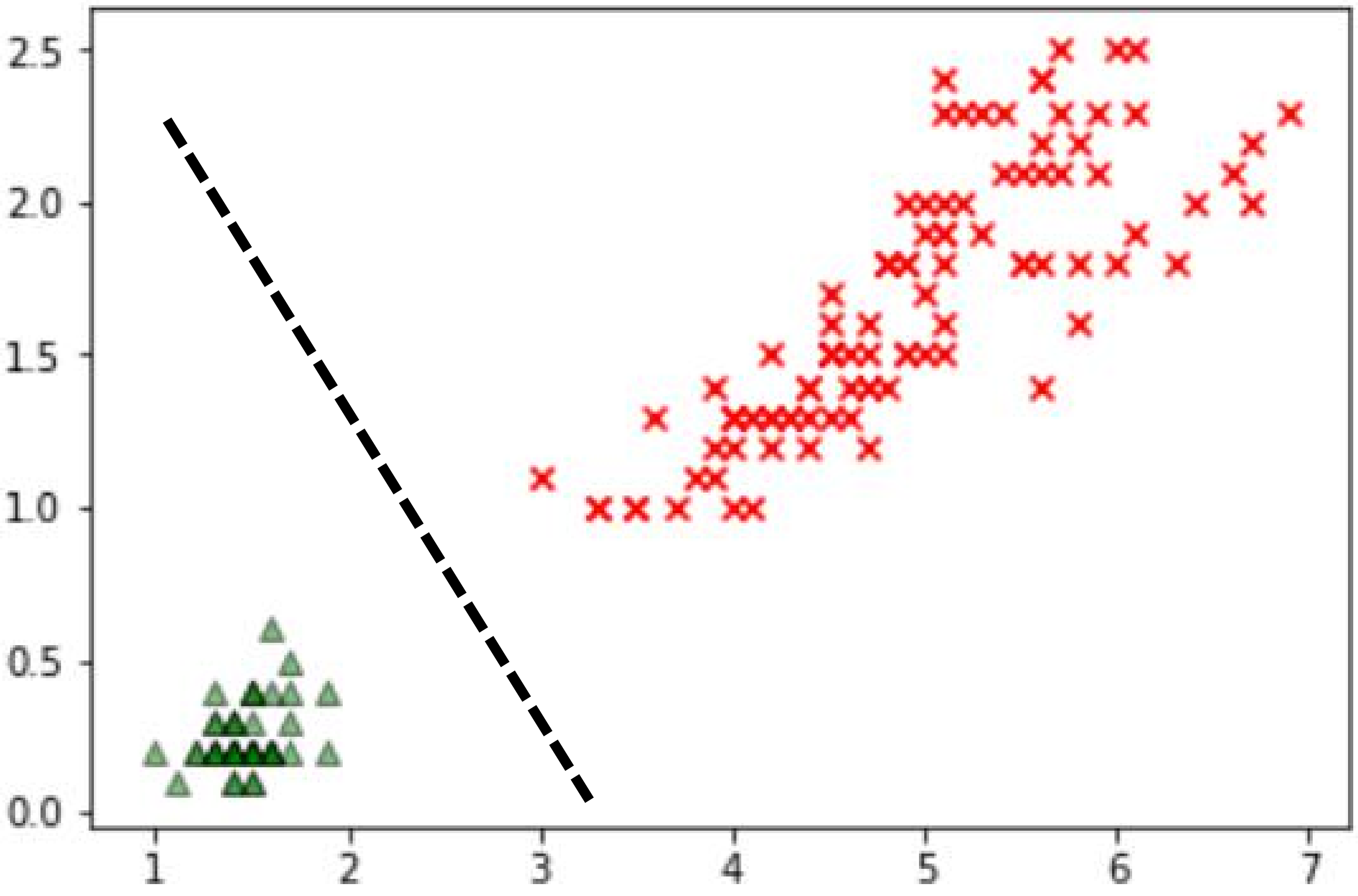
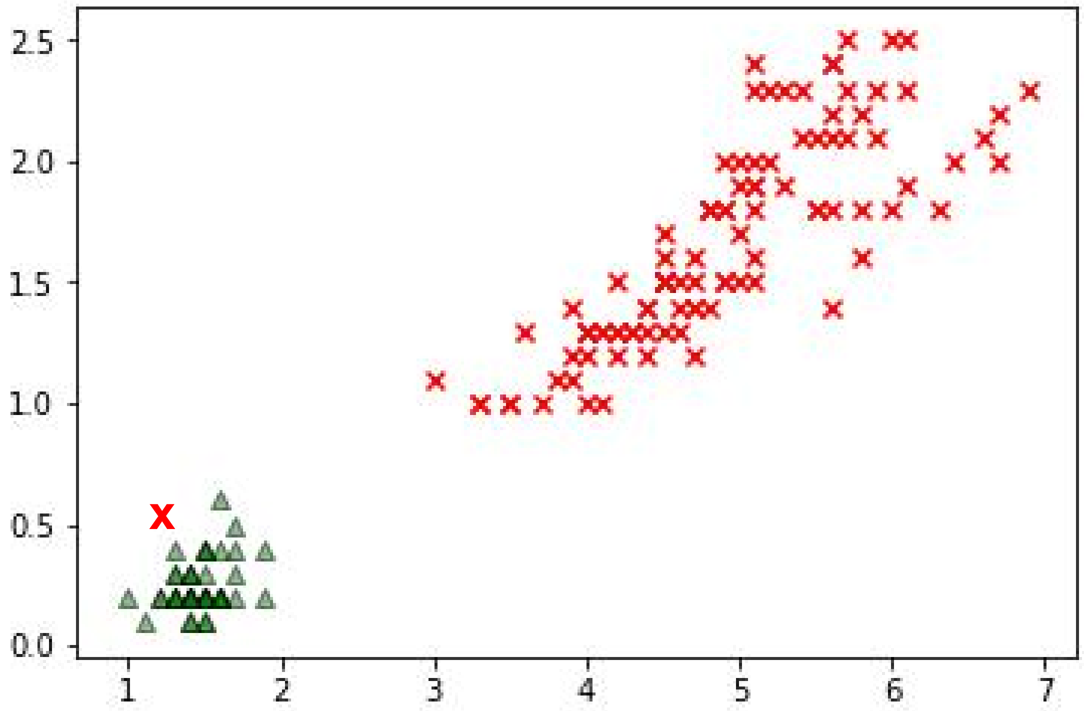
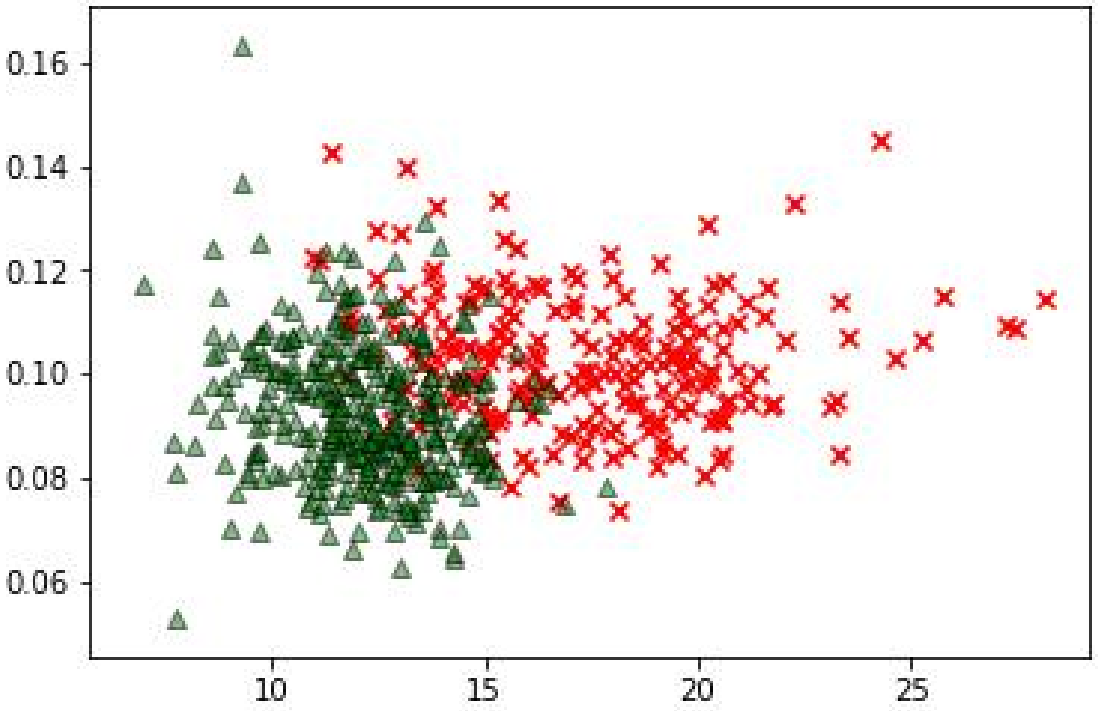
支持向量机 （Supporting Vector Machine）

