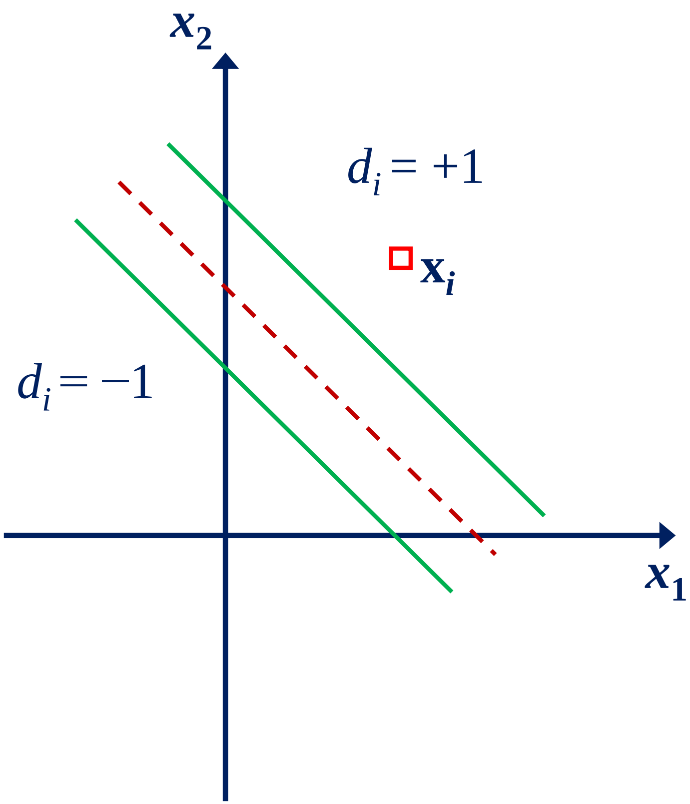
شبکه‌های عصبی مصنوعی

جلسه چهاردهم:
ماشین بردار پشتیبان (۲)
(Support Vector Machine = SVM)

محمد فرخی (farrokhi AT iust.ac.ir)

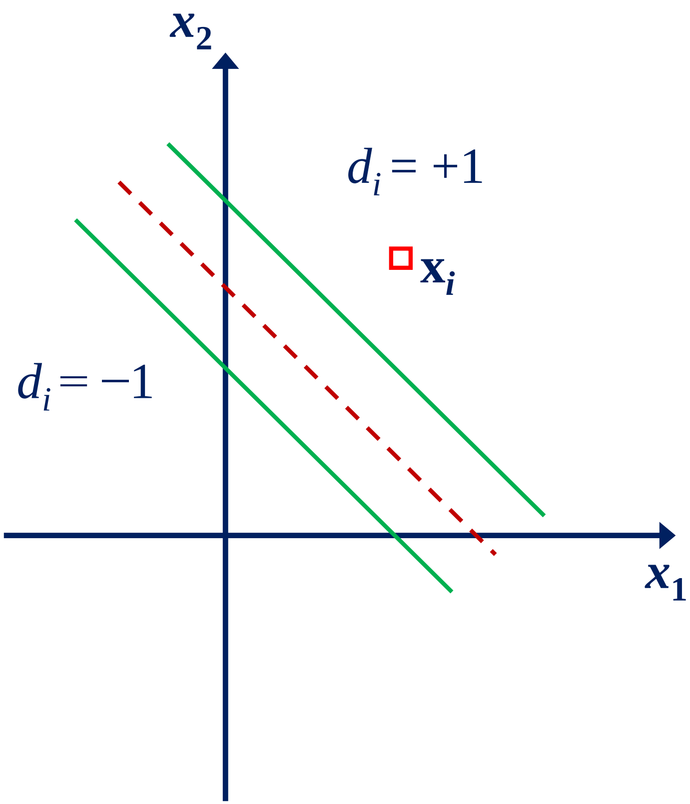
ماشین بردار پشتیبان (SVM)

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

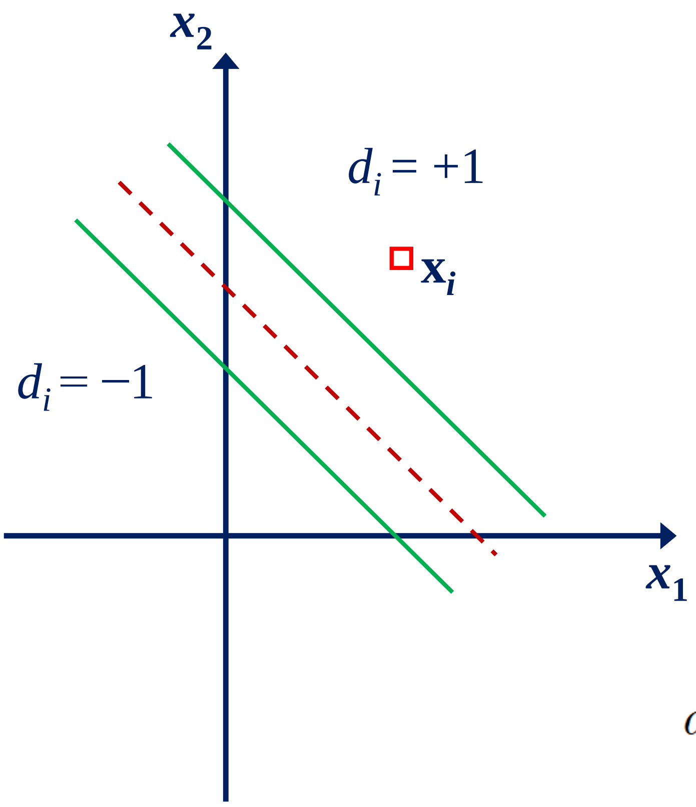
$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 \quad \text{for} \quad d_i = +1$$
$$\mathbf{w}^T\mathbf{x}_i + b \leq -1 \quad \text{for} \quad d_i = -1$$

– 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:

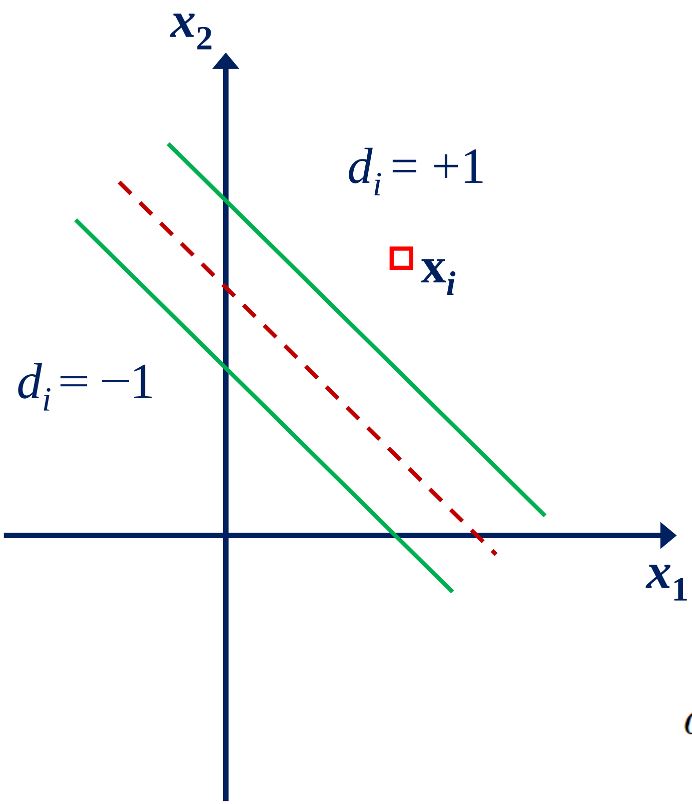
$$d_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \quad \text{for} \quad i = 1, \ldots, N$$

– 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:

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

– در نتیجه:

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

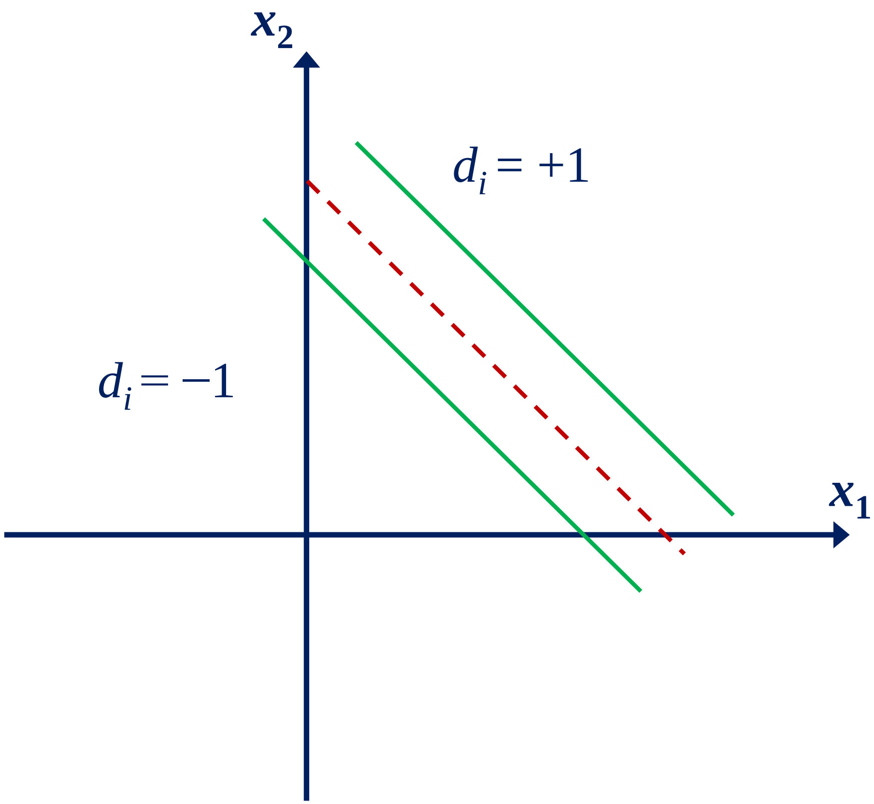
– برای داده‌هایی که برروی حاشیه جداسازی قرارمی‌گیرند

