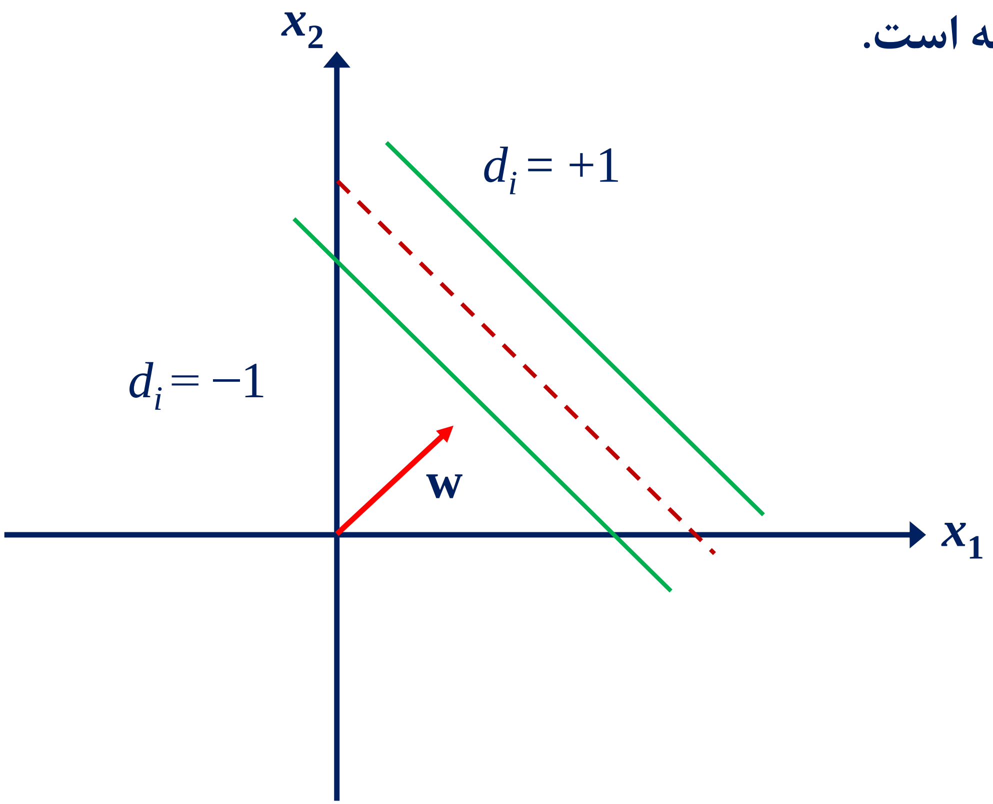


ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- این ضرب داخلی در تشکیل ابرصفحه را از دیدگاه متفاوتی بررسی می‌کنیم.

بردار $\mathbf{w}$ را درنظر بگیرید که عمود بر ابرصفحه است.



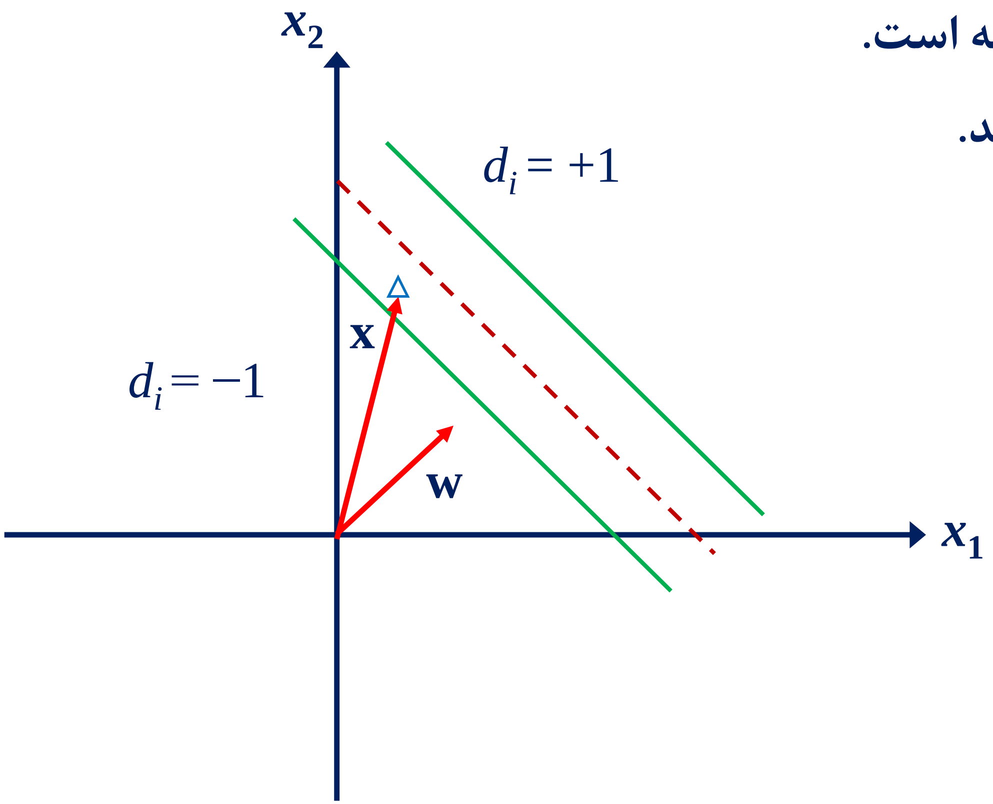
ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- این ضرب داخلی در تشکیل ابرصفحه را از دیدگاه متفاوتی بررسی می‌کنیم.

بردار $\mathbf{w}$ را درنظر بگیرید که عمود بر ابرصفحه است.

بردار $\mathbf{x}$ از مجموعه داده‌ها را نیز درنظر بگیرید.



ماشین بردار پشتیبان (SVM)

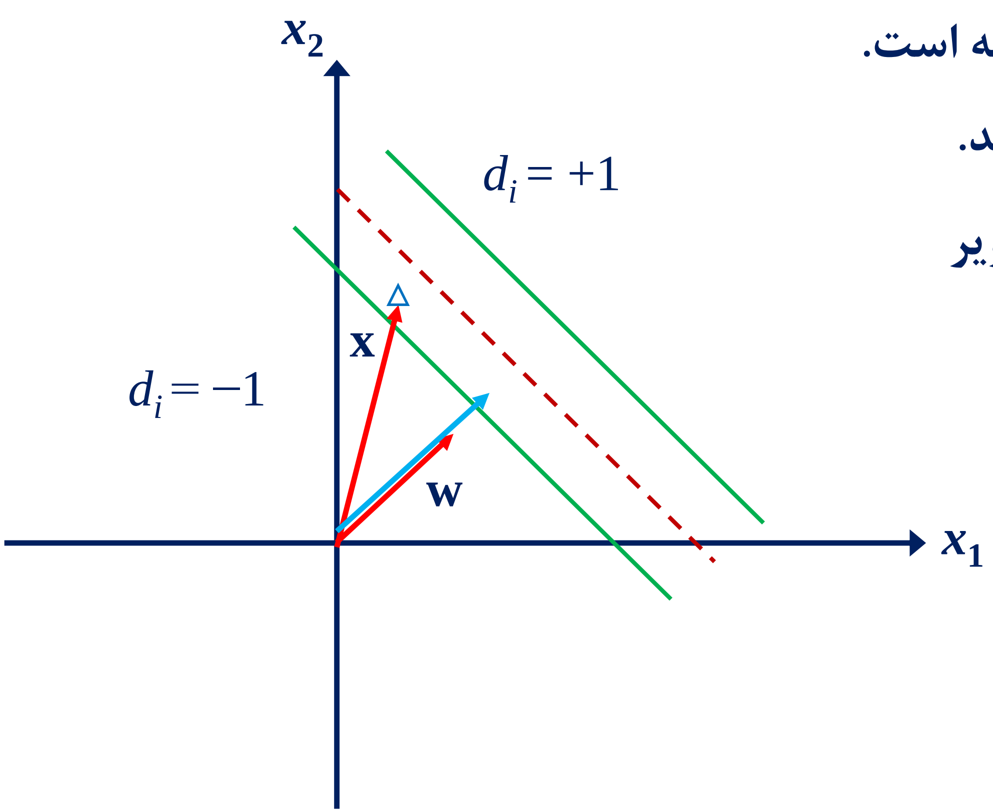
۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- این ضرب داخلی در تشکیل ابرصفحه را از دیدگاه متفاوتی بررسی می‌کنیم.

بردار **w** را درنظر بگیرید که عمود بر ابرصفحه است.

بردار **x** از مجموعه داده‌ها را نیز درنظر بگیرید.

- ضرب داخلی **w** و **x** را می‌توان به‌صورت تصویر **x** برروی **w** درنظر گرفت.



ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

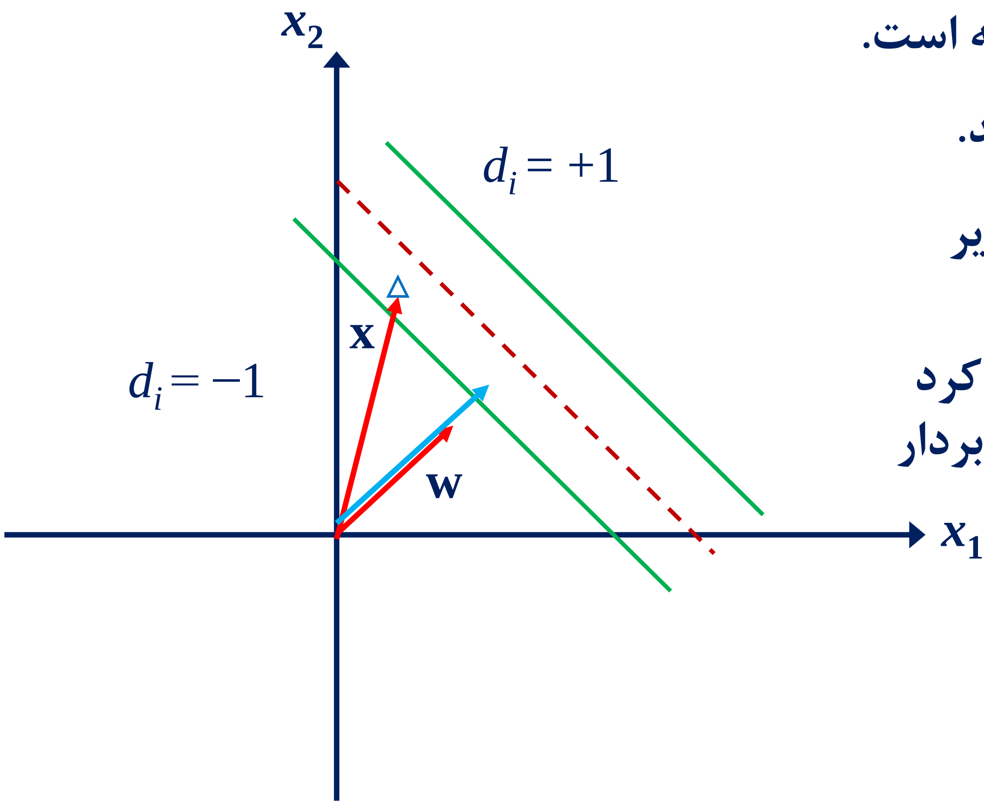
- این ضرب داخلی در تشکیل ابرصفحه را از دیدگاه متفاوتی بررسی می‌کنیم.

  بردار $\mathbf{w}$ را درنظر بگیرید که عمود بر ابرصفحه است.

  بردار $\mathbf{x}$ از مجموعه داده‌ها را نیز درنظر بگیرید.

- ضرب داخلی $\mathbf{w}$ و $\mathbf{x}$ را می‌توان به‌صورت تصویر $\mathbf{x}$ برروی $\mathbf{w}$ درنظر گرفت.

- حال باید به این ضرب داخلی، مقداری اضافه کرد (پیشقدر) یا از آن کم‌کرد (آستانه) تا این که بردار $\mathbf{x}$ به‌طور صحیح کلاسه‌بندی شود.



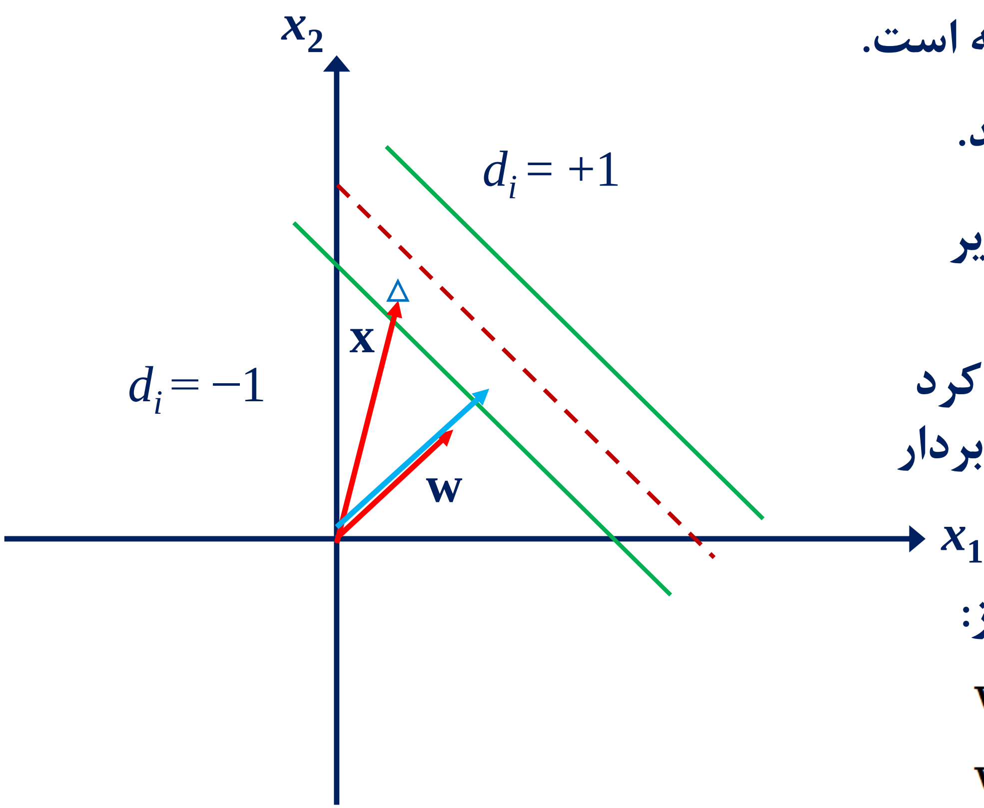
ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- این ضرب داخلی در تشکیل ابرصفحه را از دیدگاه متفاوتی بررسی می‌کنیم.

  بردار $\mathbf{w}$ را درنظر بگیرید که عمود بر ابرصفحه است.

  بردار $\mathbf{x}$ از مجموعه داده‌ها را نیز درنظر بگیرید.



- ضرب داخلی $\mathbf{w}$ و $\mathbf{x}$ را می‌توان به‌صورت تصویر $\mathbf{x}$ برروی $\mathbf{w}$ درنظر گرفت.
- حال باید به این ضرب داخلی، مقداری اضافه کرد (پیشقدر) یا از آن کم‌کرد (آستانه) تا این که بردار $\mathbf{x}$ به‌طور صحیح کلاسه‌بندی شود.
- در نتیجه، قاعده تصمیم‌گیری عبارت است از:

$$\mathbf{w}^T\mathbf{x} + b \geq 0 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x} + b < 0 \quad \text{for} \quad d_i = -1$$

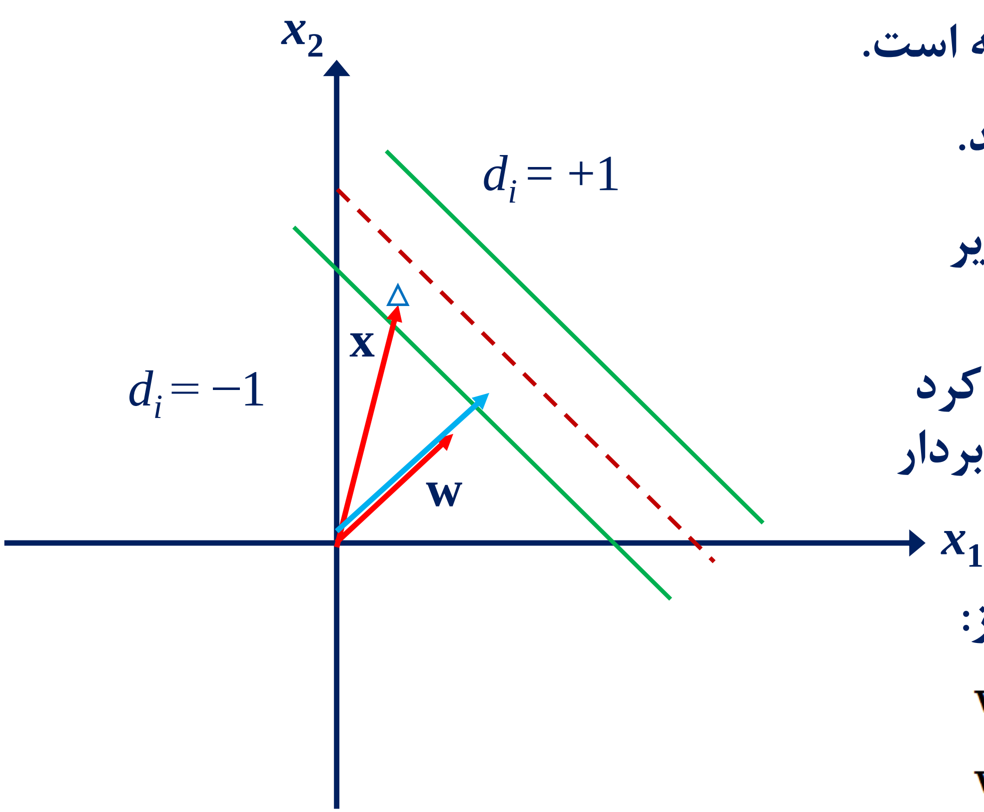
ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– این ضرب داخلی در تشکیل ابرصفحه را از دیدگاه متفاوتی بررسی می‌کنیم.

بردار $\mathbf{w}$ را درنظر بگیرید که عمود بر ابرصفحه است.

بردار $\mathbf{x}$ از مجموعه داده‌ها را نیز درنظر بگیرید.



– ضرب داخلی $\mathbf{w}$ و $\mathbf{x}$ را می‌توان به‌صورت تصویر $\mathbf{x}$ برروی $\mathbf{w}$ درنظر گرفت.

– حال باید به این ضرب داخلی، مقداری اضافه کرد (پیشقدر) یا از آن کم‌کرد (آستانه) تا این که بردار $\mathbf{x}$ به‌طور صحیح کلاسه‌بندی شود.