线性可分



线性不可分

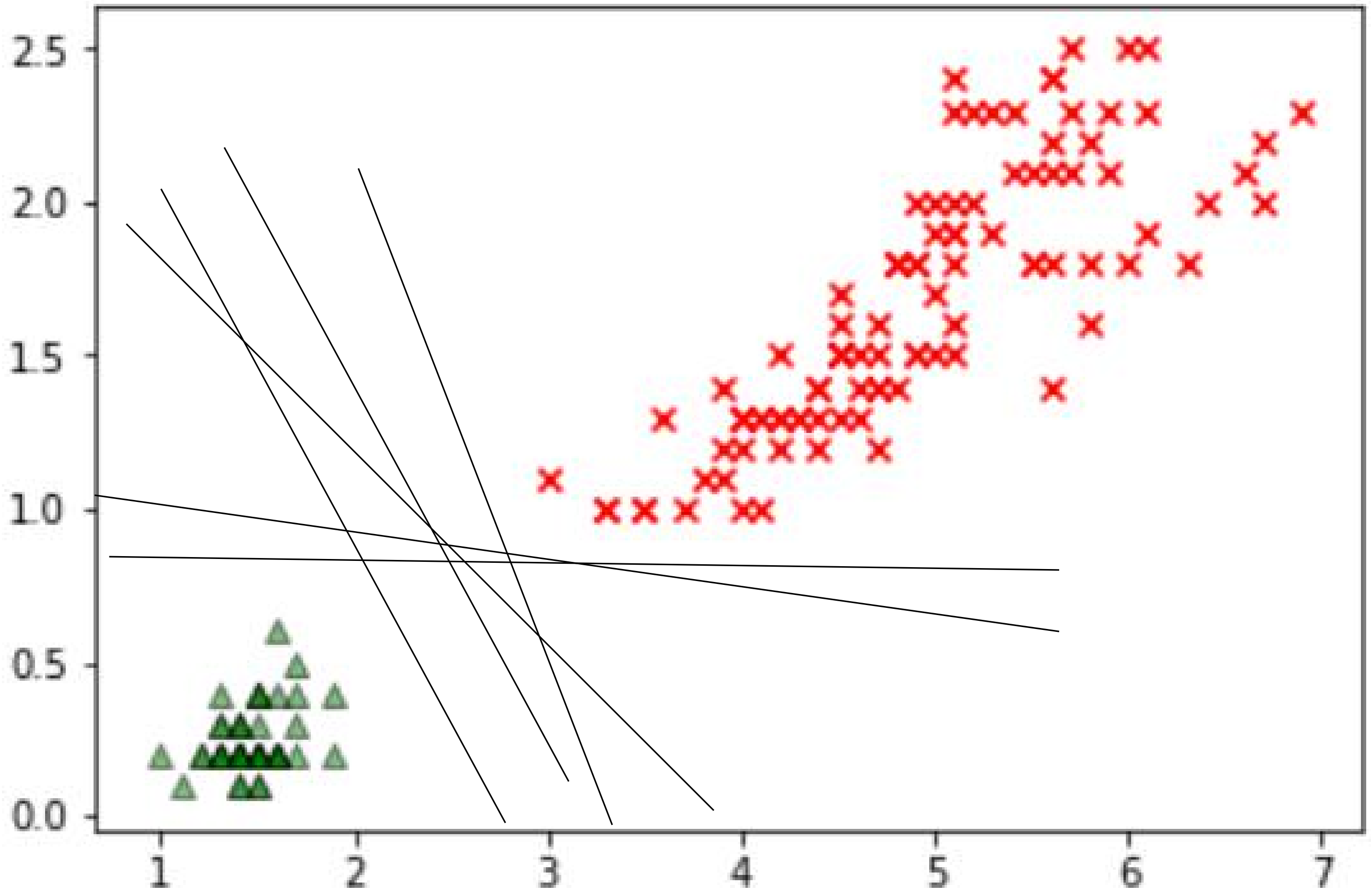




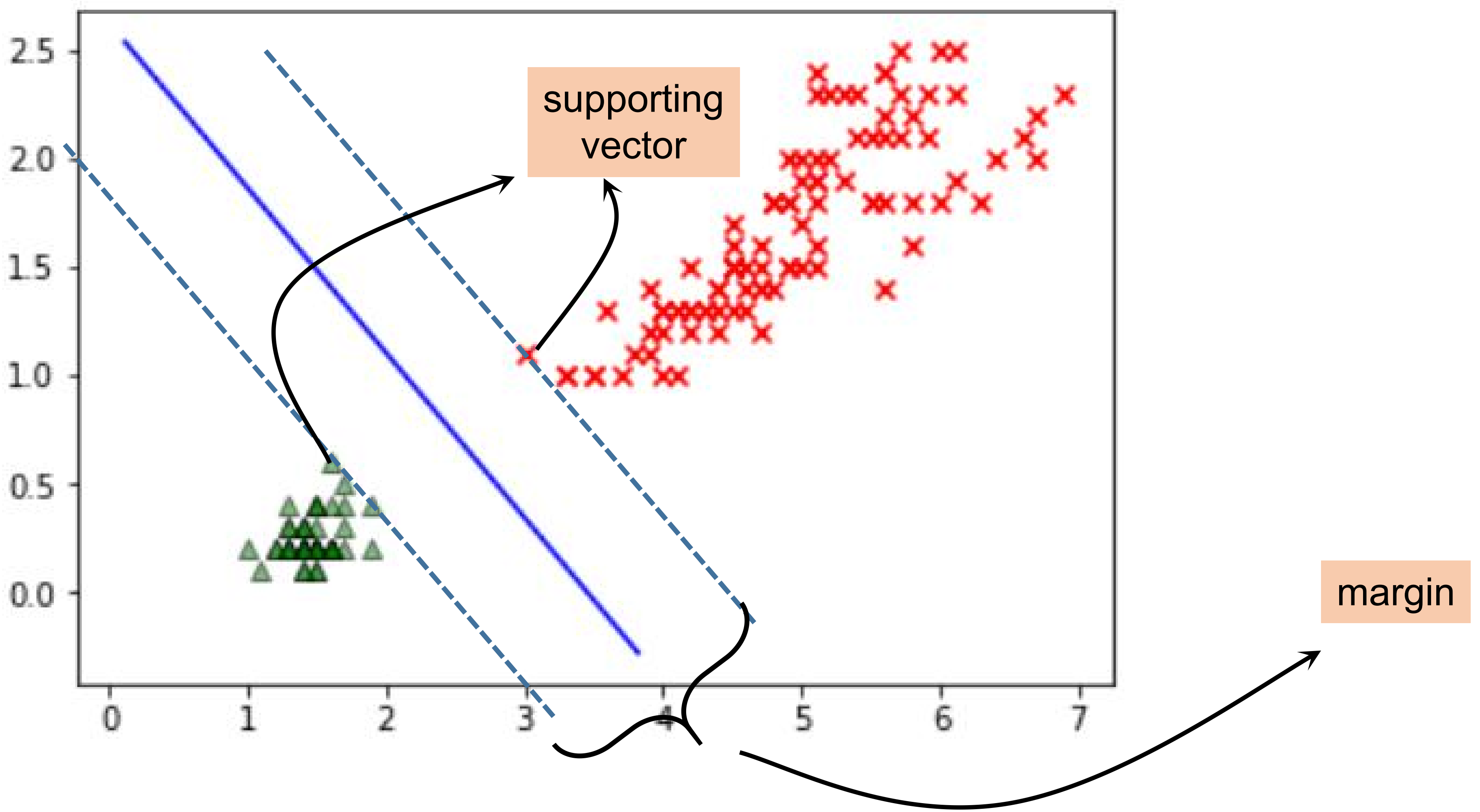
硬分类器 vs 软分类器