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

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

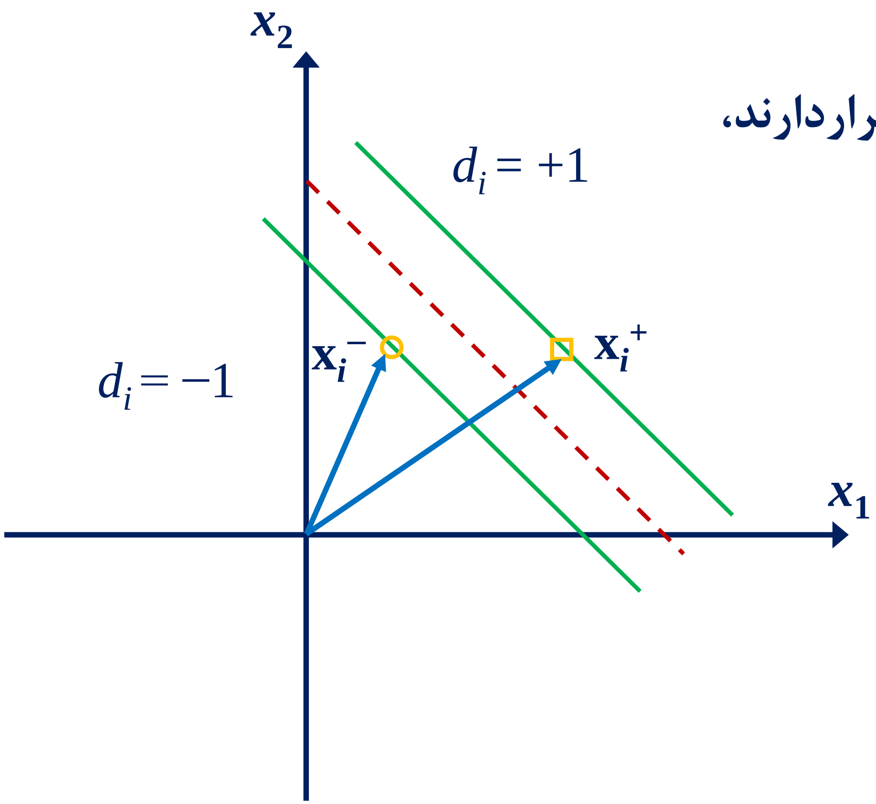
– به خاطر بیاورید که مساله بهینه‌سازی عبارت بود از کلاسه‌بندی الگوها به‌طوری که حاشیه جداسازی بیشینه شود.



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.



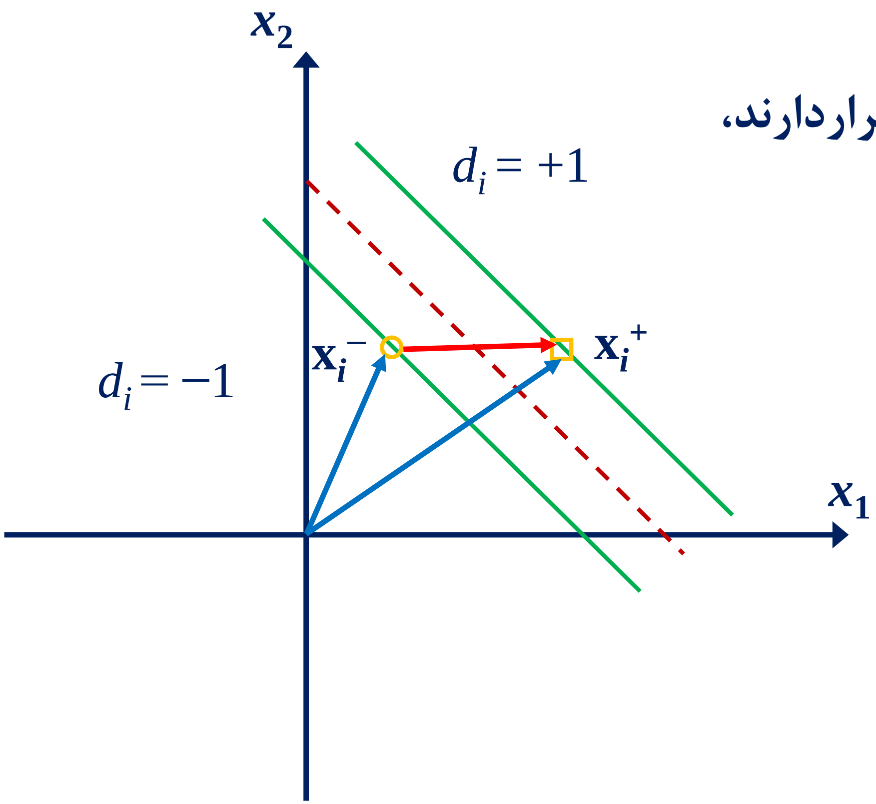
ماشین بردار پشتیبان (SVM)

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.

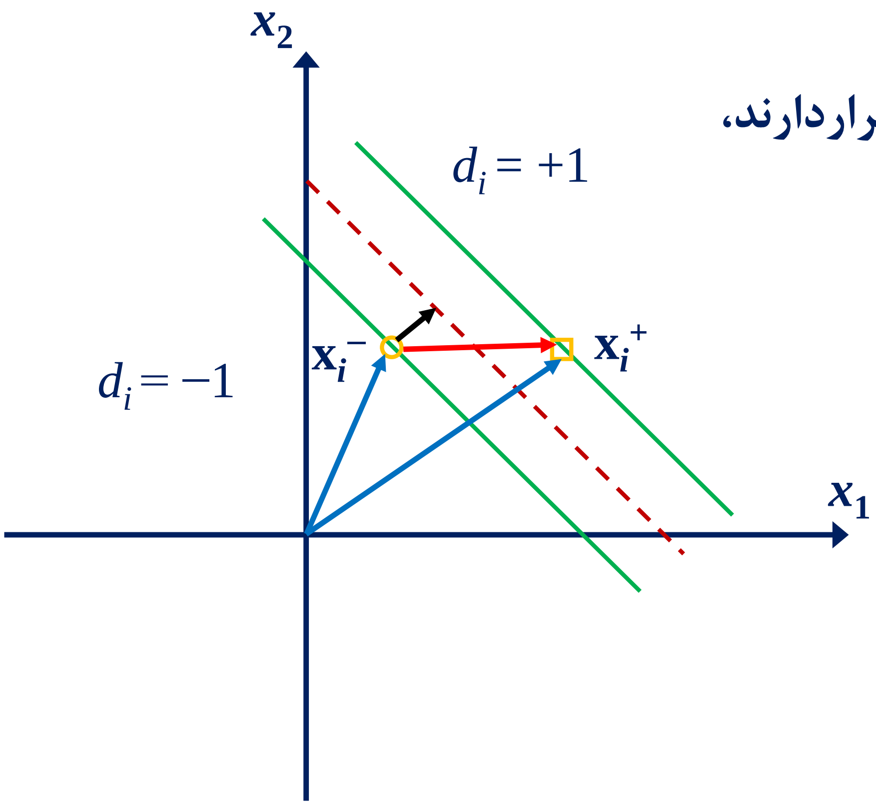
– تفاضل این دو بردار برابر است با $(\mathbf{x}_i^+ - \mathbf{x}_i^-)$



ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

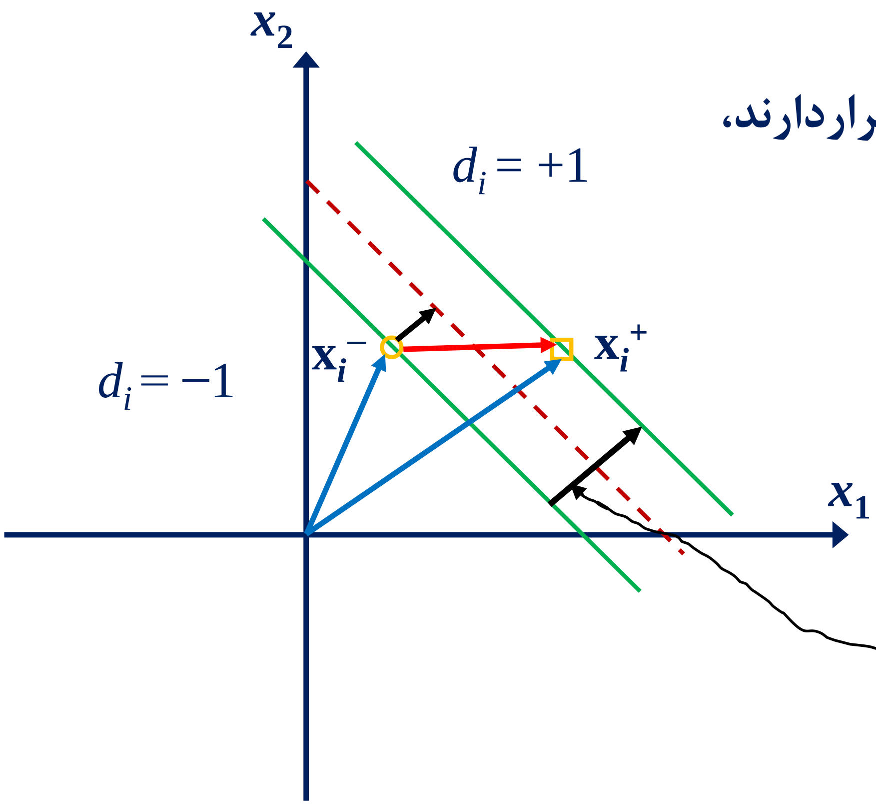
- به خاطر بیاورید که مساله بهینه‌سازی عبارت بود از کلاسه‌بندی الگوها به‌طوری که حاشیه جداسازی بیشینه شود.
- دو الگوی $\mathbf{x}_i^+$ و $\mathbf{x}_i^-$ که برروی حاشیه جداسازی قراردارند، را درنظربگیرید.
- تفاضل این دو بردار برابر است با $(\mathbf{x}_i^+ - \mathbf{x}_i^-)$
- چنانچه این تفاضل را در بردار 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:

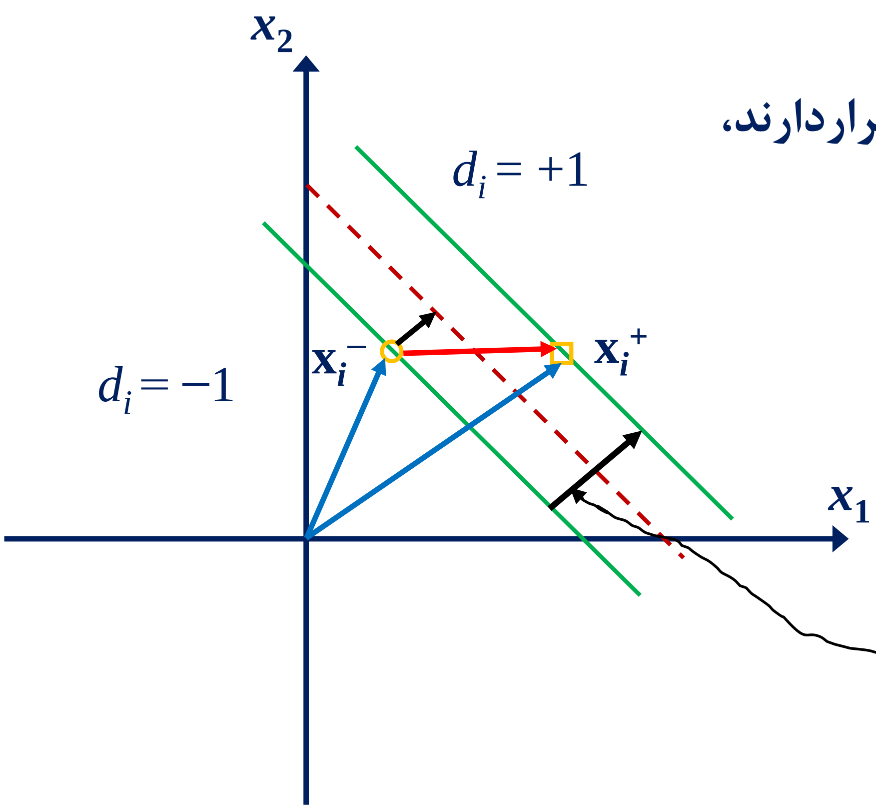


$$\text{width} = (\mathbf{x}_i^+ - \mathbf{x}_i^-)^T \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

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



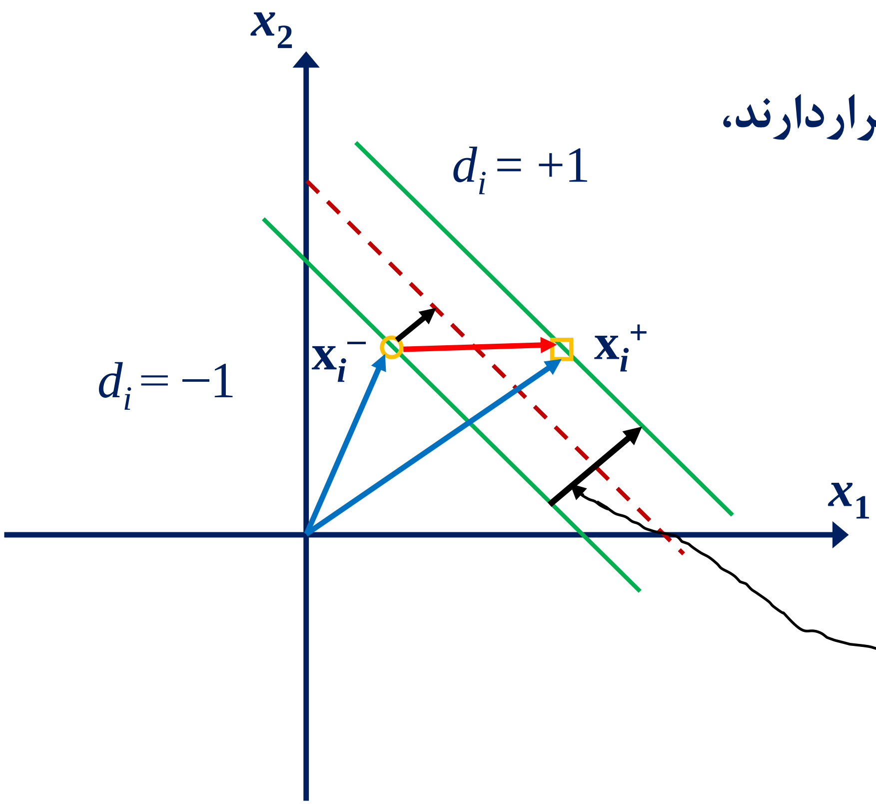
- دو الگوی $\mathbf{x}_i^+$ و $\mathbf{x}_i^-$ که برروی حاشیه جداسازی قراردارند، را درنظربگیرید.
- تفاضل این دو بردار برابر است با $(\mathbf{x}_i^+ - \mathbf{x}_i^-)$
- چنانچه این تفاضل را در بردار نرمال شده $\mathbf{w}$ ضرب داخلی‌کنیم، عرض این حاشیه جداسازی به‌دست می‌آید:

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

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

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:

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

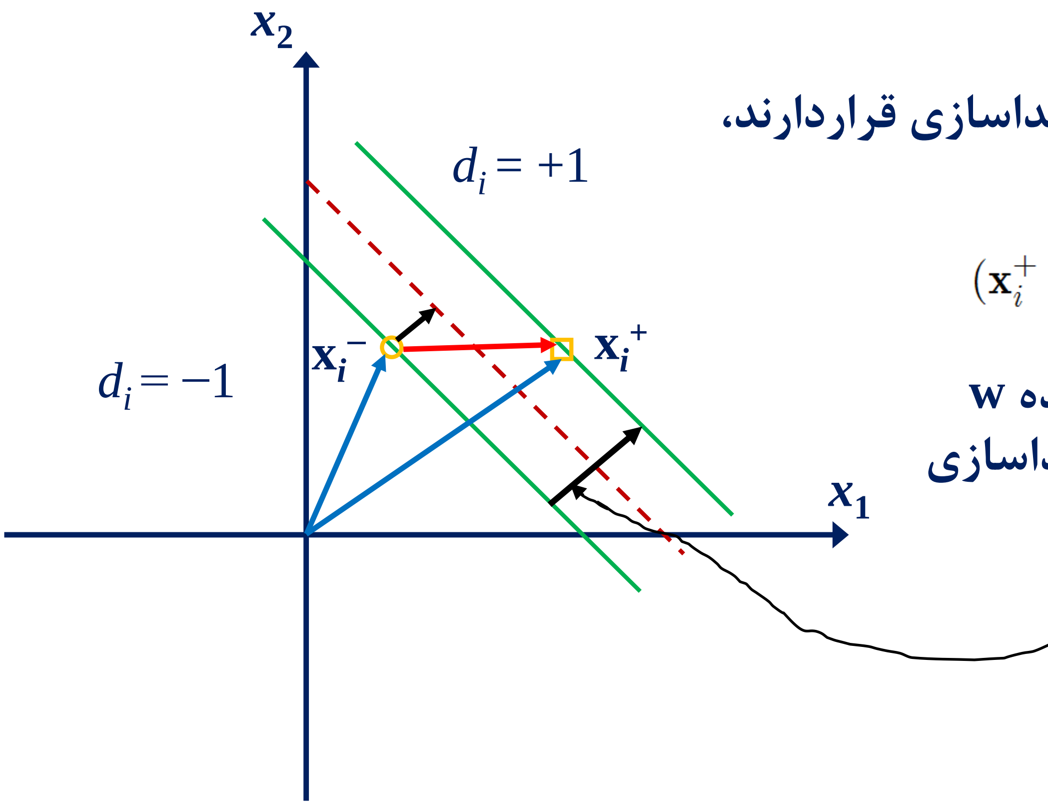
$$\text{width} = \frac{2}{\|\mathbf{w}\|}$$

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

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:



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

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

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

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

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

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

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

- 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

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

– 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:

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

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

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.

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.

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.