– در نتیجه، قاعده تصمیم‌گیری عبارت است از:

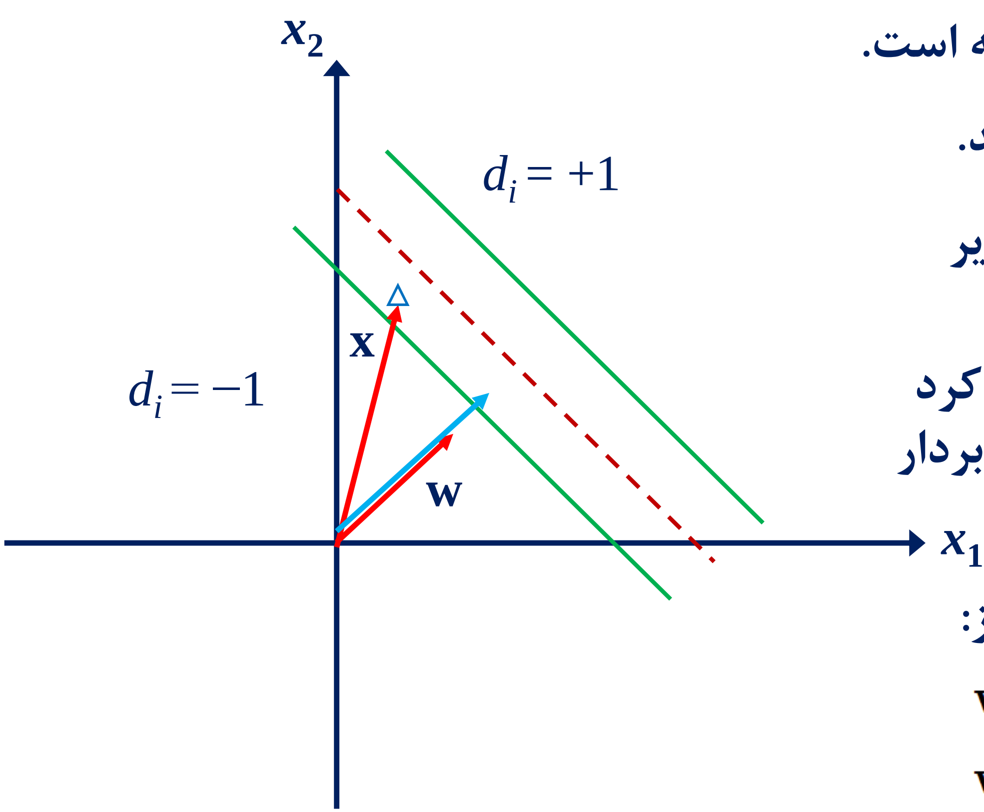
$$\mathbf{w}^T\mathbf{x} + b \geq 0 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x} + b < 0 \quad \text{for} \quad d_i = -1$$

– ولی $\mathbf{w}$ها و bهای بسیار متفاوتی می‌توانند این کار انجام دهد.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– این ضرب داخلی در تشکیل ابرصفحه را از دیدگاه متفاوتی بررسی می‌کنیم.



بردار $\mathbf{w}$ را درنظر بگیرید که عمود بر ابرصفحه است.

بردار $\mathbf{x}$ از مجموعه داده‌ها را نیز درنظر بگیرید.

– ضرب داخلی $\mathbf{w}$ و $\mathbf{x}$ را می‌توان به‌صورت تصویر $\mathbf{x}$ برروی $\mathbf{w}$ درنظر گرفت.

– حال باید به این ضرب داخلی، مقداری اضافه کرد (پیشقدر) یا از آن کم‌کرد (آستانه) تا این که بردار $\mathbf{x}$ به‌طور صحیح کلاسه‌بندی شود.

– در نتیجه، قاعده تصمیم‌گیری عبارت است از:

$$\mathbf{w}^T\mathbf{x} + b \geq 0 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x} + b < 0 \quad \text{for} \quad d_i = -1$$

– ولی $\mathbf{w}$ها و bهای بسیار متفاوتی می‌توانند این کار انجام دهد.

– به‌دنبال مقادیر بهینه برای $\mathbf{w}$و b هستیم که حاشیه جداسازی را بیشینه‌کند.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- بنابراین، مساله عبارت است از یافتن پارامترهای $\mathbf{w}_o$ و b_o برای ابرصفحه بهینه با توجه به مجموعه آموزش $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- بنابراین، مساله عبارت است از یافتن پارامترهای $\mathbf{w}_o$ و b_o برای ابرصفحه بهینه با توجه به مجموعه آموزش $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$.

- بنابراین، ابرصفحه بهینه عبارت است از:

$$\mathbf{w}_o^T \mathbf{x} + b_o = 0$$

ماشین بردار پشتیبان (SVM)

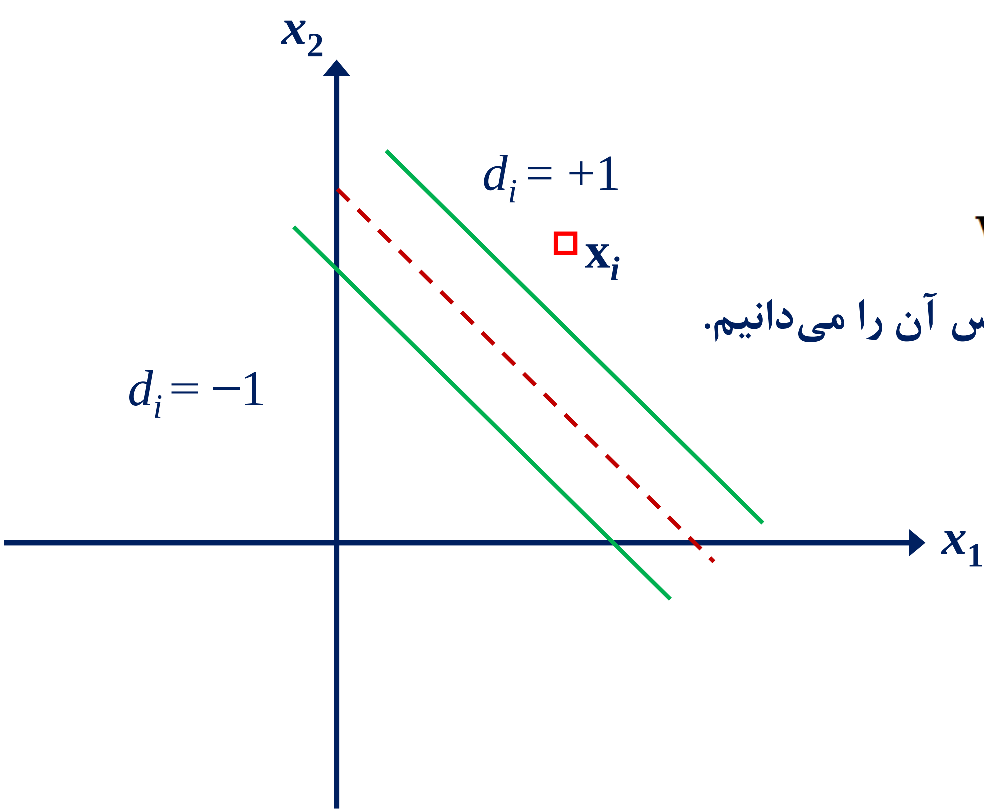
۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– بنابراین، مساله عبارت است از یافتن پارامترهای $\mathbf{w}_o$ و b_o برای ابرصفحه بهینه با توجه به مجموعه آموزش $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$.

– بنابراین، ابرصفحه بهینه عبارت است از:

$$\mathbf{w}_o^T \mathbf{x} + b_o = 0$$

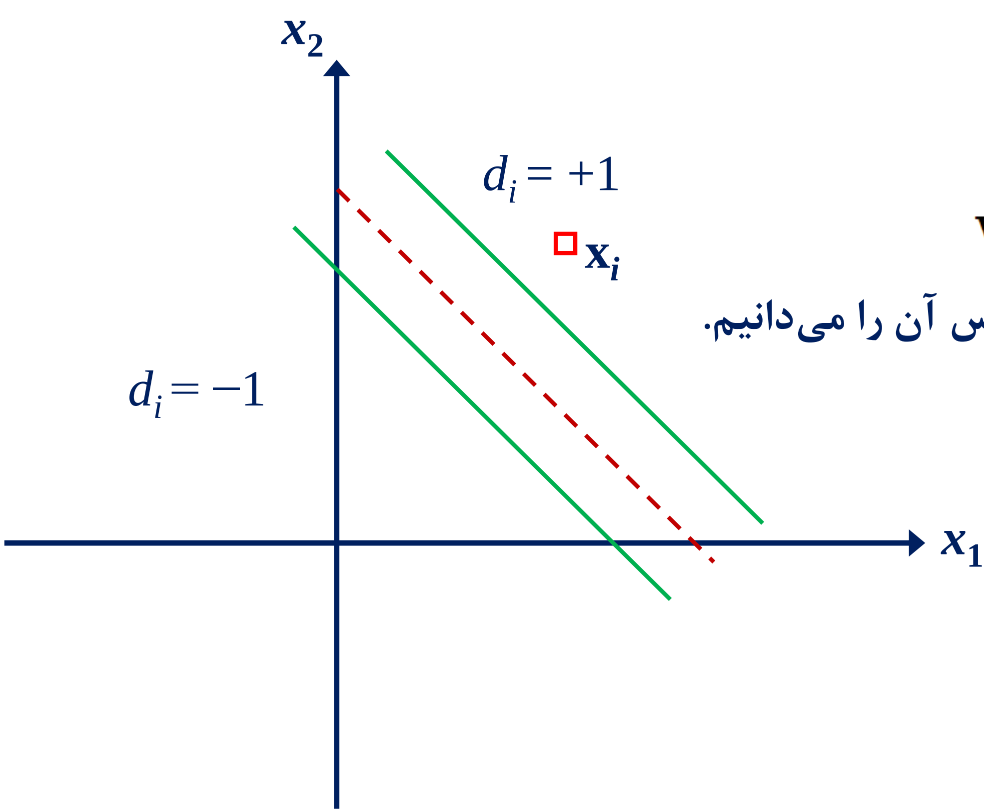
– اکنون داده مشخص $\mathbf{x}_i$ را درنظر بگیرید که کلاس آن را می‌دانیم.



ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– بنابراین، مساله عبارت است از یافتن پارامترهای $\mathbf{w}_o$ و b_o برای ابرصفحه بهینه با توجه به مجموعه آموزش $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$.



– بنابراین، ابرصفحه بهینه عبارت است از:

$$\mathbf{w}_o^T \mathbf{x} + b_o = 0$$

– اکنون داده مشخص $\mathbf{x}_i$ را درنظر بگیرید که کلاس آن را می‌دانیم.

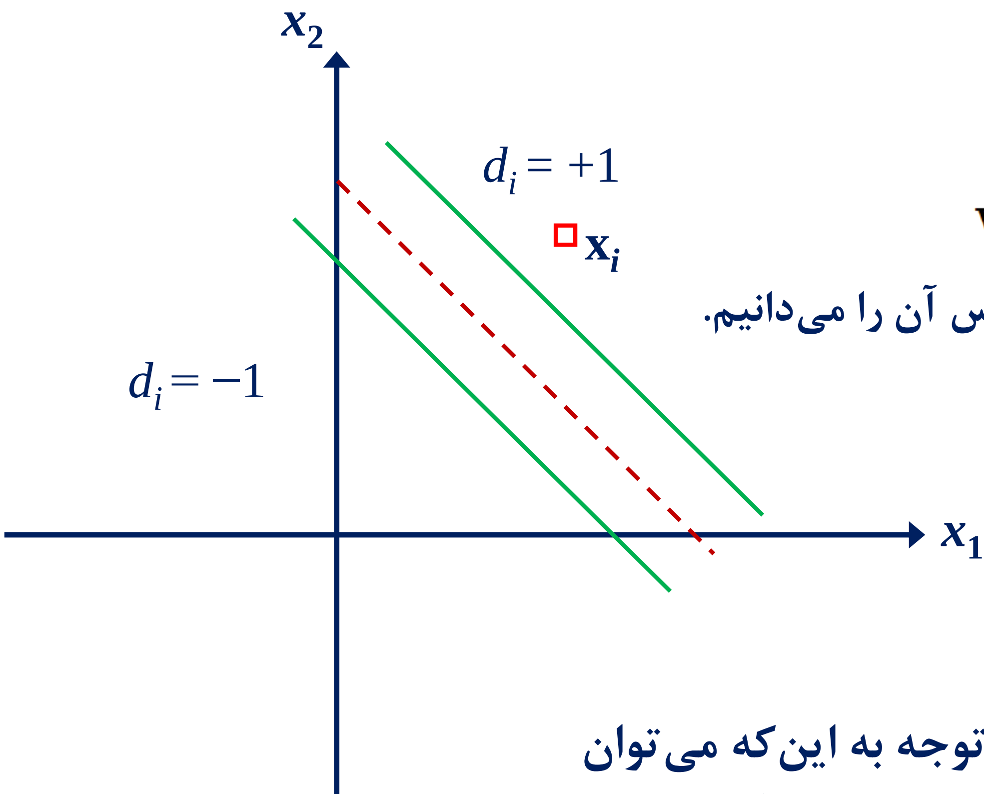
– بدون درنظرگرفتن حاشیه جداسازی، باید رابطه زیر برای $\mathbf{x}_i$ برقرارباشد:

$$\mathbf{w}^T \mathbf{x}_i + b \geq 0 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T \mathbf{x}_i + b < 0 \quad \text{for} \quad d_i = -1$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– بنابراین، مساله عبارت است از یافتن پارامترهای $\mathbf{w}_o$ و b_o برای ابرصفحه بهینه با توجه به مجموعه آموزش $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$.



– بنابراین، ابرصفحه بهینه عبارت است از:

$$\mathbf{w}_o^T \mathbf{x} + b_o = 0$$

– اکنون داده مشخص $\mathbf{x}_i$ را درنظر بگیرید که کلاس آن را می‌دانیم.

– بدون درنظرگرفتن حاشیه جداسازی،
باید رابطه زیر برای $\mathbf{x}_i$ برقرارباشد:

$$\mathbf{w}^T \mathbf{x}_i + b \geq 0 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T \mathbf{x}_i + b < 0 \quad \text{for} \quad d_i = -1$$

– لیکن با درنظرگرفتن حاشیه جداسازی، و با توجه به این‌که می‌توان بردارها را نرمال کرد، رابطه زیر برای کلاسه‌بندی $\mathbf{x}_i$ می‌توان نوشت:

$$\mathbf{w}^T \mathbf{x}_i + b \geq 1 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T \mathbf{x}_i + b \leq -1 \quad \text{for} \quad d_i = -1$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \quad \text{for} \quad d_i = -1$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \quad \text{for} \quad d_i = -1$$

- این دو شرط را می‌توان در قالب یک شرط به این فرم نوشت:

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \quad \text{for} \quad i = 1,\ldots,N$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \quad \text{for} \quad d_i = -1$$

– این دو شرط را می‌توان در قالب یک شرط به این فرم نوشت:

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \quad \text{for} \quad i = 1,\ldots,N$$

– در نتیجه:

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \geq 0 \quad \text{for} \quad i = 1,\ldots,N$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \quad \text{for} \quad d_i = -1$$

- این دو شرط را می‌توان در قالب یک شرط به این فرم نوشت:

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \quad \text{for} \quad i = 1,\dots,N$$

- در نتیجه:

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \geq 0 \quad \text{for} \quad i = 1,\dots,N$$

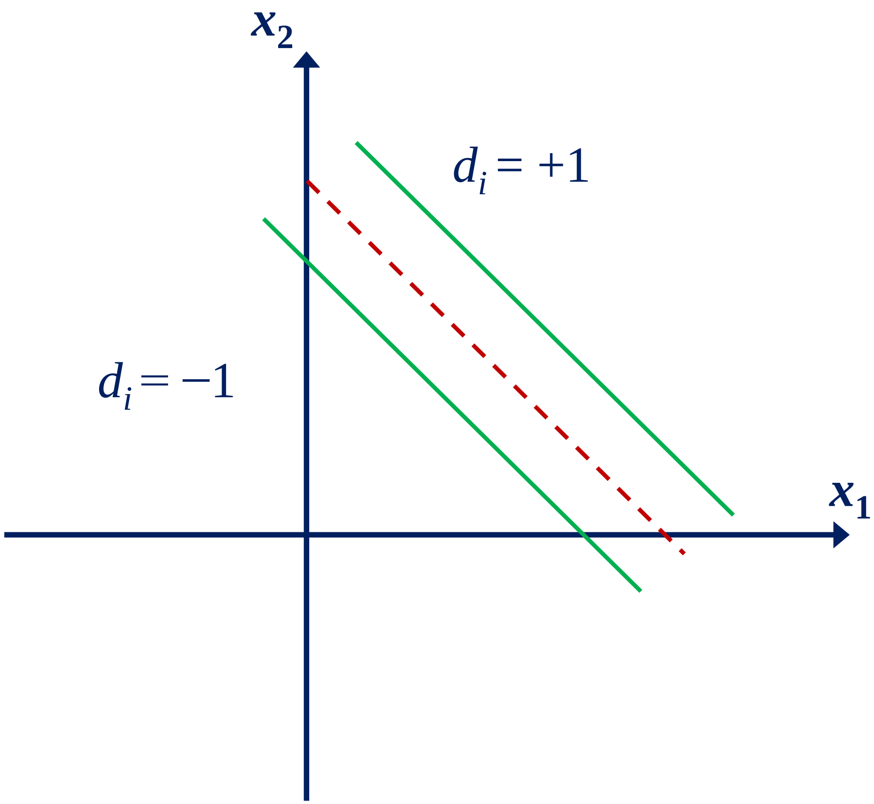
- برای داده‌هایی که برروی حاشیه جداسازی قرارمی‌گیرند

$$\boxed{d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 = 0}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

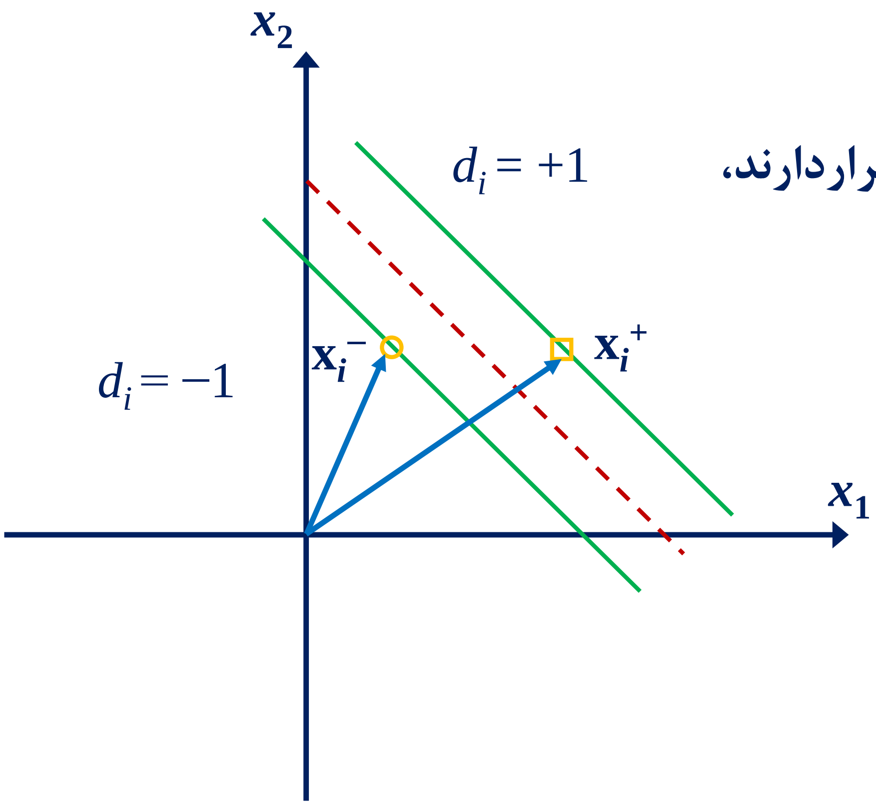
- به خاطر بیاورید که مساله بهینه‌سازی عبارت بود از کلاسه‌بندی الگوها به‌طوری که حاشیه جداسازی بیشینه شود.



ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– به خاطر بیاورید که مساله بهینه‌سازی عبارت بود از کلاسه‌بندی الگوها به‌طوری که حاشیه جداسازی بیشینه شود.

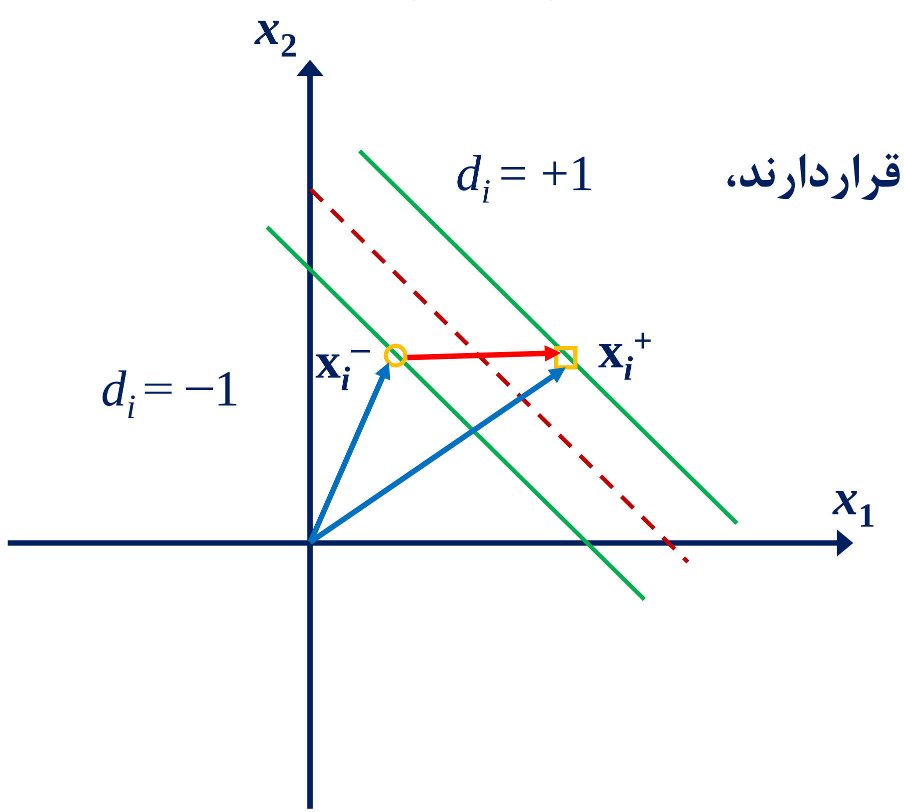


– دو الگوی $\mathbf{x}_i^+$ و $\mathbf{x}_i^-$ که برروی حاشیه جداسازی قراردارند، را درنظربگیرید.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– به خاطر بیاورید که مساله بهینه‌سازی عبارت بود از کلاسه‌بندی الگوها به‌طوری که حاشیه جداسازی بیشینه شود.



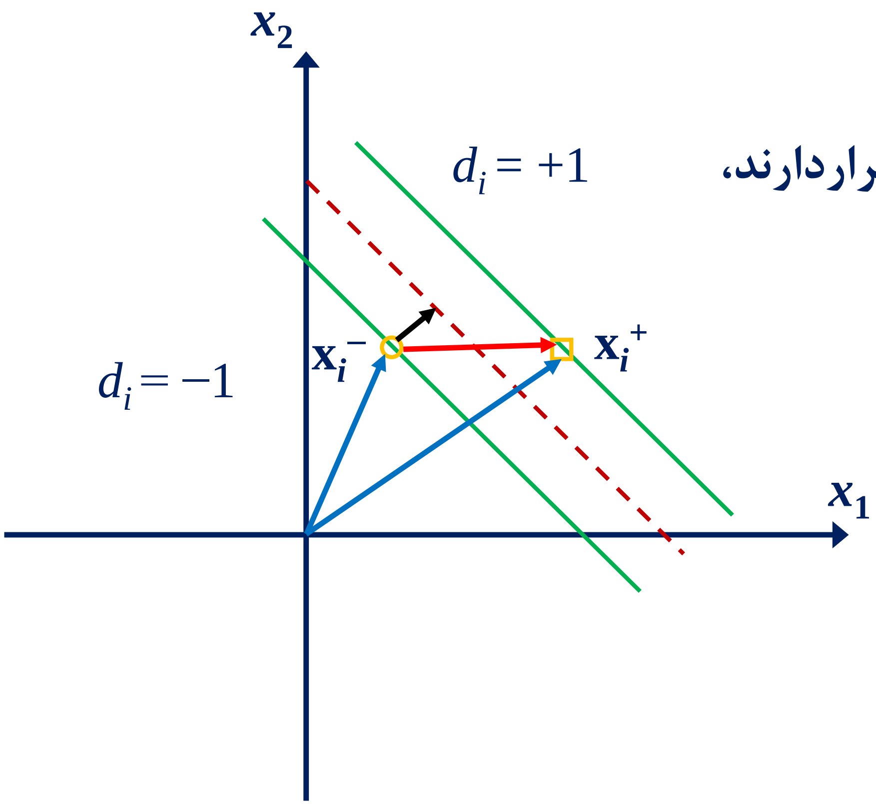
– دو الگوی $\mathbf{x}_i^+$ و $\mathbf{x}_i^-$ که برروی حاشیه جداسازی قراردارند، را درنظربگیرید.

– تفاضل این دو بردار برابر است با $(\mathbf{x}_i^+ - \mathbf{x}_i^-)$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- به خاطر بیاورید که مساله بهینه‌سازی عبارت بود از کلاسه‌بندی الگوها به‌طوری که حاشیه جداسازی بیشینه شود.



- دو الگوی $\mathbf{x}_i^+$ و $\mathbf{x}_i^-$ که برروی حاشیه جداسازی قراردارند، را درنظربگیرید.

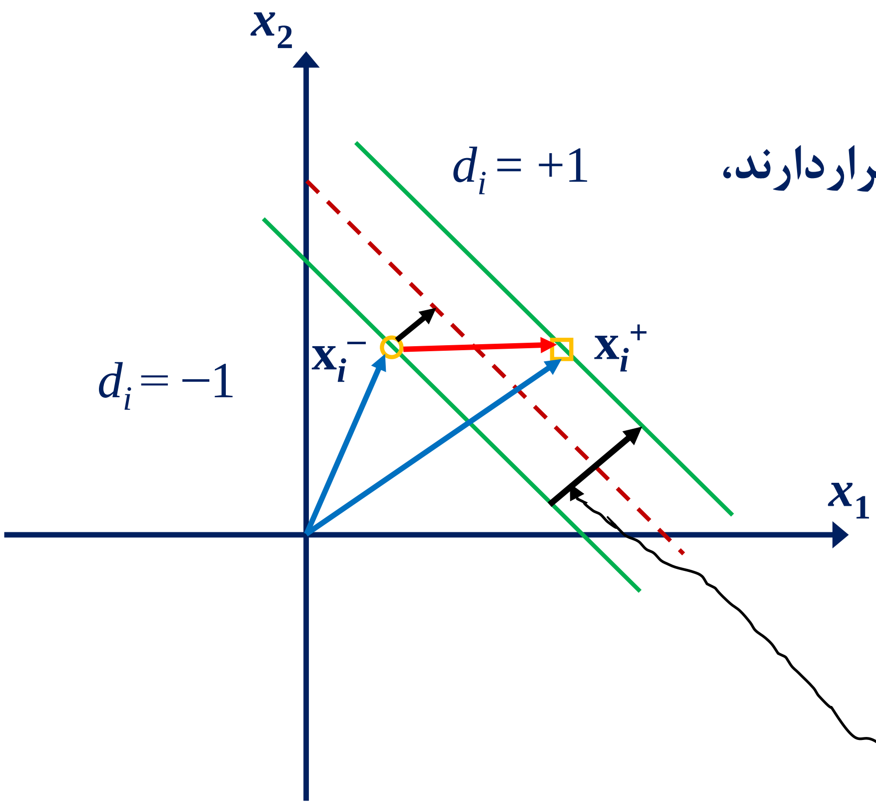
- تفاضل این دو بردار برابر است با $(\mathbf{x}_i^+ - \mathbf{x}_i^-)$

- چنانچه این تفاضل را در بردار نرمال شده $\mathbf{w}$ ضرب داخلی‌کنیم، عرض این حاشیه جداسازی به‌دست می‌آید:

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- به خاطر بیاورید که مساله بهینه‌سازی عبارت بود از کلاسه‌بندی الگوها به‌طوری که حاشیه جداسازی بیشینه شود.



- دو الگوی $\mathbf{x}_i^+$ و $\mathbf{x}_i^-$ که برروی حاشیه جداسازی قراردارند، را درنظربگیرید.