硬分类器	软分类器
将所有点均正确分类 仅能用于线性可分问题 不可用于线性不可分问题	允许一些点分类错误 可用于线性不可分问题

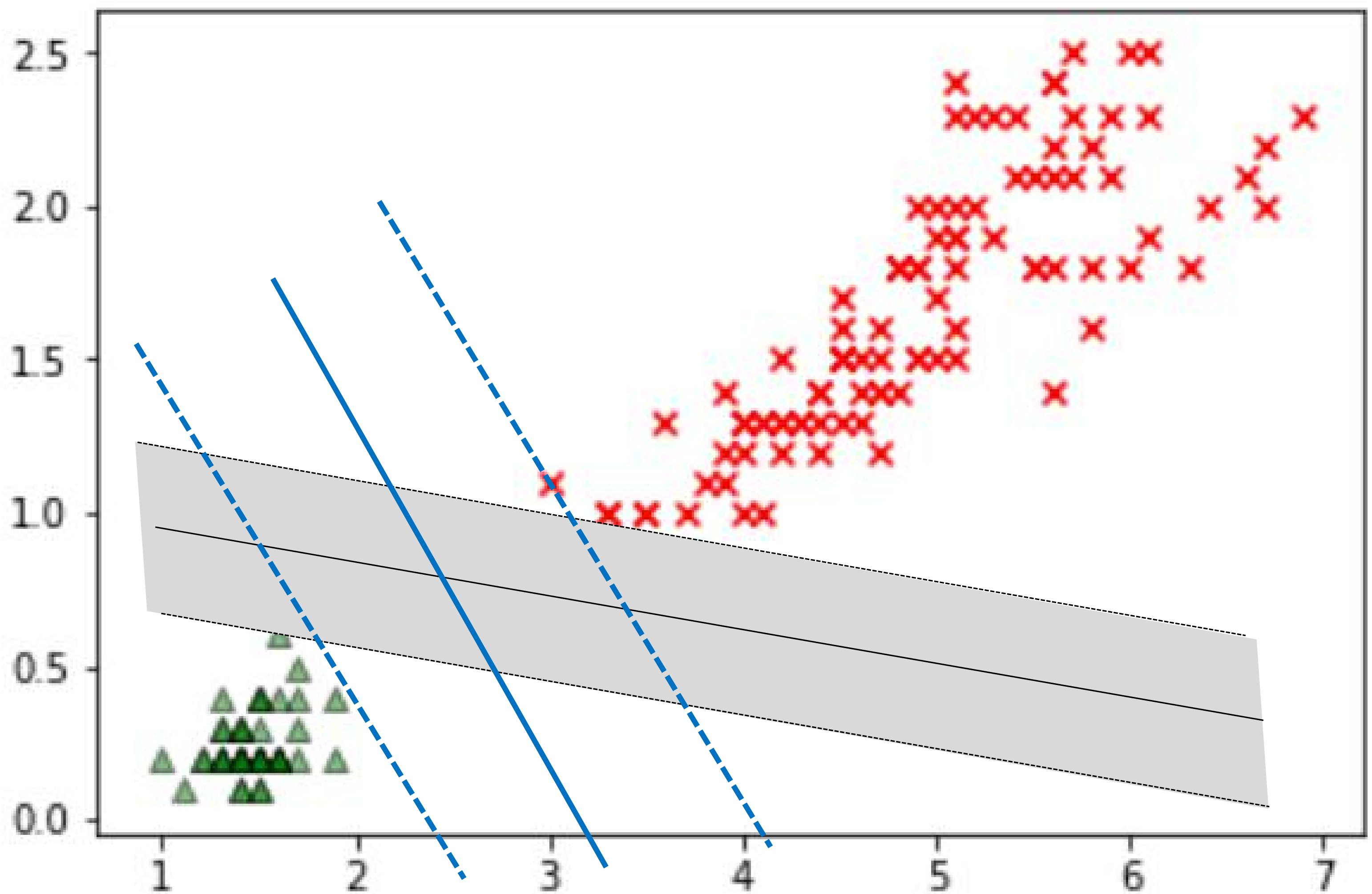
最优线性分类器?