– 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:

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

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

– 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.

– 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

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

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:

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

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- برای بهینه‌سازی نسبت به ضرایب لاگرانژ از «مساله دوگان» (Dual Problem) استفاده می‌شود.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- برای بهینه‌سازی نسبت به ضرایب لاگرانژ از «مساله دوگان» (Dual Problem) استفاده می‌شود.

- مساله دوگان همان جواب بهینه مساله اولیه را دارد ولی در آن جواب ضرایب لاگرانژ نیز بهینه است.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- برای بهینه‌سازی نسبت به ضرایب لاگرانژ از «مساله دوگان» (Dual Problem) استفاده می‌شود.

- مساله دوگان همان جواب بهینه مساله اولیه را دارد ولی در آن جواب ضرایب لاگرانژ نیز بهینه است.

- قضیه دوگانگی (Duality Theorem) به شکل زیر بیان می‌شود:

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- برای بهینه‌سازی نسبت به ضرایب لاگرانژ از «مساله دوگان» (Dual Problem) استفاده می‌شود.

- مساله دوگان همان جواب بهینه مساله اولیه را دارد ولی در آن جواب ضرایب لاگرانژ نیز بهینه است.

- قضیه دوگانگی (Duality Theorem) به شکل زیر بیان می‌شود:

آ- اگر مساله اولیه دارای جواب بهینه باشد، مساله دوگان نیز جواب بهینه دارد و مقادیر بهینه نظیر برابراند.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- برای بهینه‌سازی نسبت به ضرایب لاگرانژ از «مساله دوگان» (Dual Problem) استفاده می‌شود.

- مساله دوگان همان جواب بهینه مساله اولیه را دارد ولی در آن جواب ضرایب لاگرانژ نیز بهینه است.

- قضیه دوگانگی (Duality Theorem) به شکل زیر بیان می‌شود:

آ- اگر مساله اولیه دارای جواب بهینه باشد، مساله دوگان نیز جواب بهینه دارد و مقادیر بهینه نظیر برابراند.

ب- به‌منظور این که $\mathbf{w}_o$ جواب بهینه اولیه باشد و α_i جواب بهینه دوگان، لازم و کافی است که $\mathbf{w}_o$ برای مساله اولیه شدنی (feasible) باشد و

$$\phi(\mathbf{w}_o) = J(\mathbf{w}_o, b_o, \alpha_o) = \min_{\mathbf{w}}(\mathbf{w}, b, \alpha)$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- برای یافتن α_i های بهینه، تابع هزینه بهینه‌شده برحسب $\mathbf{w}$ و b را درنظر بگیرید

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

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- برای یافتن α_i های بهینه، تابع هزینه بهینه‌شده برحسب $\mathbf{w}$ و b را درنظر بگیرید

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

- مساله دوگان:

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– برای یافتن α_i های بهینه، تابع هزینه بهینه‌شده برحسب $\mathbf{w}$ و b را درنظر بگیرید

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

– مساله دوگان:

با توجه به داده‌های $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$، مطلوب است یافتن ضرایب لاگرانژ $\{\alpha_i\}_{i=1}^{N}$ به‌طوری‌که تابع هزینه بالا را بیشینه کند، با توجه به قیود زیر:

$$\sum_{i=1}^{N} \alpha_i d_i = 0$$

$$\alpha_i \geq 0 \qquad i = 1, \ldots, N$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- برای یافتن α_i های بهینه، تابع هزینه بهینه‌شده برحسب $\mathbf{w}$ و b را درنظر بگیرید

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

- مساله دوگان:

با توجه به داده‌های $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$، مطلوب است یافتن ضرایب لاگرانژ $\{\alpha_i\}_{i=1}^{N}$ به‌طوری‌که تابع هزینه بالا را بیشینه کند، با توجه به قیود زیر:

$$\sum_{i=1}^{N} \alpha_i d_i = 0$$

$$\alpha_i \geq 0 \qquad i = 1, \ldots, N$$

- توجه کنید که برخلاف مساله بهینه‌سازی اولیه که بر مبنای تابع لاگرانژ بود، مساله دوگان کاملا برحسب داده‌های آموزش است.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– برای یافتن α_i های بهینه، تابع هزینه بهینه‌شده برحسب $\mathbf{w}$ و b را درنظر بگیرید

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

– مساله دوگان:

با توجه به داده‌های $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$، مطلوب است یافتن ضرایب لاگرانژ $\{\alpha_i\}_{i=1}^{N}$ به‌طوری‌که تابع هزینه بالا را بیشینه کند، با توجه به قیود زیر:

$$\sum_{i=1}^{N} \alpha_i d_i = 0$$

$$\alpha_i \geq 0 \qquad i = 1, \ldots, N$$

– توجه کنید که برخلاف مساله بهینه‌سازی اولیه که بر مبنای تابع لاگرانژ بود، مساله دوگان کاملا برحسب داده‌های آموزش است.

– علاوه بر آن، تابع هزینه که باید بیشینه شود، فقط وابسته به ضرب داخلی الگوهای ورودی است

$$\{\mathbf{x}_i^T \mathbf{x}_j\}_{i,j=1}^{N}$$

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

- نکته مهم:

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– **نکته مهم:**

بردارهای پشتیبان $\mathbf{x}^{(s)}$ زیرمجموعه ای از داده‌های آموزش هستند.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– نکته مهم:

بردارهای پشتیبان $\mathbf{x}^{(s)}$ زیرمجموعه ای از داده‌های آموزش هستند.

– بنابراین، با درنظرگرفتن قید $\alpha_i \geq 0$ ، فقط تعدادی از α_i ها غیرصفراند و بقیه صفراند.

یعنی این که بردار جواب، حالت پراکنده (sparse) دارد.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– نکته مهم:
بردارهای پشتیبان $\mathbf{x}^{(s)}$ زیرمجموعه ای از داده‌های آموزش هستند.

– بنابراین، با درنظرگرفتن قید $\alpha_i \geq 0$ ، فقط تعدادی از α_i ها غیرصفراند و بقیه صفراند.
یعنی این که بردار جواب، حالت پراکنده (sparse) دارد.

– با حل مساله بهینه‌سازی مقید، ضرایب بهینه لاگرانژ $(\alpha_{i,o})$ به‌دست می‌آید.

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– نکته مهم:
بردارهای پشتیبان $\mathbf{x}^{(s)}$ زیرمجموعه ای از داده‌های آموزش هستند.

– بنابراین، با درنظرگرفتن قید $\alpha_i \geq 0$ ، فقط تعدادی از α_i ها غیرصفراند و بقیه صفراند. یعنی این که بردار جواب، حالت پراکنده (sparse) دارد.

– با حل مساله بهینه‌سازی مقید، ضرایب بهینه لاگرانژ $(\alpha_{i,o})$ به‌دست می‌آید.

– در نتیجه، بردار وزن‌ها برابر خواهدشد با

$$\mathbf{w}_o = \sum_{i=1}^{Ns} \alpha_{i,o} d_i \mathbf{x}_i$$

Ns تعداد بردارهای پشتیبان (یعنی تعداد ضرایب غیرصفر لاگرانژ)

ماشین بردار پشتیبان (SVM)

۱- ابرصفحه بهینه برای الگوهای جداپذیر خطی

– نکته مهم:
بردارهای پشتیبان $\mathbf{x}^{(s)}$ زیرمجموعه ای از داده‌های آموزش هستند.

– بنابراین، با درنظرگرفتن قید $\alpha_i \geq 0$ ، فقط تعدادی از α_i ها غیرصفراند و بقیه صفراند.
یعنی این که بردار جواب، حالت پراکنده (sparse) دارد.

– با حل مساله بهینه‌سازی مقید، ضرایب بهینه لاگرانژ $(\alpha_{i,o})$ به‌دست می‌آید.

– در نتیجه، بردار وزن‌ها برابر خواهدشد با

$$\mathbf{w}_o = \sum_{i=1}^{Ns} \alpha_{i,o} d_i \mathbf{x}_i$$

Ns تعداد بردارهای پشتیبان (یعنی تعداد ضرایب غیرصفر لاگرانژ)

– برای محاسبه b_o ها:

$$\mathbf{w}_o^T \mathbf{x}^{(s)} + b_o = \pm 1 \quad \text{for} \quad d(s) = \pm 1$$

$$b_o = 1 - \sum_{i=1}^{Ns} \alpha_{i,o} d_i \mathbf{x}_i^T \mathbf{x}^{(s)}$$

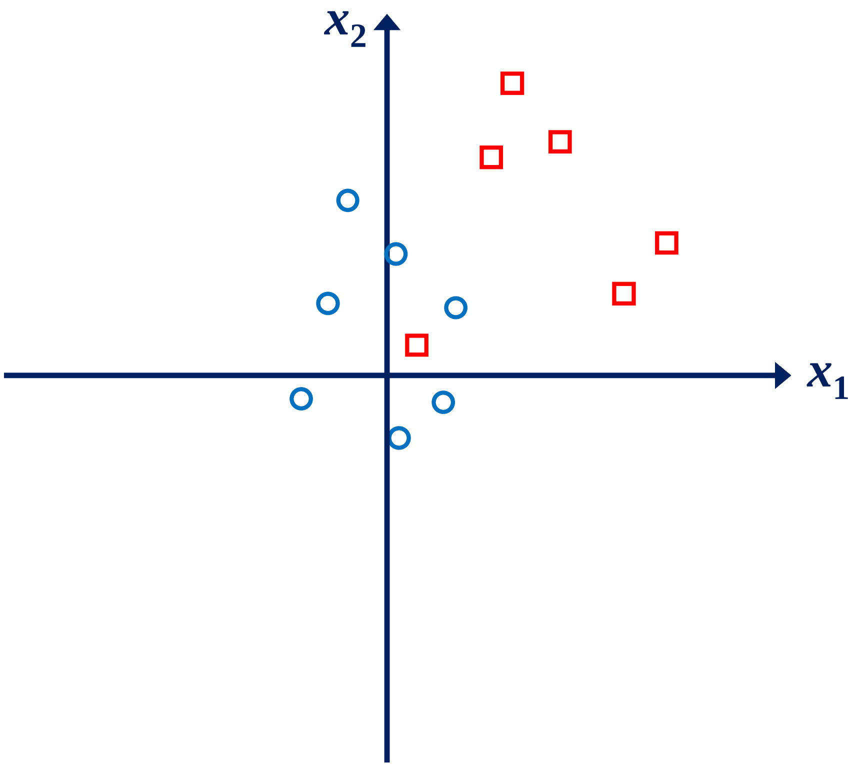
ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