- تفاضل این دو بردار برابر است با $(\mathbf{x}_i^+ - \mathbf{x}_i^-)$

- چنانچه این تفاضل را در بردار نرمال شده $\mathbf{w}$ ضرب داخلی‌کنیم، عرض این حاشیه جداسازی به‌دست می‌آید:

$$\text{width} = (\mathbf{x}_i^+ - \mathbf{x}_i^-)^T \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- از قبل داشتیم:

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \quad \text{for} \quad d_i = -1$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- از قبل داشتیم:

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \quad \text{for} \quad d_i = -1$$

- بنابراین، برای داده‌هایی که برروی حاشیه قرارگرفته‌اند

$$\mathbf{w}^T\mathbf{x}_i^+ = 1 - b$$
$$\mathbf{w}^T\mathbf{x}_i^- = -1 - b$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– از قبل داشتیم:

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \quad \text{for} \quad d_i = -1$$

– بنابراین، برای داده‌هایی که برروی حاشیه قرارگرفته‌اند

$$\mathbf{w}^T\mathbf{x}_i^+ = 1 - b$$
$$\mathbf{w}^T\mathbf{x}_i^- = -1 - b$$

– درنتیجه

$$\text{width} = (\mathbf{x}_i^+ - \mathbf{x}_i^-)^T \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– از قبل داشتیم:

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \quad \text{for} \quad d_i = -1$$

– بنابراین، برای داده‌هایی که برروی حاشیه قرارگرفته‌اند

$$\mathbf{w}^T\mathbf{x}_i^+ = 1 - b$$
$$\mathbf{w}^T\mathbf{x}_i^- = -1 - b$$

– درنتیجه

$$\text{width} = (\mathbf{x}_i^+ - \mathbf{x}_i^-)^T \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

$$\text{width} = (1 - b + 1 + b)\frac{1}{\|\mathbf{w}\|}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– از قبل داشتیم:

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \quad \text{for} \quad d_i = -1$$

– بنابراین، برای داده‌هایی که برروی حاشیه قرارگرفته‌اند

$$\mathbf{w}^T\mathbf{x}_i^+ = 1 - b$$
$$\mathbf{w}^T\mathbf{x}_i^- = -1 - b$$

– درنتیجه

$$\text{width} = (\mathbf{x}_i^+ - \mathbf{x}_i^-)^T \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

$$\text{width} = (1 - b + 1 + b)\frac{1}{\|\mathbf{w}\|}$$

– و در نهایت

$$\text{width} = \frac{2}{\|\mathbf{w}\|}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- با حذف ضریب ۲

$$\text{width} = \frac{1}{\|\mathbf{w}\|}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- با حذف ضریب ۲

$$\text{width} = \frac{1}{\|\mathbf{w}\|}$$

- بیشینه‌کردن رابطه فوق معادل است با کمینه‌کردن اندازه بردار w

$$\min \|\mathbf{w}\|$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- با حذف ضریب ۲

$$\text{width} = \frac{1}{\|\mathbf{w}\|}$$

- بیشینه‌کردن رابطه فوق معادل است با کمینه‌کردن اندازه بردار w

$$\min \|\mathbf{w}\| \implies \min \frac{1}{2}\|\mathbf{w}\|^2$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- با حذف ضریب ۲

$$\text{width} = \frac{1}{\|\mathbf{w}\|}$$

- بیشینه‌کردن رابطه فوق معادل است با کمینه‌کردن اندازه بردار w

$$\min \|\mathbf{w}\| \implies \min \frac{1}{2}\|\mathbf{w}\|^2 \implies \boxed{\min \frac{1}{2}\mathbf{w}^T\mathbf{w}}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– با حذف ضریب ۲

$$\text{width} = \frac{1}{\|\mathbf{w}\|}$$

– بیشینه‌کردن رابطه فوق معادل است با کمینه‌کردن اندازه بردار w

$$\min \|\mathbf{w}\| \implies \min \frac{1}{2}\|\mathbf{w}\|^2 \implies \boxed{\min \frac{1}{2}\mathbf{w}^T\mathbf{w}}$$

بیشینه‌کردن حاشیه جداسازی بین الگوهای مثبت و منفی معادل است با کمینه‌سازی اندازه بردار w.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- با حذف ضریب ۲

$$\text{width} = \frac{1}{\|\mathbf{w}\|}$$

- بیشینه‌کردن رابطه فوق معادل است با کمینه‌کردن اندازه بردار w

$$\min \|\mathbf{w}\| \implies \min \frac{1}{2}\|\mathbf{w}\|^2 \implies \boxed{\min \frac{1}{2}\mathbf{w}^T\mathbf{w}}$$

بیشینه‌کردن حاشیه جداسازی بین الگوهای مثبت و منفی معادل است با کمینه‌سازی اندازه بردار w.

- به‌طور خلاصه:

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– با حذف ضریب ۲

$$\text{width} = \frac{1}{\|\mathbf{w}\|}$$

– بیشینه‌کردن رابطه فوق معادل است با کمینه‌کردن اندازه بردار **w**

$$\min \|\mathbf{w}\| \implies \min \frac{1}{2}\|\mathbf{w}\|^2 \implies \boxed{\min \frac{1}{2}\mathbf{w}^T\mathbf{w}}$$

بیشینه‌کردن حاشیه جداسازی بین الگوهای مثبت و منفی معادل است با کمینه‌سازی اندازه بردار w.

– به‌طور خلاصه:

- ابر صفحه بهینه $\mathbf{w}_o^T\mathbf{x} + b_o = 0$ همواره برای داده‌های موردنظر،یکتا است.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– با حذف ضریب ۲

$$\text{width} = \frac{1}{\|\mathbf{w}\|}$$

– بیشینه‌کردن رابطه فوق معادل است با کمینه‌کردن اندازه بردار $\mathbf{w}$

$$\min \|\mathbf{w}\| \implies \min \frac{1}{2}\|\mathbf{w}\|^2 \implies \boxed{\min \frac{1}{2}\mathbf{w}^T\mathbf{w}}$$

بیشینه‌کردن حاشیه جداسازی بین الگوهای مثبت و منفی معادل است با کمینه‌سازی اندازه بردار $\mathbf{w}$.

– به‌طور خلاصه:

- ابر صفحه بهینه $\mathbf{w}_o^T\mathbf{x} + b_o = 0$ همواره برای داده‌های موردنظر،یکتا است.
- یعنی بردار بهینه $\mathbf{w}_o$،بیشترین جداسازی ممکن را بین نمونه‌های مثبت و منفی به‌دست می‌دهد.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– با حذف ضریب ۲

$$\text{width} = \frac{1}{\|\mathbf{w}\|}$$

– بیشینه‌کردن رابطه فوق معادل است با کمینه‌کردن اندازه بردار $\mathbf{w}$

$$\min \|\mathbf{w}\| \implies \min \frac{1}{2}\|\mathbf{w}\|^2 \implies \boxed{\min \frac{1}{2}\mathbf{w}^T\mathbf{w}}$$

بیشینه‌کردن حاشیه جداسازی بین الگوهای مثبت و منفی معادل است با کمینه‌سازی اندازه بردار $\mathbf{w}$.

– به‌طور خلاصه:

- ابر صفحه بهینه $\mathbf{w}_o^T\mathbf{x} + b_o = 0$ همواره برای داده‌های موردنظر،یکتا است.
- یعنی بردار بهینه $\mathbf{w}_o$،بیشترین جداسازی ممکن را بین نمونه‌های مثبت و منفی به‌دست می‌دهد.
- این بهینگی با کمینه‌کردن اندازه بردار $\mathbf{w}$ حاصل می‌شود.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- بنابراین، مساله بهینه‌سازی مقید به‌صورت زیر بیان می‌شود:

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- بنابراین، مساله بهینه‌سازی مقید به‌صورت زیر بیان می‌شود:

برای داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$ مطلوب است یافتن مقدار بهینه برای $\mathbf{w}_o$ و b_o به‌طوری که تابع هزینه و قید زیر برآورده شوند:

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– بنابراین، مساله بهینه‌سازی مقید به‌صورت زیر بیان می‌شود:

برای داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$ مطلوب است یافتن مقدار بهینه برای $\mathbf{w}_o$ و b_o به‌طوری که تابع هزینه و قید زیر برآورده شوند:

$$\phi(\mathbf{w}) = \min \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– بنابراین، مساله بهینه‌سازی مقید به‌صورت زیر بیان می‌شود:

برای داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$ مطلوب است یافتن مقدار بهینه برای $\mathbf{w}_o$ و b_o به‌طوری که تابع هزینه و قید زیر برآورده شوند:

$$\phi(\mathbf{w}) = \min \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 = 0 \qquad \text{for} \qquad i = 1, \ldots, N$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– بنابراین، مساله بهینه‌سازی مقید به‌صورت زیر بیان می‌شود:

برای داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$ مطلوب است یافتن مقدار بهینه برای $\mathbf{w}_o$ و b_o به‌طوری که تابع هزینه و قید زیر برآورده شوند:

$$\phi(\mathbf{w}) = \min \frac{1}{2} \mathbf{w}^T \mathbf{w}$$

$$d_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 = 0 \qquad \text{for} \qquad i = 1, \ldots, N$$

– به این نوع مساله بهینه‌سازی مقید، «مساله اولیه» (Primal Problem) می‌گویند که به شکل زیر مشخص می‌شود:

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– بنابراین، مساله بهینه‌سازی مقید به‌صورت زیر بیان می‌شود:

برای داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$ مطلوب است یافتن مقدار بهینه برای $\mathbf{w}_o$ و b_o به‌طوری که تابع هزینه و قید زیر برآورده شوند:

$$\phi(\mathbf{w}) = \min \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 = 0 \qquad \text{for} \qquad i = 1, \ldots, N$$

– به این نوع مساله بهینه‌سازی مقید، «مساله اولیه» (Primal Problem) می‌گویند که به شکل زیر مشخص می‌شود:

- تابع هزینه، تابعی محدب از بردار $\mathbf{w}$ است.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– بنابراین، مساله بهینه‌سازی مقید به‌صورت زیر بیان می‌شود:

برای داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$ مطلوب است یافتن مقدار بهینه برای $\mathbf{w}_o$ و b_o به‌طوری که تابع هزینه و قید زیر برآورده شوند:

$$\phi(\mathbf{w}) = \min \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 = 0 \qquad \text{for} \qquad i = 1, \dots, N$$

– به این نوع مساله بهینه‌سازی مقید، «مساله اولیه» (Primal Problem) می‌گویند که به شکل زیر مشخص می‌شود:

- تابع هزینه، تابعی محدب از بردار $\mathbf{w}$ است.
- قیود، برحسب $\mathbf{w}$ خطی‌اند.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$\phi(\mathbf{w}) = \min \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 = 0 \qquad \text{for} \qquad i = 1, \ldots, N$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$\phi(\mathbf{w}) = \min \frac{1}{2} \mathbf{w}^T \mathbf{w}$$

$$d_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 = 0 \qquad \text{for} \qquad i = 1, \ldots, N$$

- مساله بهینه‌سازی مقید با استفاده از روش لاگرانژ حل می‌شود:

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$\phi(\mathbf{w}) = \min \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 = 0 \qquad \text{for} \qquad i = 1, \ldots, N$$

- مساله بهینه‌سازی مقید با استفاده از روش لاگرانژ حل می‌شود:

- تابع لاگرانژ

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1\right]$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$\phi(\mathbf{w}) = \min \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 = 0 \qquad \text{for} \qquad i = 1, \ldots, N$$

– مساله بهینه‌سازی مقید با استفاده از روش لاگرانژ حل می‌شود:

– تابع لاگرانژ

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i \left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \right]$$

– توجه شود که جمله اول تابع لاگرانژ باید کمینه شود. ولی جمله دوم باید بیشینه‌شود (بیشینه‌کردن حاشیه‌جداسازی یا همان عرض خیابان).

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$\phi(\mathbf{w}) = \min \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 = 0 \qquad \text{for} \qquad i = 1,\ldots,N$$

– مساله بهینه‌سازی مقید با استفاده از روش لاگرانژ حل می‌شود:

– تابع لاگرانژ

$$J(\mathbf{w},b,\alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1\right]$$

– توجه شود که جمله اول تابع لاگرانژ باید کمینه شود. ولی جمله دوم باید بیشینه‌شود (بیشینه‌کردن حاشیه‌جداسازی یا همان عرض خیابان).

– یعنی تابع هزینه باید نسبت به $\mathbf{w}$ و b کمینه شود ولی نسبت به آلفا بیشینه شود.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$\phi(\mathbf{w}) = \min \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 = 0 \qquad \text{for} \qquad i = 1, \ldots, N$$

- مساله بهینه‌سازی مقید با استفاده از روش لاگرانژ حل می‌شود:
- تابع لاگرانژ

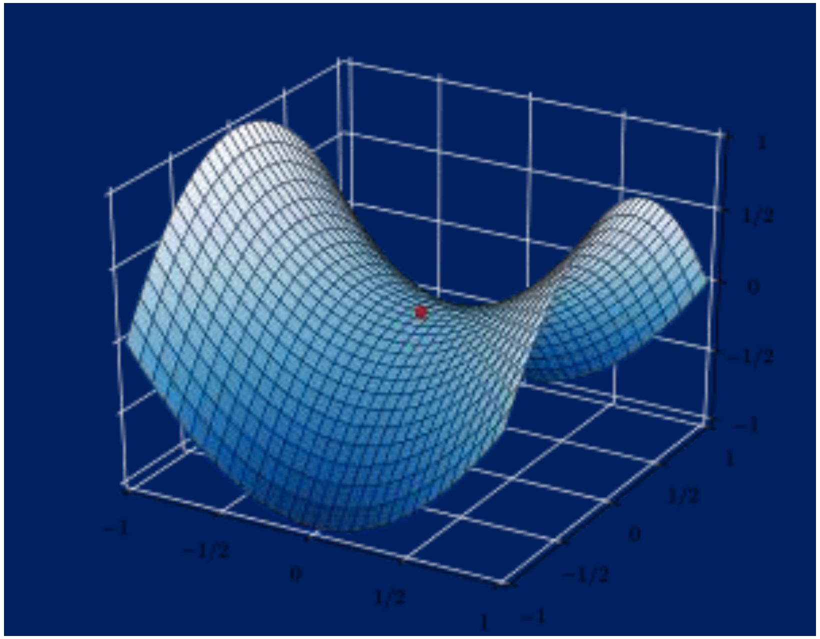
$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i \left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \right]$$

- توجه شود که جمله اول تابع لاگرانژ باید کمینه شود. ولی جمله دوم باید بیشینه‌شود (بیشینه‌کردن حاشیه‌جداسازی یا همان عرض خیابان).
- یعنی تابع هزینه باید نسبت به $\mathbf{w}$ و b کمینه شود ولی نسبت به آلفا بیشینه شود.
- این مساله باعث به‌وجود آمدن نقطه زینی (Saddle Point) می‌شود.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

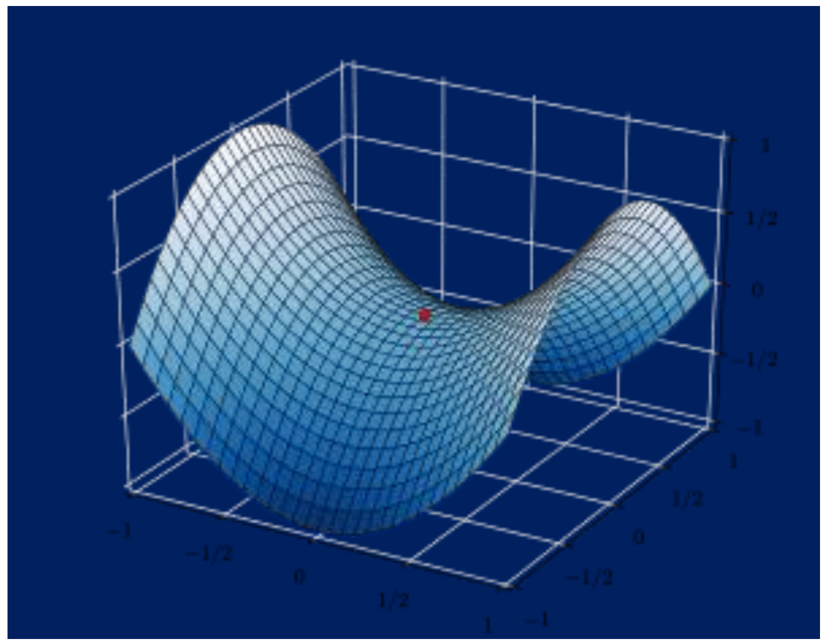
$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i \left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \right]$$



ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i \left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \right]$$

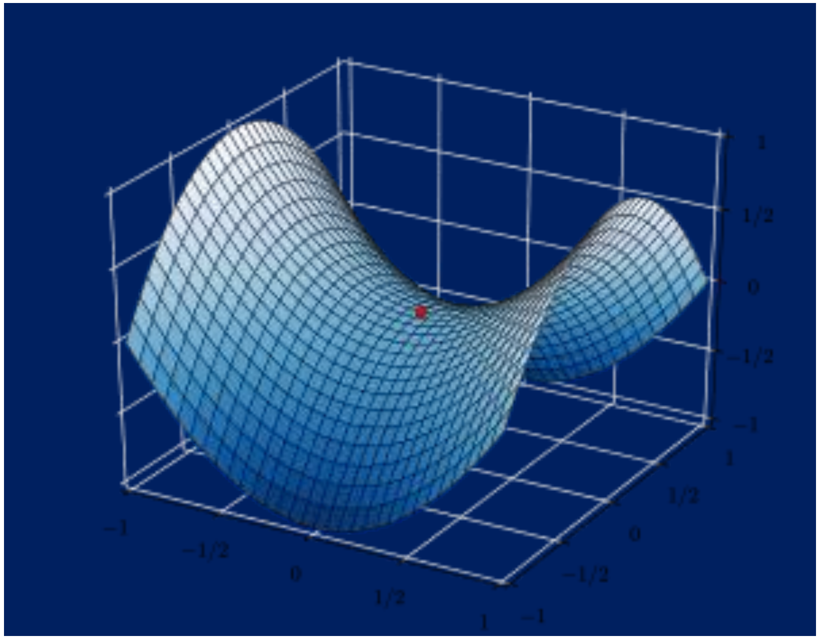


– نقطه زینی تابع لاگرانژ بالا، نقطه‌ای است با ریشه‌های حقیقی ولی علامت مخالف.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha)=\frac{1}{2}\mathbf{w}^T\mathbf{w}-\sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i+b)-1\right]$$



- **نقطه زینی تابع لاگرانژ بالا، نقطه‌ای است با ریشه‌های حقیقی ولی علامت مخالف.**
- **یک چنین تکینگی همواره ناپایدار است.**

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i \left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \right]$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha)=\frac{1}{2}\mathbf{w}^T\mathbf{w}-\sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i+b)-1\right]$$

– فعلا از تابع لاگرانژ نسبت به $\mathbf{w}$ و b مشتق گرفته و برابر صفر قرار می‌دهیم:

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i \left[d_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 \right]$$

– فعلا از تابع لاگرانژ نسبت به $\mathbf{w}$ و b مشتق گرفته و برابر صفر قرار می‌دهیم:

$$\frac{\partial J(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = 0 \quad \Longrightarrow \quad \mathbf{w}_o = \sum_{i=1}^{N} \alpha_i d_i \mathbf{x}_i$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1\right]$$