Margin



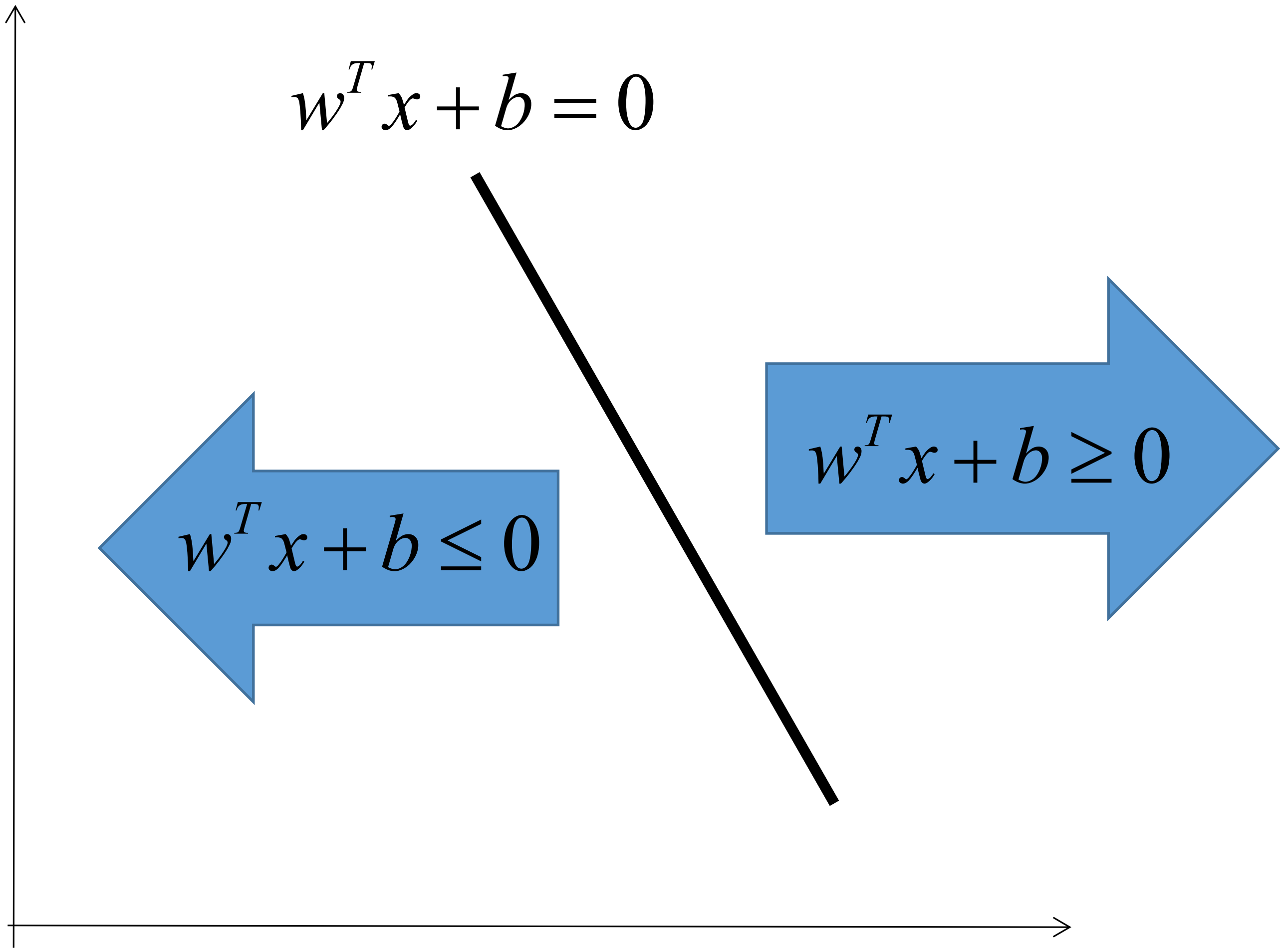
Large Margin Classification



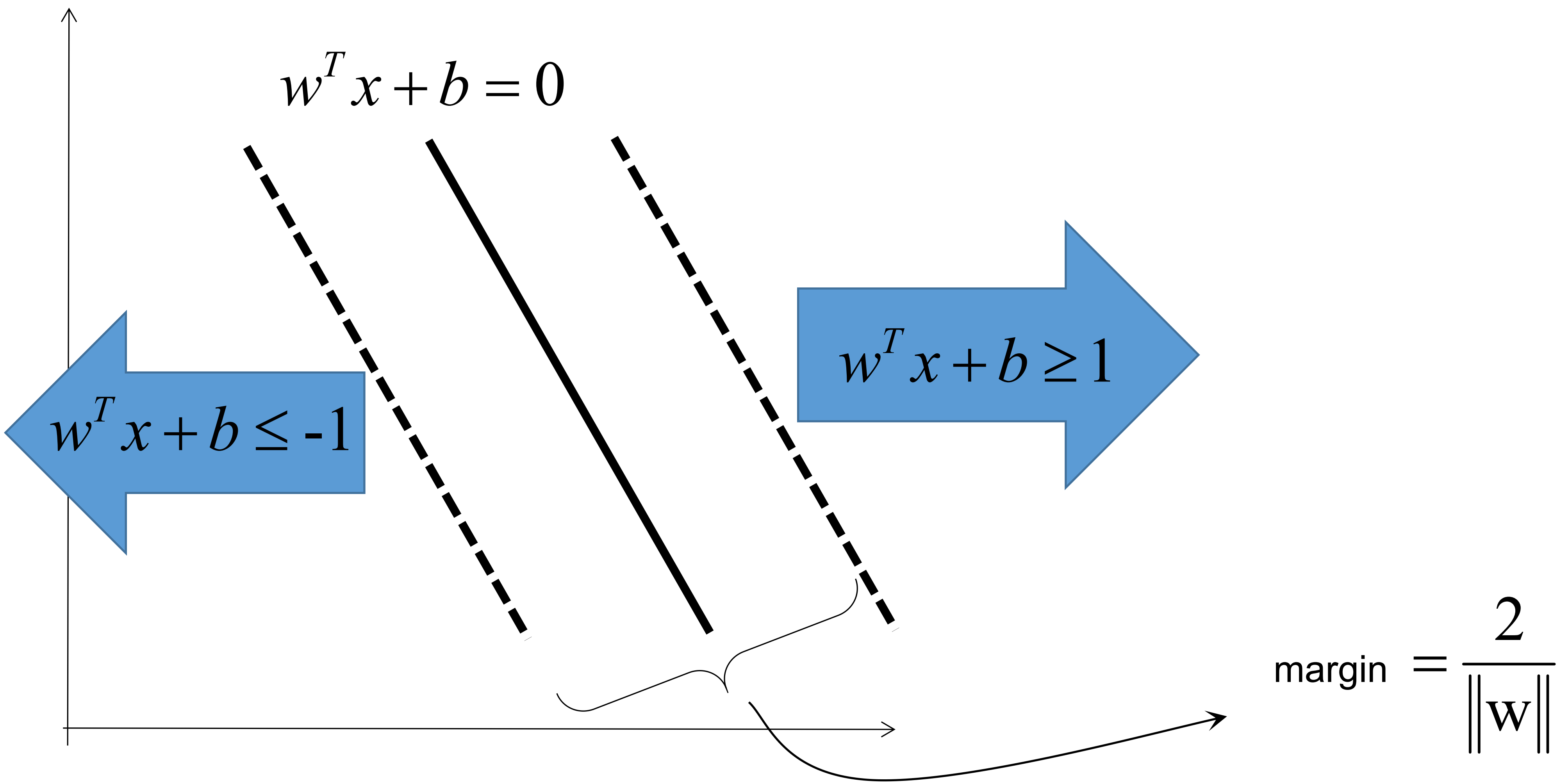
Large Margin Classification

Hard Margin Classification

Linear Classification



SVM Classification



数学表示

Hard Classification

$$\begin{cases} w^T x_i + b \ge 1 & (y_i = +1) \\ w^T x_i + b \le -1 & (y_i = -1) \end{cases} \Rightarrow y_i\left(w^T x_i + b\right) \ge 1$$