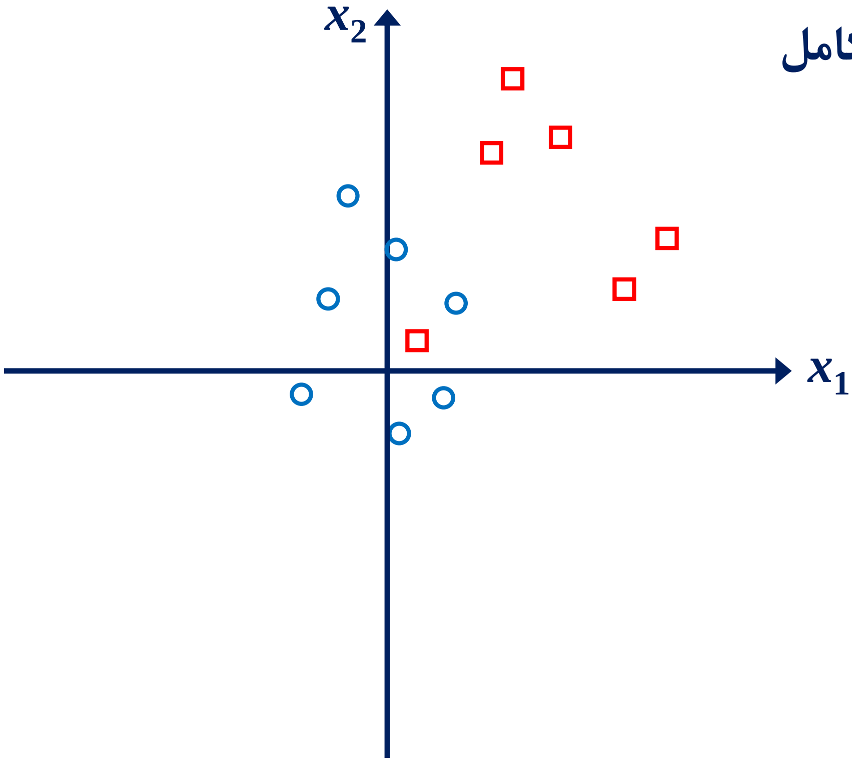
ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر



ماشین بردار پشتیبان (SVM)

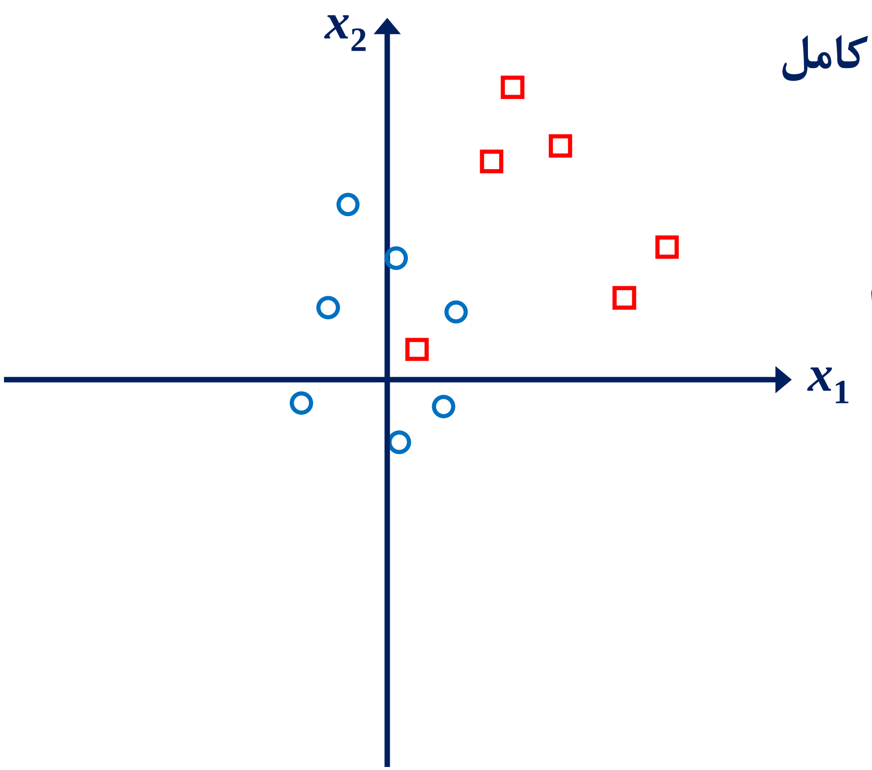
۲- ابرصفحه بهینه برای الگوهای جداناپذیر



- در این حالت، ابرصفحه نمی‌تواند الگوها را به‌طور کامل از یکدیگر جداکند و بنابراین، خطای کلاسه‌بندی خواهیم داشت.

ماشین بردار پشتیبان (SVM)

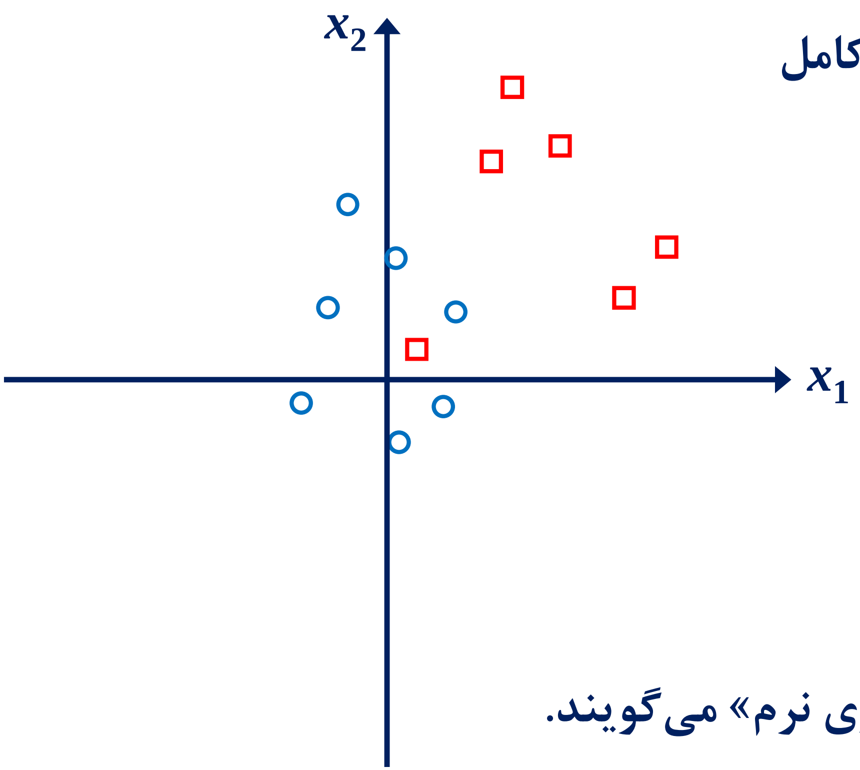
۲- ابرصفحه بهینه برای الگوهای جداناپذیر



- در این حالت، ابرصفحه نمی‌تواند الگوها را به‌طور کامل از یکدیگر جداکند و بنابراین، خطای کلاسه‌بندی خواهیم داشت.
- اگرچه مایلیم ابرصفحه‌ای بهینه‌ای به‌دست آوریم که خطای کلاسه‌بندی کمینه‌شود.

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر



- در این حالت، ابرصفحه نمی‌تواند الگوها را به‌طور کامل از یکدیگر جداکند و بنابراین، خطای کلاسه‌بندی خواهیم داشت.
- اگرچه مایلیم ابرصفحه‌ای بهینه‌ای به‌دست آوریم که خطای کلاسه‌بندی کمینه‌شود.
- در این جا، چنانچه حاشیه جداسازی شرط

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

را نقض کند، در این‌صورت به آن «حاشیه جداسازی نرم» می‌گویند.