- فعلا از تابع لاگرانژ نسبت به $\mathbf{w}$ و b مشتق گرفته و برابر صفر قرار می‌دهیم:

$$\frac{\partial J(\mathbf{w},b,\alpha)}{\partial \mathbf{w}} = 0 \quad \Longrightarrow \quad \mathbf{w}_o = \sum_{i=1}^{N}\alpha_i d_i \mathbf{x}_i$$

$$\frac{\partial J(\mathbf{w},b,\alpha)}{\partial b} = 0 \quad \Longrightarrow \quad \sum_{i=1}^{N}\alpha_i d_i = 0$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha)=\frac{1}{2}\mathbf{w}^T\mathbf{w}-\sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i+b)-1\right]$$

– فعلا از تابع لاگرانژ نسبت به $\mathbf{w}$ و b مشتق گرفته و برابر صفر قرار می‌دهیم:

$$\frac{\partial J(\mathbf{w},b,\alpha)}{\partial \mathbf{w}}=0 \quad\Longrightarrow\quad \mathbf{w}_o=\sum_{i=1}^{N}\alpha_i d_i \mathbf{x}_i$$

$$\frac{\partial J(\mathbf{w},b,\alpha)}{\partial b}=0 \quad\Longrightarrow\quad \sum_{i=1}^{N}\alpha_i d_i=0$$

– اگرچه جواب بردار $\mathbf{w}$ یکتا است (برطبق محدب‌بودن لاگرانژ) ولی در مورد آلفا نمی‌توان چنین مطلبی را گفت.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha)=\frac{1}{2}\mathbf{w}^T\mathbf{w}-\sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i+b)-1\right]$$

- فعلا از تابع لاگرانژ نسبت به $\mathbf{w}$ و b مشتق گرفته و برابر صفر قرار می‌دهیم:

$$\frac{\partial J(\mathbf{w},b,\alpha)}{\partial \mathbf{w}}=0 \quad\Longrightarrow\quad \mathbf{w}_o=\sum_{i=1}^{N}\alpha_i d_i \mathbf{x}_i$$

$$\frac{\partial J(\mathbf{w},b,\alpha)}{\partial b}=0 \quad\Longrightarrow\quad \sum_{i=1}^{N}\alpha_i d_i=0$$

- اگرچه جواب بردار $\mathbf{w}$ یکتا است (برطبق محدب‌بودن لاگرانژ) ولی در مورد آلفا نمی‌توان چنین مطلبی را گفت.
- بنابراین، برای شروطی که تساوی آن‌ها برآورده نمی‌شود، باید آلفای نظیر را برابر صفر قرارداد.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha)=\frac{1}{2}\mathbf{w}^T\mathbf{w}-\sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i+b)-1\right]$$

– فعلا از تابع لاگرانژ نسبت به $\mathbf{w}$ و b مشتق گرفته و برابر صفر قرار می‌دهیم:

$$\frac{\partial J(\mathbf{w},b,\alpha)}{\partial \mathbf{w}}=0 \;\Longrightarrow\; \mathbf{w}_o=\sum_{i=1}^{N}\alpha_i d_i \mathbf{x}_i$$

$$\frac{\partial J(\mathbf{w},b,\alpha)}{\partial b}=0 \;\Longrightarrow\; \sum_{i=1}^{N}\alpha_i d_i=0$$

– اگرچه جواب بردار $\mathbf{w}$ یکتا است (برطبق محدب‌بودن لاگرانژ) ولی در مورد آلفا نمی‌توان چنین مطلبی را گفت.

– بنابراین، برای شروطی که تساوی آن‌ها برآورده نمی‌شود، باید آلفای نظیر را برابر صفر قرارداد.

– به‌عبارت دیگر، فقط ضرایبی که دقیقا شرط $\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i+b)-1\right]=0$ را برآورده می‌کنند، می‌توانند مقدار غیرصفر داشته باشند.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha)=\frac{1}{2}\mathbf{w}^T\mathbf{w}-\sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i+b)-1\right]$$

– فعلا از تابع لاگرانژ نسبت به $\mathbf{w}$ و b مشتق گرفته و برابر صفر قرار می‌دهیم:

$$\frac{\partial J(\mathbf{w},b,\alpha)}{\partial \mathbf{w}}=0 \implies \mathbf{w}_o=\sum_{i=1}^{N}\alpha_i d_i \mathbf{x}_i$$

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

– اگرچه جواب بردار $\mathbf{w}$ یکتا است (برطبق محدب‌بودن لاگرانژ) ولی در مورد آلفا نمی‌توان چنین مطلبی را گفت.

– بنابراین، برای شروطی که تساوی آن‌ها برآورده نمی‌شود، باید آلفای نظیر را برابر صفر قرارداد.

– به‌عبارت دیگر، فقط ضرایبی که دقیقا شرط $\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i+b)-1\right]=0$ را برآورده می‌کنند، می‌توانند مقدار غیرصفر داشته باشند.

– این شروط را شروط کاروش-کان-تاکر (Karush-Kahn-Tucker = KKT) می‌نامند.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i \left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \right]$$

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i d_i \mathbf{x}_i \qquad \sum_{i=1}^{N} \alpha_i d_i = 0$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1\right]$$

$$\mathbf{w} = \sum_{i=1}^{N}\alpha_i d_i \mathbf{x}_i \qquad \sum_{i=1}^{N}\alpha_i d_i = 0$$

- با قراردادن دو شرط اخیر در تابع لاگرانژ، نتیجه می‌شود:

$$\begin{aligned} J(\mathbf{w},b,\alpha) = &\frac{1}{2}\left(\sum_{i=1}^{N}\alpha_i d_i \mathbf{x}_i\right)^T\left(\sum_{i=1}^{N}\alpha_i d_i \mathbf{x}_i\right) \\ &- \left(\sum_{i=1}^{N}\alpha_i d_i \mathbf{x}_i\right)^T\left(\sum_{i=1}^{N}\alpha_i d_i \mathbf{x}_i\right) - \sum_{i=1}^{N}\alpha_i d_i b + \sum_{i=1}^{N}\alpha_i \end{aligned}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha)=\frac{1}{2}\mathbf{w}^T\mathbf{w}-\sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i+b)-1\right]$$

$$\mathbf{w}=\sum_{i=1}^{N}\alpha_i d_i\mathbf{x}_i \qquad \sum_{i=1}^{N}\alpha_i d_i=0$$

– با قراردادن دو شرط اخیر در تابع لاگرانژ، نتیجه می‌شود:

$$\begin{aligned}J(\mathbf{w},b,\alpha)=&\frac{1}{2}\left(\sum_{i=1}^{N}\alpha_i d_i\mathbf{x}_i\right)^T\left(\sum_{i=1}^{N}\alpha_i d_i\mathbf{x}_i\right)\\&-\left(\sum_{i=1}^{N}\alpha_i d_i\mathbf{x}_i\right)^T\left(\sum_{i=1}^{N}\alpha_i d_i\mathbf{x}_i\right)-\sum_{i=1}^{N}\alpha_i d_i b+\sum_{i=1}^{N}\alpha_i\end{aligned}$$

– بعد از خلاصه‌سازی:

$$J(\mathbf{w},b,\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j d_i d_j\mathbf{x}_i^T\mathbf{x}_j$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j d_i d_j \mathbf{x}_i^T\mathbf{x}_j$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w}, b, \alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j$$

- نکته بسیار مهم: تابع لاگرانژ بهینه شده، تابعی است از ضرب داخلی بردارهای ورودی.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j d_i d_j \mathbf{x}_i^T\mathbf{x}_j$$

- نکته بسیار مهم: تابع لاگرانژ بهینه شده، تابعی است از ضرب داخلی بردارهای ورودی.

- در مورد ابر صفحه بهینه نیز می‌توان همین امر را مشاهده کرد:

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w}, b, \alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j$$

- نکته بسیار مهم: تابع لاگرانژ بهینه شده، تابعی است از ضرب داخلی بردارهای ورودی.

- در مورد ابر صفحه بهینه نیز می‌توان همین امر را مشاهده کرد:

$$\mathbf{w}_o^T \mathbf{x} + b_o = 0$$

$$\mathbf{w}_o = \sum_{i=1}^{N} \alpha_i d_i \mathbf{x}_i$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w}, b, \alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j$$

– نکته بسیار مهم: تابع لاگرانژ بهینه شده، تابعی است از ضرب داخلی بردارهای ورودی.

– در مورد ابر صفحه بهینه نیز می‌توان همین امر را مشاهده کرد:

$$\begin{aligned} \mathbf{w}_o^T \mathbf{x} + b_o &= 0 \\ \mathbf{w}_o &= \sum_{i=1}^{N} \alpha_i d_i \mathbf{x}_i \end{aligned} \quad \Rightarrow \quad \begin{aligned} &\sum_{i=1}^{N} \alpha_i d_i \mathbf{x}_i^T \mathbf{x} + b \geq 0 \quad \text{for} \quad d_i = +1 \\ &\sum_{i=1}^{N} \alpha_i d_i \mathbf{x}_i^T \mathbf{x} + b < 0 \quad \text{for} \quad d_i = -1 \end{aligned}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

$$J(\mathbf{w},b,\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j d_i d_j \mathbf{x}_i^T\mathbf{x}_j$$

- نکته بسیار مهم: تابع لاگرانژ بهینه شده، تابعی است از ضرب داخلی بردارهای ورودی.

- در مورد ابر صفحه بهینه نیز می‌توان همین امر را مشاهده کرد:

$$\begin{aligned}\mathbf{w}_o^T\mathbf{x}+b_o&=0\\ \mathbf{w}_o&=\sum_{i=1}^{N}\alpha_i d_i\mathbf{x}_i\end{aligned}\quad\Rightarrow\quad\begin{aligned}\sum_{i=1}^{N}\alpha_i d_i\mathbf{x}_i^T\mathbf{x}+b\geq 0 &\quad \text{for}\quad d_i=+1\\ \sum_{i=1}^{N}\alpha_i d_i\mathbf{x}_i^T\mathbf{x}+b< 0 &\quad \text{for}\quad d_i=-1\end{aligned}$$