SVM

min (w,b) $\frac{1}{2}\|w\|^2$

s.t. $y_i\left(w^T x_i + b\right) \ge 1,$ i = 1,2,...,m

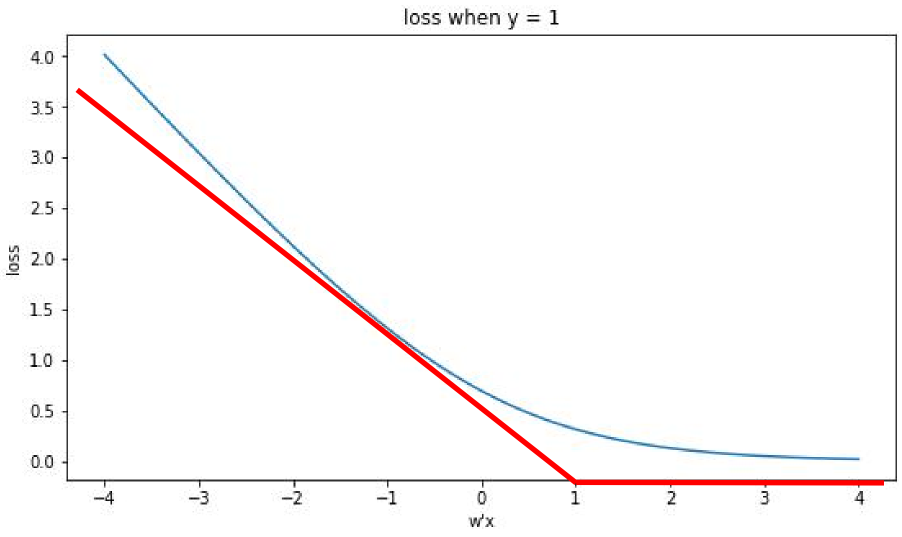
From LogisticRegression to SVM

Objective function of logistic regression with L2 penalty

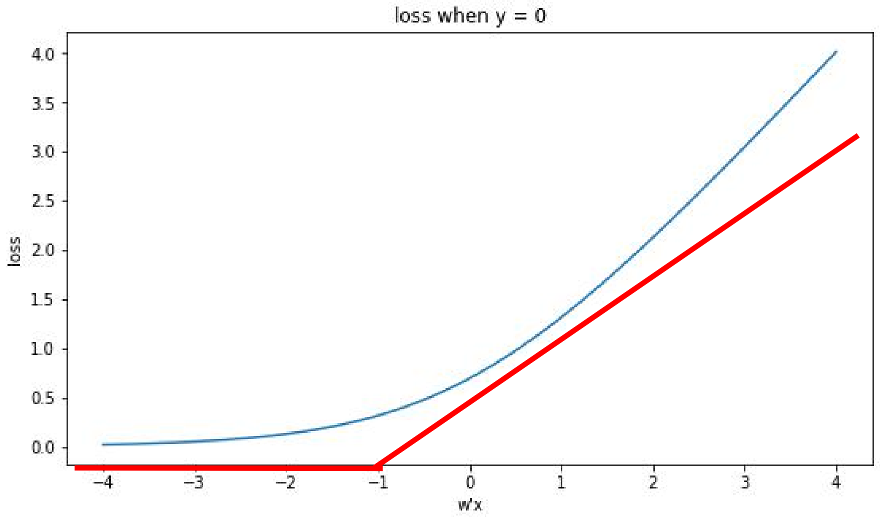
$$C\sum_{i=1}^{m}\left[-y_i\log\left(\sigma\left(w^T x_i\right)\right)-\left(1-y_i\right)\log\left(1-\sigma\left(w^T x_i\right)\right)\right]+\sum_{i}^{d} w_i^{\,2}$$

yi=1

yi=0



$\mathrm{cost}_1\left(w^T x_i\right)$



$\mathrm{cost}_0\left(w^T x_i\right)$

Adjusted Loss Function

$$C\sum_{i=1}^{m}\left[\underline{-y_i\log\left(\sigma\left(w^T x_i\right)\right)-\left(1-y_i\right)\log\left(1-\sigma\left(w^T x_i\right)\right)}\right]+\sum_{i}^{d} w_i^{\,2}$$

$$\Downarrow$$

$$C\sum_{i=1}^{m}\left[-y_i\bullet\mathrm{cost}_1\left(w^T x_i\right)-\left(1-y_i\right)\bullet\mathrm{cost}_0\left(w^T x_i\right)\right]+\sum_{i}^{d} w_i^{\,2}$$

SVM Objective Function

$$L = C\sum_{i=1}^{m}\left[-y_i \bullet \text{cost}_1\left(w^T x_i\right) - \left(1-y_i\right)\bullet \text{cost}_0\left(w^T x_i\right)\right] + \sum_{i}^{d} w_i^{\,2}$$