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

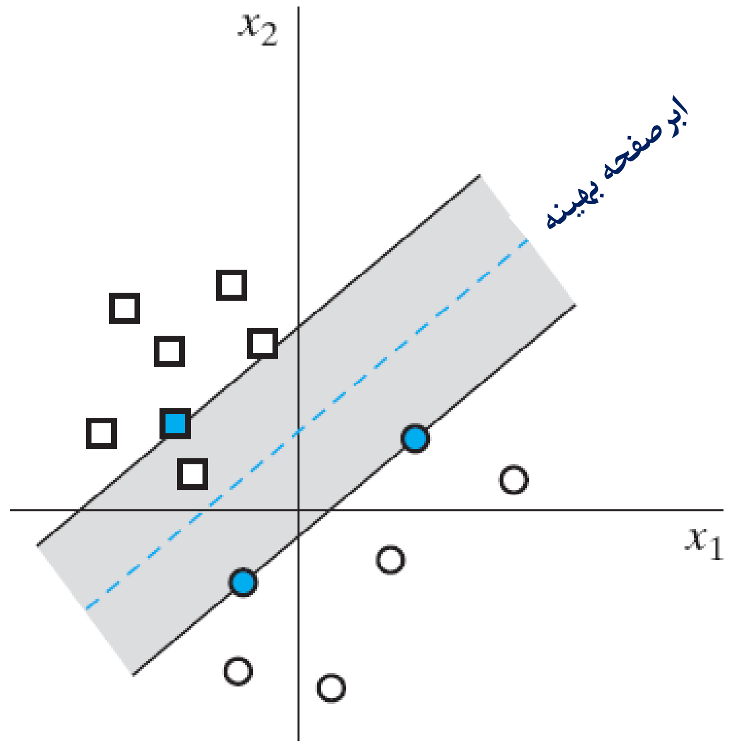
- این نقض‌کردن به دو صورت می‌تواند اتفاق بیافتد:

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- این نقض‌کردن به دو صورت می‌تواند اتفاق بیافتد:

۱- داده $(\mathbf{x}_i\,, d_i)$ داخل ناحیه جداسازی قرارگرفته ولی در سمت صحیح سطح تصمیم‌گیری است.



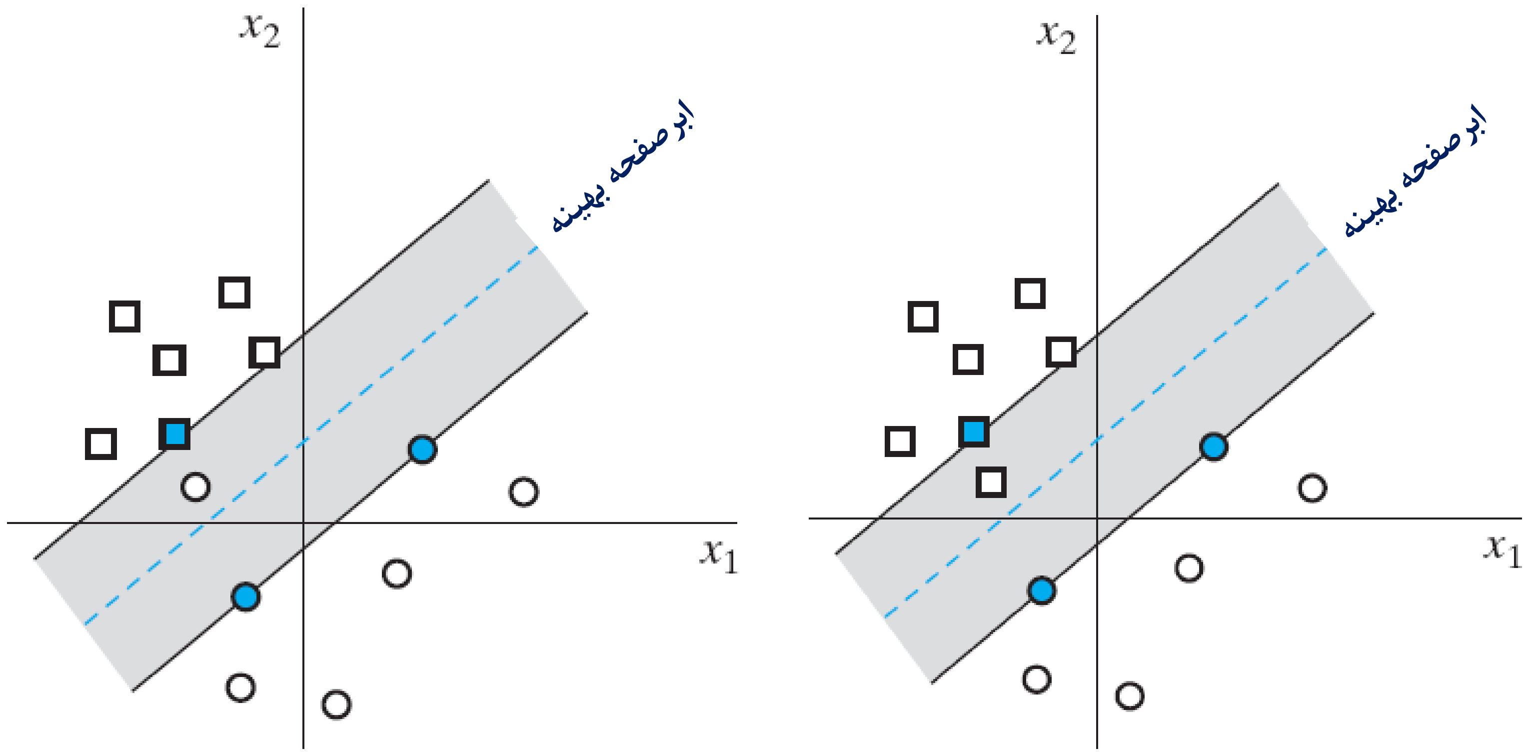
ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- این نقش‌کردن به دو صورت می‌تواند اتفاق بیافتد:

۱- داده $(\mathbf{x}_i\,, d_i)$ داخل ناحیه جداسازی قرارگرفته ولی در سمت صحیح سطح تصمیم‌گیری است.

۲- داده $(\mathbf{x}_i\,, d_i)$ در سمت غلط سطح تصمیم‌گیری قرارگرفته است.



ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- برای این حالت، متغیر غیرمنفی $\{\xi_i\}_{i=1}^{N}$ را به‌صورت زیر به سطح تصمیم‌گیری اضافه می‌کنیم:

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- برای این حالت، متغیر غیرمنفی $\{\xi_i\}_{i=1}^{N}$ را به‌صورت زیر به سطح تصمیم‌گیری اضافه می‌کنیم:

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

ξ_i متغیرهای شل (Slack Variables)

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- برای این حالت، متغیر غیرمنفی $\{\xi_i\}_{i=1}^{N}$ را به‌صورت زیر به سطح تصمیم‌گیری اضافه می‌کنیم:

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

ξ_i متغیرهای شل (Slack Variables)

$0 < \xi_i \leq 1$ داده داخل ناحیه جداسازی ولی در سمت صحیح سطح تصمیم گیری قراردارد

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- برای این حالت، متغیر غیرمنفی $\{\xi_i\}_{i=1}^{N}$ را به‌صورت زیر به سطح تصمیم‌گیری اضافه می‌کنیم:

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

ξ_i متغیرهای شل (Slack Variables)

$0 < \xi_i \leq 1$ داده داخل ناحیه جداسازی ولی در سمت صحیح سطح تصمیم گیری قراردارد

$\xi_i > 1$ داده در سمت غلط سطح تصمیم‌گیری قراردارد

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

– برای این حالت، متغیر غیرمنفی $\{\xi_i\}_{i=1}^{N}$ را به‌صورت زیر به سطح تصمیم‌گیری اضافه می‌کنیم:

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

ξ_i متغیرهای شل (Slack Variables)

$0 < \xi_i \leq 1$ داده داخل ناحیه جداسازی ولی در سمت صحیح سطح تصمیم گیری قراردارد

$\xi_i > 1$ داده در سمت غلط سطح تصمیم‌گیری قراردارد

– در این جا، بردارهای پشتیبان آن‌هایی هستند که نامعادله بالا را برای $\xi_i > 0$ برآورده کنند.

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

– برای این حالت، متغیر غیرمنفی $\{\xi_i\}_{i=1}^{N}$ را به‌صورت زیر به سطح تصمیم‌گیری اضافه می‌کنیم:

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

ξ_i متغیرهای شل (Slack Variables)

$0 < \xi_i \leq 1$ داده داخل ناحیه جداسازی ولی در سمت صحیح سطح تصمیم گیری قراردارد

$\xi_i > 1$ داده در سمت غلط سطح تصمیم‌گیری قراردارد

– در این جا، بردارهای پشتیبان آن‌هایی هستند که نامعادله بالا را برای $\xi_i > 0$ برآورده کنند.

– بنابراین، بردارهای پشتیبان دقیقا به همان شکل حالت جداپذیر خطی تعریف و تعیین می‌شوند.

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- هدف یافتن ابرصفحه جداکننده‌ای است که خطای کلاسه‌بندی را برای داده‌های آموزش کمینه‌کند:

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- هدف یافتن ابرصفحه جداکننده‌ای است که خطای کلاسه‌بندی را برای داده‌های آموزش کمینه‌کند:

$$\phi(\xi) = \sum_{i=1}^{N} I(\xi_i - 1)$$

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

– هدف یافتن ابرصفحه جداکننده‌ای است که خطای کلاسه‌بندی را برای داده‌های آموزش کمینه‌کند:

$$\phi(\xi) = \sum_{i=1}^{N} I(\xi_i - 1)$$

– این بهنیه‌سازی با توجه به بردار وزن w، قید

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

و قید برروی $\|\mathbf{w}\|^2$ انجام می‌شود و

$$I(\xi) = \begin{cases} 0 & \text{if} \quad \xi \leq 0 \\ 1 & \text{if} \quad \xi > 0 \end{cases}$$

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- هدف یافتن ابرصفحه جداکننده‌ای است که خطای کلاسه‌بندی را برای داده‌های آموزش کمینه‌کند:

$$\phi(\xi) = \sum_{i=1}^{N} I(\xi_i - 1)$$

- این بهنیه‌سازی با توجه به بردار وزن w، قید

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

و قید برروی $\|\mathbf{w}\|^2$ انجام می‌شود و

$$I(\xi) = \begin{cases} 0 & \text{if} \quad \xi \leq 0 \\ 1 & \text{if} \quad \xi > 0 \end{cases}$$

- متاسفانه این کمینه‌سازی با درنظرگرفتن w، مساله بهینه‌سازی نامحدب است.