假设 C 很大，

$$y_i = 1,\ \text{want } w^T x_i \geq 1$$

$$y_i = 0,\ \text{want } w^T x_i \leq -1$$

此时，

$$L = C \bullet 0 + \sum_{i=1}^{d} w_i^{\,2} = \sum_{i=1}^{d} w_i^{\,2}$$

min (w) $\quad \left\|w\right\|^2$

s.t. $\quad w^T x_i \geq 1,\quad y_i = 1 \qquad$ i = 1,2,...,m

$\quad w^T x_i \leq \text{-}1,\quad y_i = 0$

结论

Hard SVM = Logistic Regression with L2 penatly

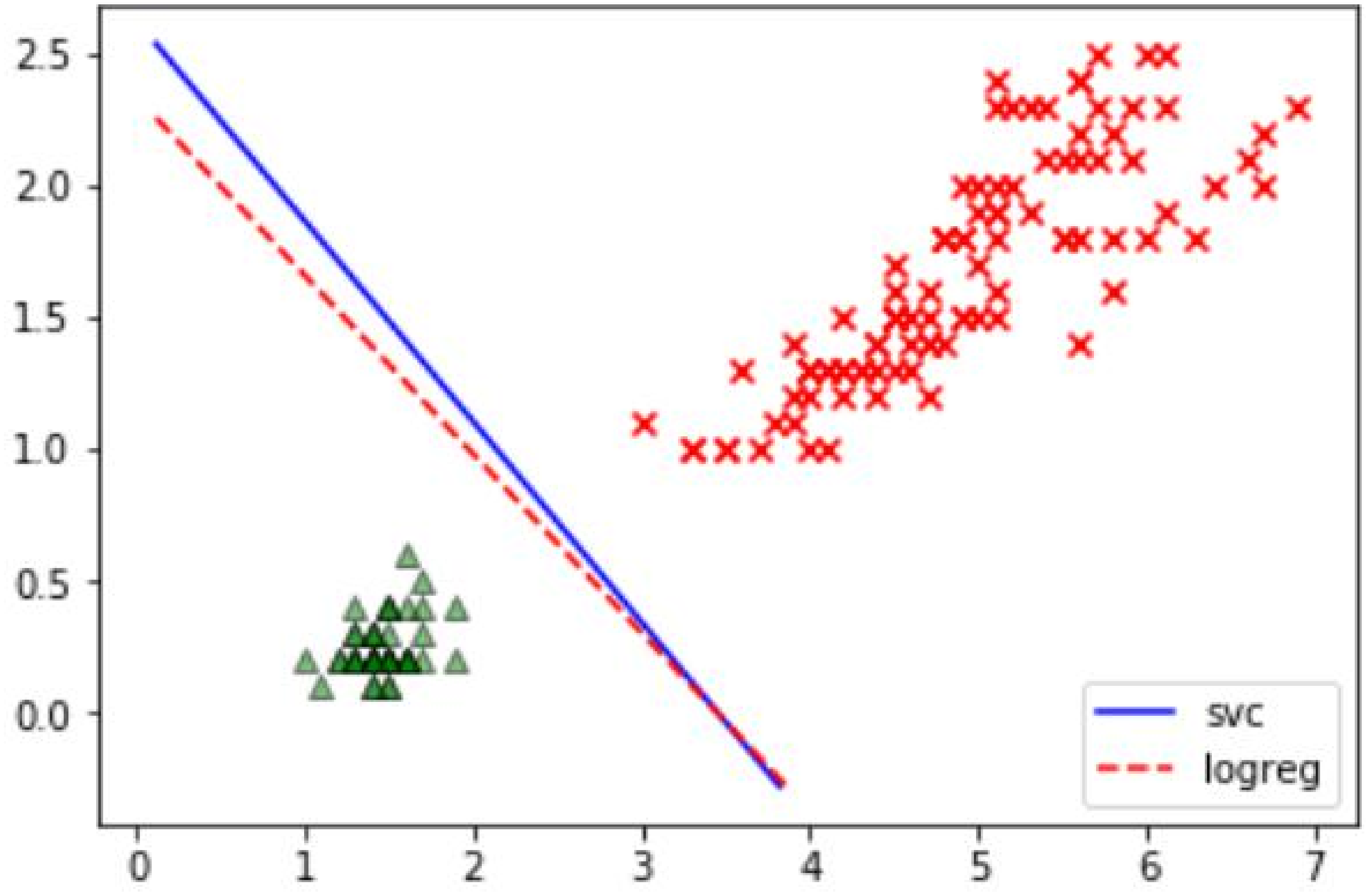
+

Adjusted cost function

+

Severe penalty

SVM vs LogisticRegression

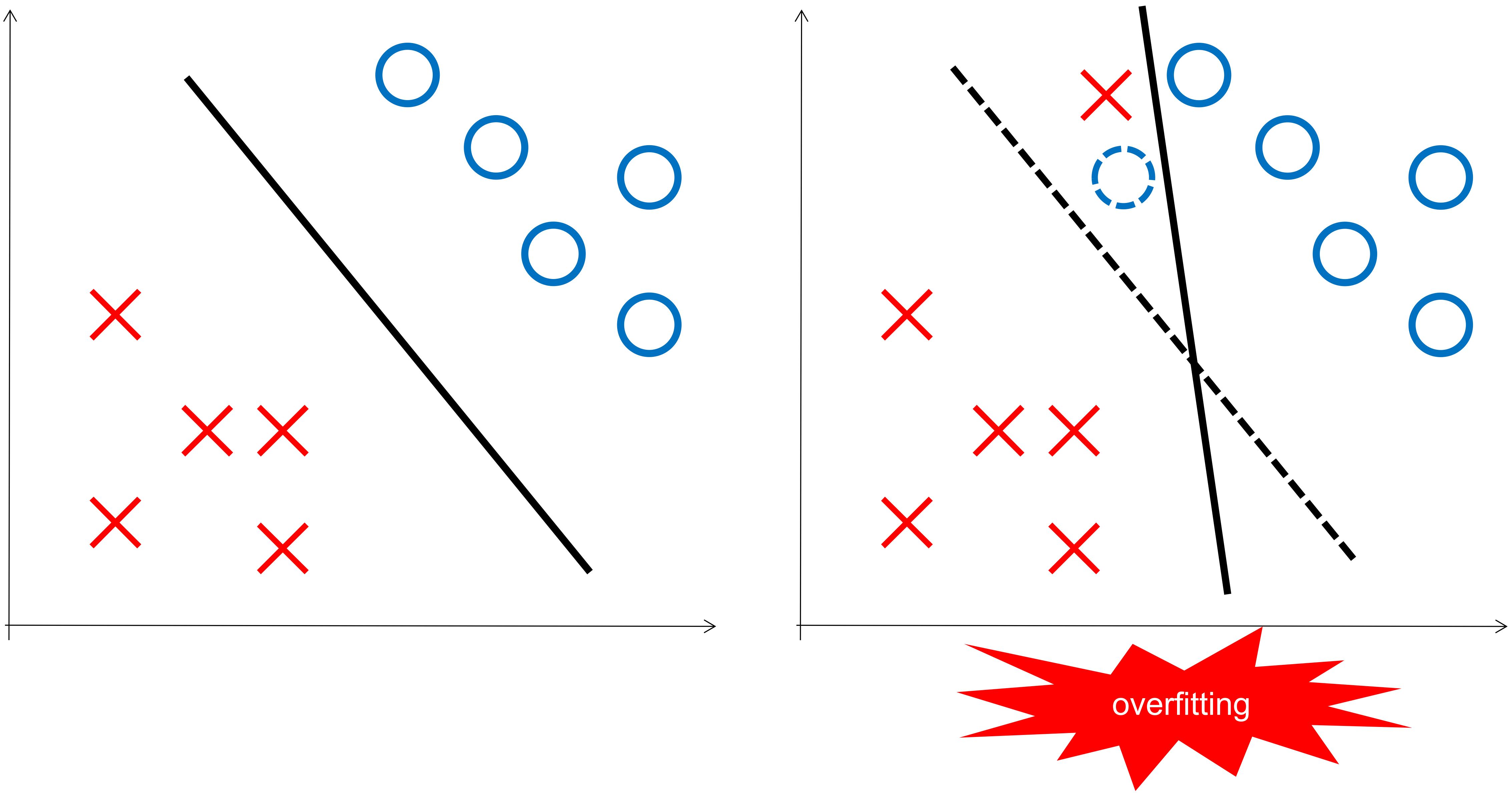


Soft Margin Classification

Why Need Soft Margin Classification?

- 线性不可分问题，只能用 soft margin classification
- Hard margin classification 对于异常值比较敏感

异常值问题



Soft SVM

Hard SVM

min $\frac{1}{2}\|w\|^2$
(w,b)

s.t. $y_i\left(w^T x_i + b\right) \geq 1,$ i = 1,2,...,m

Soft SVM

min $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} l\left(y_i\left(w^T x_i + b\right)\right)$
(w,b)

$$l(z) = \begin{cases} 1, & \text{if } z < 1 \\ 0, & \text{otherwise} \end{cases}$$

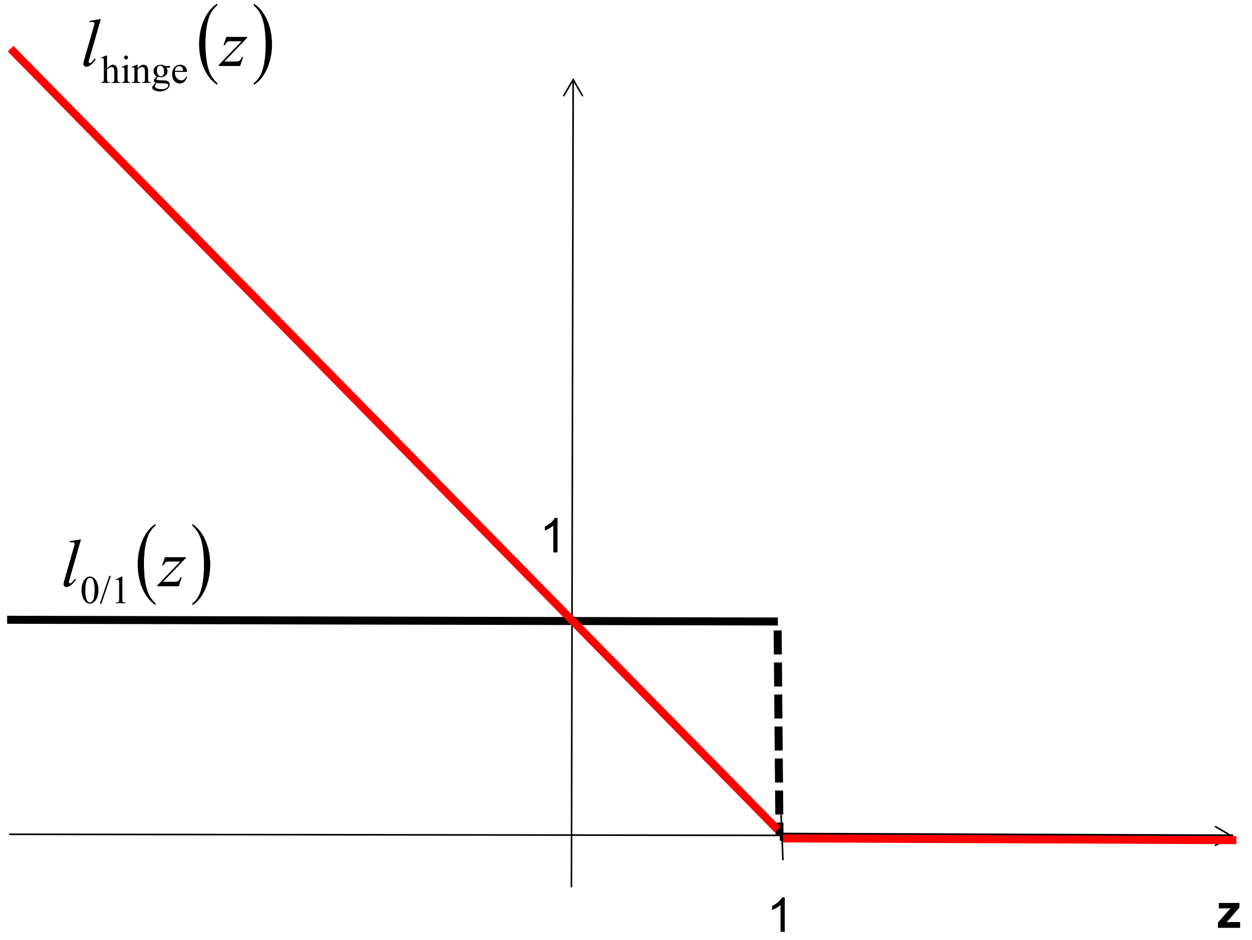
$$C \to +\infty$$

$$\text{Soft SVM} \to \text{Hard SVM}$$

选择损失函数

$$l_{0/1}(z)=\begin{cases}1, & \text{if } z<1\\ 0, & \text{otherwise}\end{cases}$$

$$l_{\text{hinge}}(z)=\max(0,1-z)$$



Soft SVM with Hinge Loss Function

$$\min_{(w,b)} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} l_{\text{hinge}}\left(y_i\left(w^T x_i + b\right)\right)$$

$$\min_{(w,b)} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \max\left(0, 1 - y_i\left(w^T x_i + b\right)\right)$$

$$\min_{w,b,\varepsilon_i} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \varepsilon_i$$

$$\text{s.t} \quad y_i\left(w^T x_i + b\right) \geq 1 - \varepsilon_i$$

$$\varepsilon_i \geq 0,\ i = 1,2,\ldots,m$$

Loss Function 的特征

单调递减函数

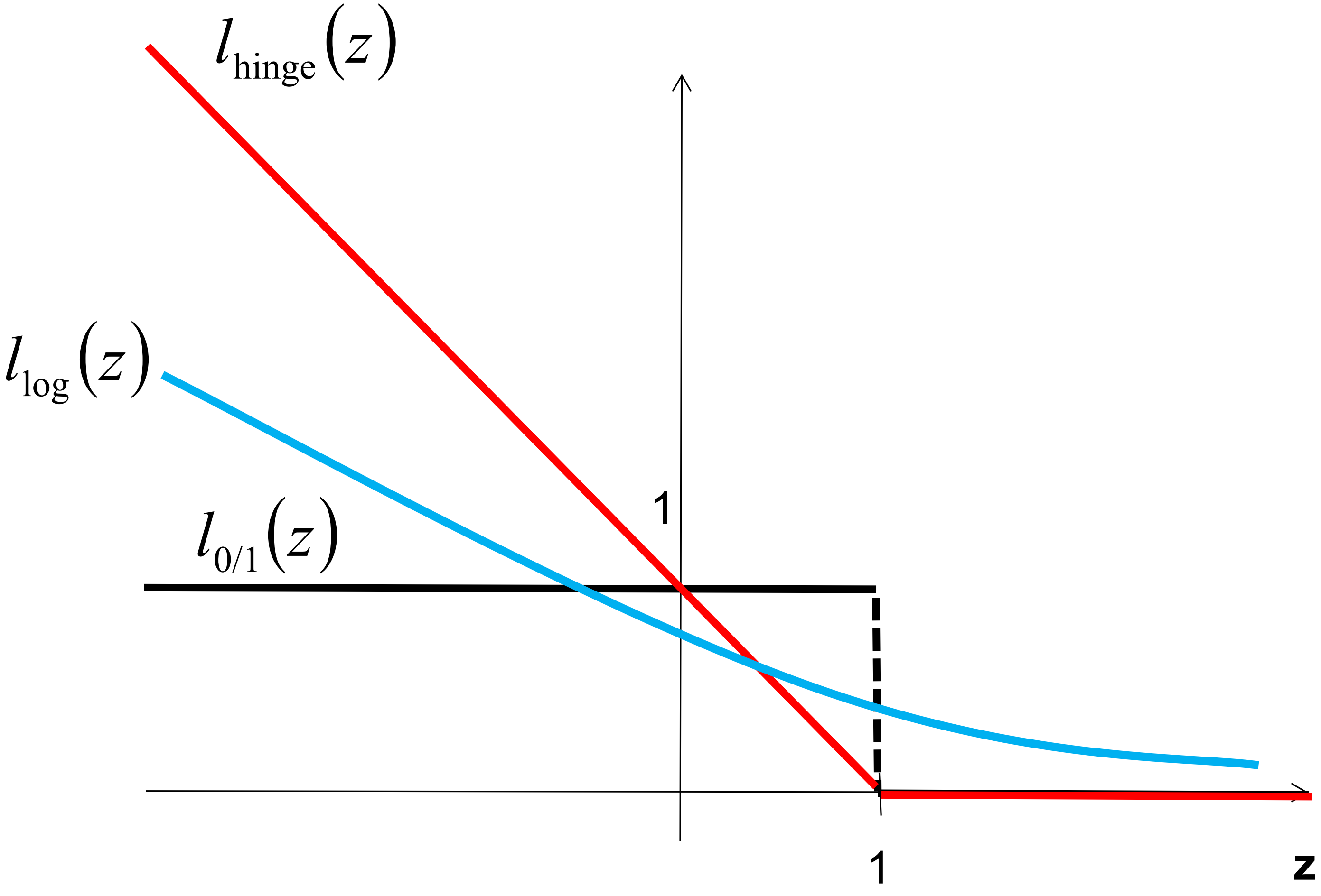
$$z \to +\infty, \quad l(z) \to 0$$
$$z \to -\infty, \quad l(z) \to +\infty$$