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- برای رفع این مشکل، تابع هزینه را به‌صورت زیر درنظر می‌گیریم:

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- برای رفع این مشکل، تابع هزینه را به‌صورت زیر درنظر می‌گیریم:

$$\phi(\xi) = \sum_{i=1}^{N} \xi_i$$

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- برای رفع این مشکل، تابع هزینه را به‌صورت زیر درنظر می‌گیریم:

$$\phi(\xi) = \sum_{i=1}^{N} \xi_i$$

علاوه بر آن، برای بهینه‌سازی $\mathbf{w}$

$$\phi(\mathbf{w}, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N} \xi_i$$

C پارامتر متقابل تنظیم‌کننده
(Reciprocal Regularization Parameter)

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- برای رفع این مشکل، تابع هزینه را به‌صورت زیر درنظر می‌گیریم:

$$\phi(\xi) = \sum_{i=1}^{N} \xi_i$$

علاوه بر آن، برای بهینه‌سازی $\mathbf{w}$

$$\phi(\mathbf{w}, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N} \xi_i$$

C پارامتر متقابل تنظیم‌کننده
(Reciprocal Regularization Parameter)

- C بزرگ ≡ طراح SVM اعتماد زیادی به داده‌های آموزش دارد.

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- برای رفع این مشکل، تابع هزینه را به‌صورت زیر درنظر می‌گیریم:

$$\phi(\xi) = \sum_{i=1}^{N} \xi_i$$

علاوه بر آن، برای بهینه‌سازی $\mathbf{w}$

$$\phi(\mathbf{w}, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N} \xi_i$$

C پارامتر متقابل تنظیم‌کننده
(Reciprocal Regularization Parameter)

- C بزرگ ≡ طراح SVM اعتماد زیادی به داده‌های آموزش دارد.
- C کوچک ≡ داده های آموزش نویزی است و اعتماد کمی باید به آن داشت.

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

– برای رفع این مشکل، تابع هزینه را به‌صورت زیر درنظر می‌گیریم:

$$\phi(\xi) = \sum_{i=1}^{N} \xi_i$$

علاوه بر آن، برای بهینه‌سازی $\mathbf{w}$

$$\phi(\mathbf{w}, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N} \xi_i$$

C پارامتر متقابل تنظیم‌کننده
(Reciprocal Regularization Parameter)

- C بزرگ ≡ طراح SVM اعتماد زیادی به داده‌های آموزش دارد.
- C کوچک ≡ داده های آموزش نویزی است و اعتماد کمی باید به آن داشت.

– پارامتر C مصالحه‌ای بین پیچیدگی ماشین و تعداد نقاط جداناپذیر می‌کند.

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- مساله اولیه برای الگوهای جداناپذیر:

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- مساله اولیه برای الگوهای جداناپذیر:

با توجه به داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$ ، مطلوب است یافتن مقدار بهینه $\mathbf{w}$ و b به‌طوری که قیود زیررا برآورده کرده:

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

$$\xi_i > 0 \quad \forall i$$

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- مساله اولیه برای الگوهای جداناپذیر:

با توجه به داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$ ، مطلوب است یافتن مقدار بهینه $\mathbf{w}$ و b به‌طوری که قیود زیررا برآورده کرده:

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

و به‌طوری که بردار $\mathbf{w}$ و متغیرهای شُل تابع هزینه زیر را کمینه‌کنند:

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

که در آن $C > 0$ توسط طراح تعیین می‌شود.

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- مساله دوگان برای الگوهای جداناپذیر:

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- مساله دوگان برای الگوهای جداناپذیر:

با توجه به داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$ ، مطلوب است یافتن ضرایب لاگرانژ $\{\alpha_i\}_{i=1}^{N}$ که تابع هزینه زیر را بیشینه‌کند:

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

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

– مساله دوگان برای الگوهای جداناپذیر:

با توجه به داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$ ، مطلوب است یافتن ضرایب لاگرانژ $\{\alpha_i\}_{i=1}^{N}$ که تابع هزینه زیر را بیشینه‌کند:

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

با درنظرگرفتن قیود

$$\sum_{i=1}^{N} \alpha_i d_i = 0$$

$$0 \le \alpha_i \le C$$

که در آن $C > 0$ توسط طراحی تعیین می‌شود.

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- مساله دوگان برای الگوهای جداناپذیر:

با توجه به داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^N$ ، مطلوب است یافتن ضرایب لاگرانژ $\{\alpha_i\}_{i=1}^N$ که تابع هزینه زیر را بیشینه‌کند:

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

با درنظرگرفتن قیود

$$\sum_{i=1}^N \alpha_i d_i = 0$$

$$0 \le \alpha_i \le C$$

که در آن $C > 0$ توسط طراحی تعیین می‌شود.

- توجه کنید:

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- مساله دوگان برای الگوهای جداناپذیر:

با توجه به داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^N$ ، مطلوب است یافتن ضرایب لاگرانژ $\{\alpha_i\}_{i=1}^N$ که تابع هزینه زیر را بیشینه‌کند:

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

با درنظرگرفتن قیود

$$\sum_{i=1}^N \alpha_i d_i = 0$$

$$0 \le \alpha_i \le C$$

که در آن $C > 0$ توسط طراحی تعیین می‌شود.

- توجه کنید:

۱- نه ξ_iها در مساله دوگان ظاهرمی‌شوند و نه ضرایب لاگرانژ آن‌ها

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

- مساله دوگان برای الگوهای جداناپذیر:

با توجه به داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^N$ ، مطلوب است یافتن ضرایب لاگرانژ $\{\alpha_i\}_{i=1}^N$ که تابع هزینه زیر را بیشینه‌کند:

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

با درنظرگرفتن قیود

$$\sum_{i=1}^N \alpha_i d_i = 0$$

$$0 \le \alpha_i \le C$$

که در آن $C > 0$ توسط طراحی تعیین می‌شود.

- توجه کنید:

۱- نه ξ_iها در مساله دوگان ظاهرمی‌شوند و نه ضرایب لاگرانژ آن‌ها

۲- مساله دوگان برای الگوهای جداناپذیر همانند حالت جداپذیر است به غیر از $0 \le \alpha_i \le C$

ماشین بردار پشتیبان (SVM)

۲- ابرصفحه بهینه برای الگوهای جداناپذیر

– مساله دوگان برای الگوهای جداناپذیر:

با توجه به داده‌های موجود $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$ ، مطلوب است یافتن ضرایب لاگرانژ $\{\alpha_i\}_{i=1}^{N}$ که تابع هزینه زیر را بیشینه‌کند:

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

با درنظرگرفتن قیود

$$\sum_{i=1}^{N} \alpha_i d_i = 0$$

$$0 \leq \alpha_i \leq C$$

که در آن $C > 0$ توسط طراحی تعیین می‌شود.

– توجه کنید:

۱- نه ξ_iها در مساله دوگان ظاهرمی‌شوند و نه ضرایب لاگرانژ آن‌ها

۲- مساله دوگان برای الگوهای جداناپذیر همانند حالت جداپذیر است به غیر از $0 \leq \alpha_i \leq C$

۳- بردارهای پشتیبان همانند حالت قبل تعیین می‌شوند.

ماشین بردار پشتیبان (SVM)

– فلسفه SVM برای کلاسه‌بندی الگوها

ماشین بردار پشتیبان (SVM)

– فلسفه SVM برای کلاسه‌بندی الگوها

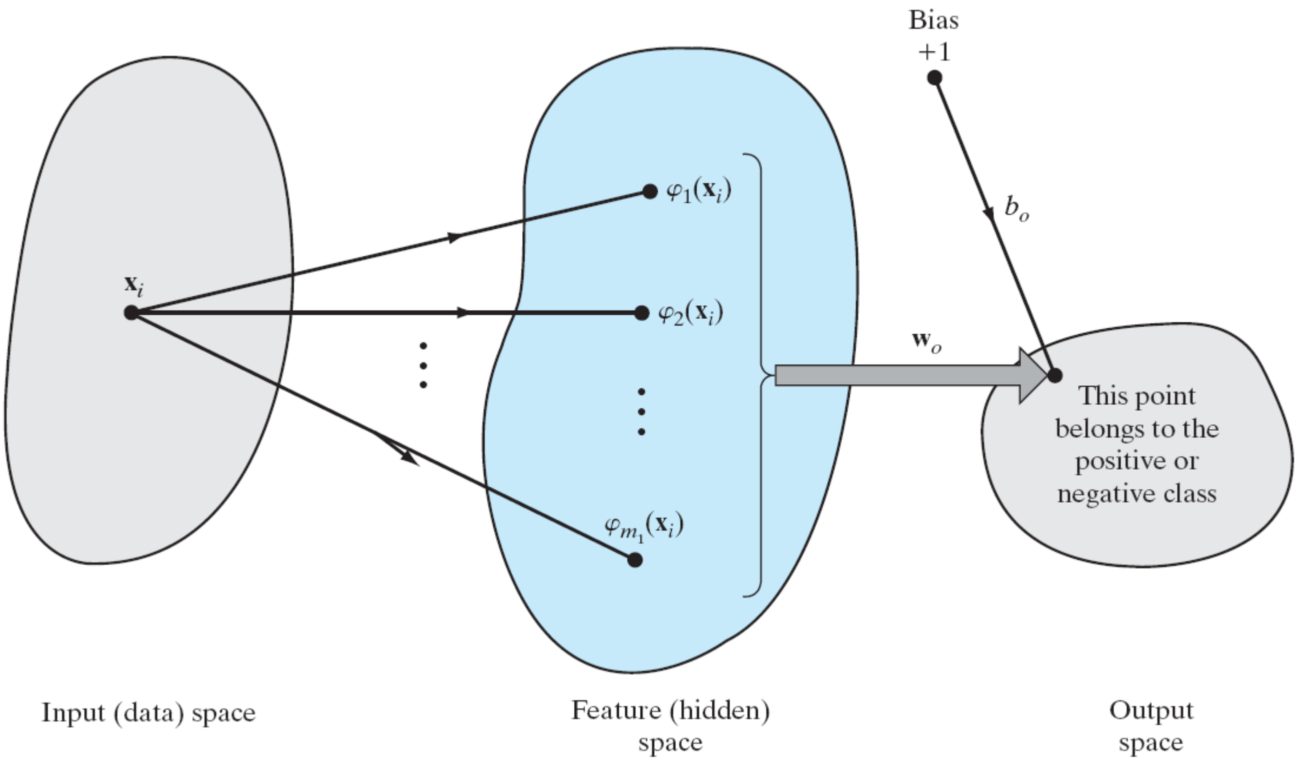
– اصول SVM

ماشین بردار پشتیبان (SVM)

– فلسفه SVM برای کلاسه‌بندی الگوها

– اصول SVM

۱– نگاشت غیرخطی بردار ورودی به فضای ویژگی‌ها



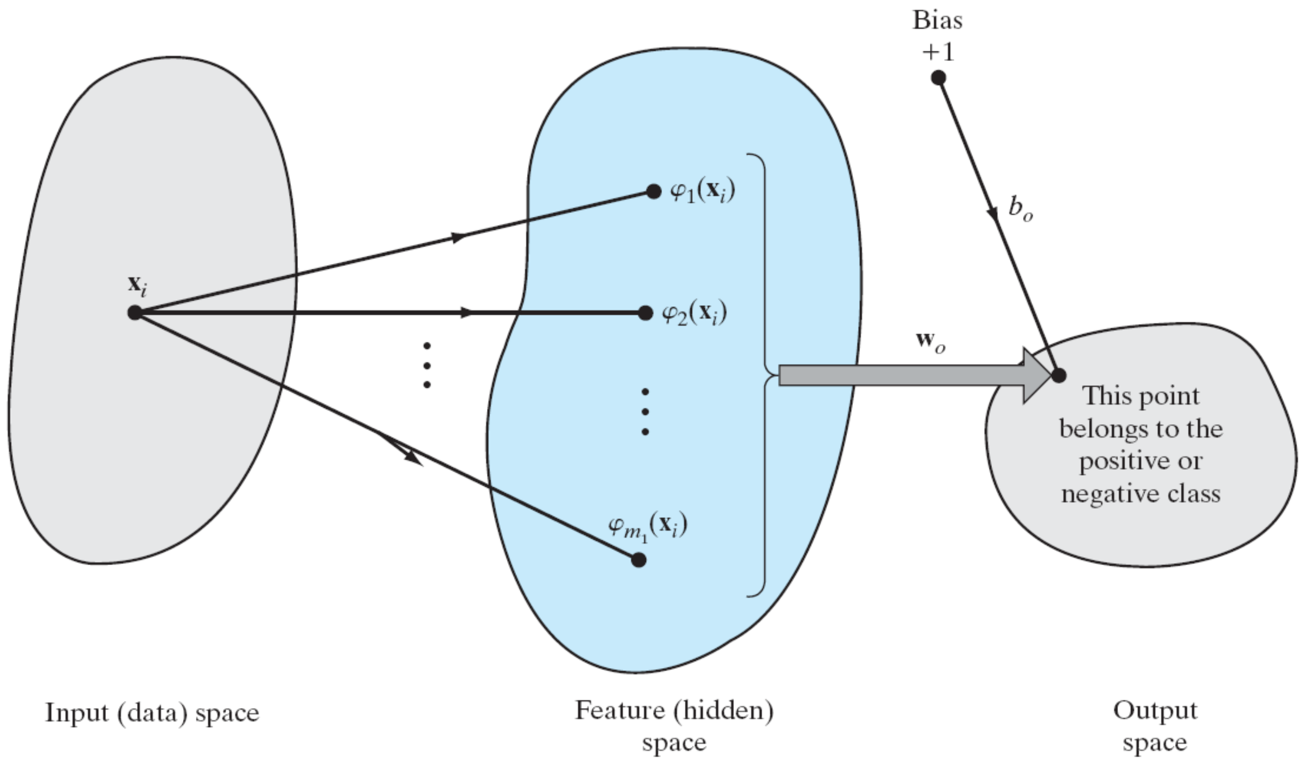
ماشین بردار پشتیبان (SVM)

– فلسفه SVM برای کلاسه‌بندی الگوها

– اصول SVM

۱– نگاشت غیرخطی بردار ورودی به فضای ویژگی‌ها

۲– تشکیل ابرصفحه بهینه برای جداسازی ویژگی‌ها که در قسمت ۱ کشف شدند.

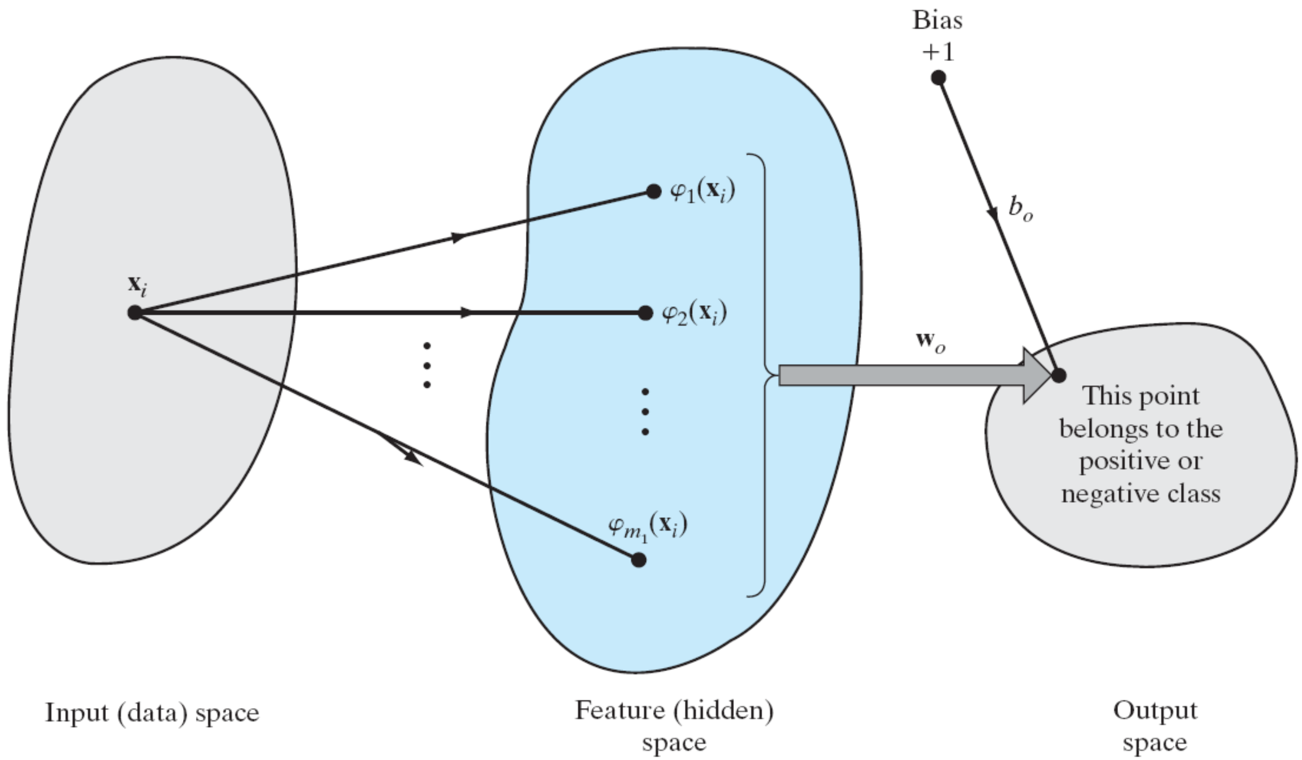


ماشین بردار پشتیبان (SVM)

- فلسفه SVM برای کلاسه‌بندی الگوها
- اصول SVM

۱- نگاشت غیرخطی بردار ورودی به فضای ویژگی‌ها

۲- تشکیل ابرصفحه بهینه برای جداسازی ویژگی‌ها که در قسمت ۱ کشف شدند.



نکته مهم: نظریه SVM، به‌طور تحلیلی ابعاد بهینه برای فضای پنهان پیدامی‌کند.