Logistic Loss Function

$$l_{0/1}(z) = \begin{cases} 1, & \text{if } z < 1 \\ 0, & \text{otherwise} \end{cases}$$

$$l_{\text{hinge}}(z) = \max(0, 1 - z)$$

$$l_{\log}(z) = \log(1 + \exp(-z))$$



Logistic Regession and SVM

Logistic Regression
with L2 penalty

=

Soft SVM
with logistic loss function

Logistic Regession and SVM (Optional)

Logistic Regression with L2 penalty

$$C\sum_{i=1}^{m}\left[-y_i\log\left(\sigma\left(w^T x_i\right)\right)-\left(1-y_i\right)\log\left(1-\sigma\left(w^T x_i\right)\right)\right]+\sum_{i}^{d} w_i^{\ 2}$$

正类，$y_i = 1$

负类，$y_i = 0$

$$\log\left(1+\mathrm{e}^{-\mathrm{v}_\mathrm{i}}\right)$$

$$\log\left(1+\mathrm{e}^{\mathrm{v}_\mathrm{i}}\right)$$

正类，$y_i = 1$

负类，$y_i = 0$

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\log\left(1+e^{-y^i\left(w^T x_i + b\right)}\right)$$

Soft SVM with logistic loss function

$$v_i = w^T x_i$$

$$||$$

在 logistic regression 中，
xi 包含了截距项

$$||$$

$$v_i = w^T x_i + b$$

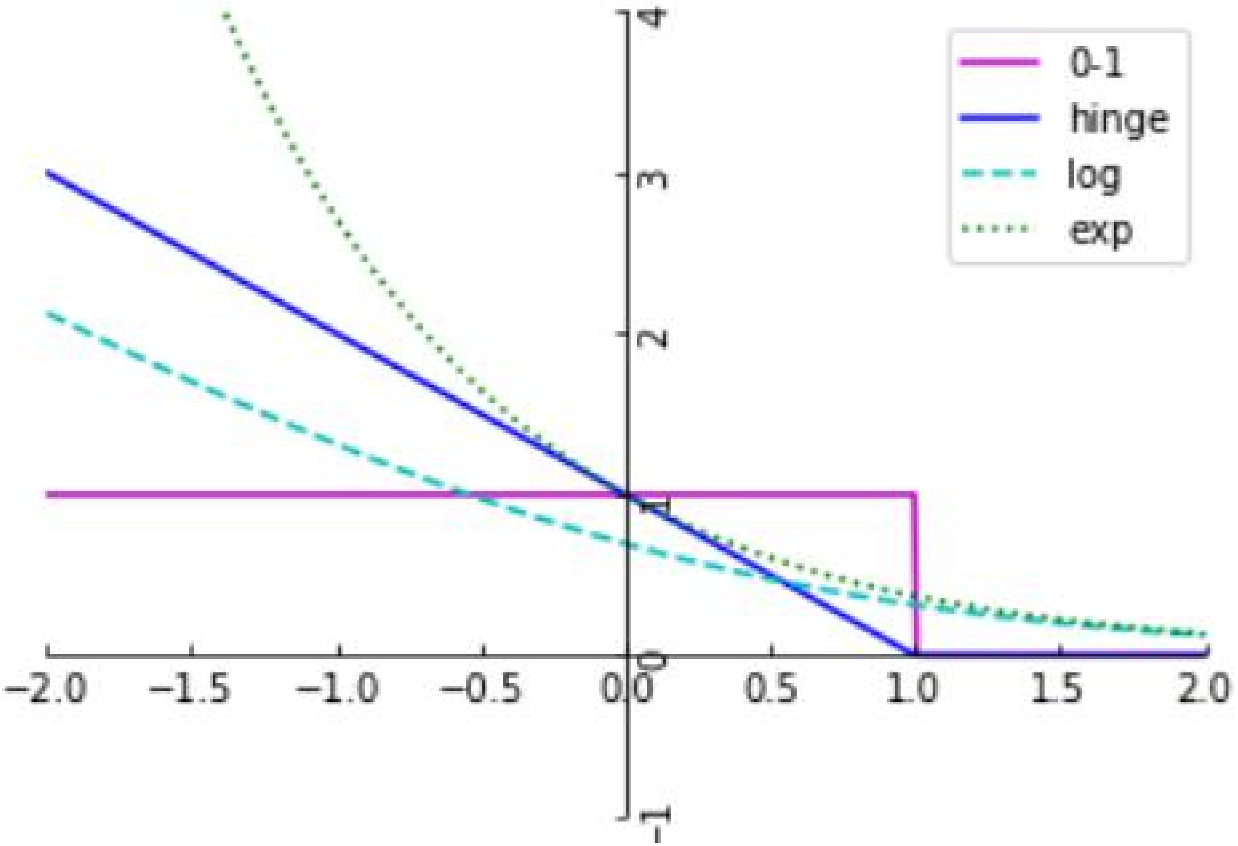
Other Loss Functions

$$l_{0/1}(z) = \begin{cases} 1, & \text{if } z < 1 \\ 0, & \text{otherwise} \end{cases}$$

$$l_{\text{hinge}}(z) = \max(0, 1 - z)$$

$$l_{\log}(z) = \log(1 + \exp(-z))$$

$$l_{\exp}(z) = \exp(-z)$$



使用 LinearSVC

LinearSVC

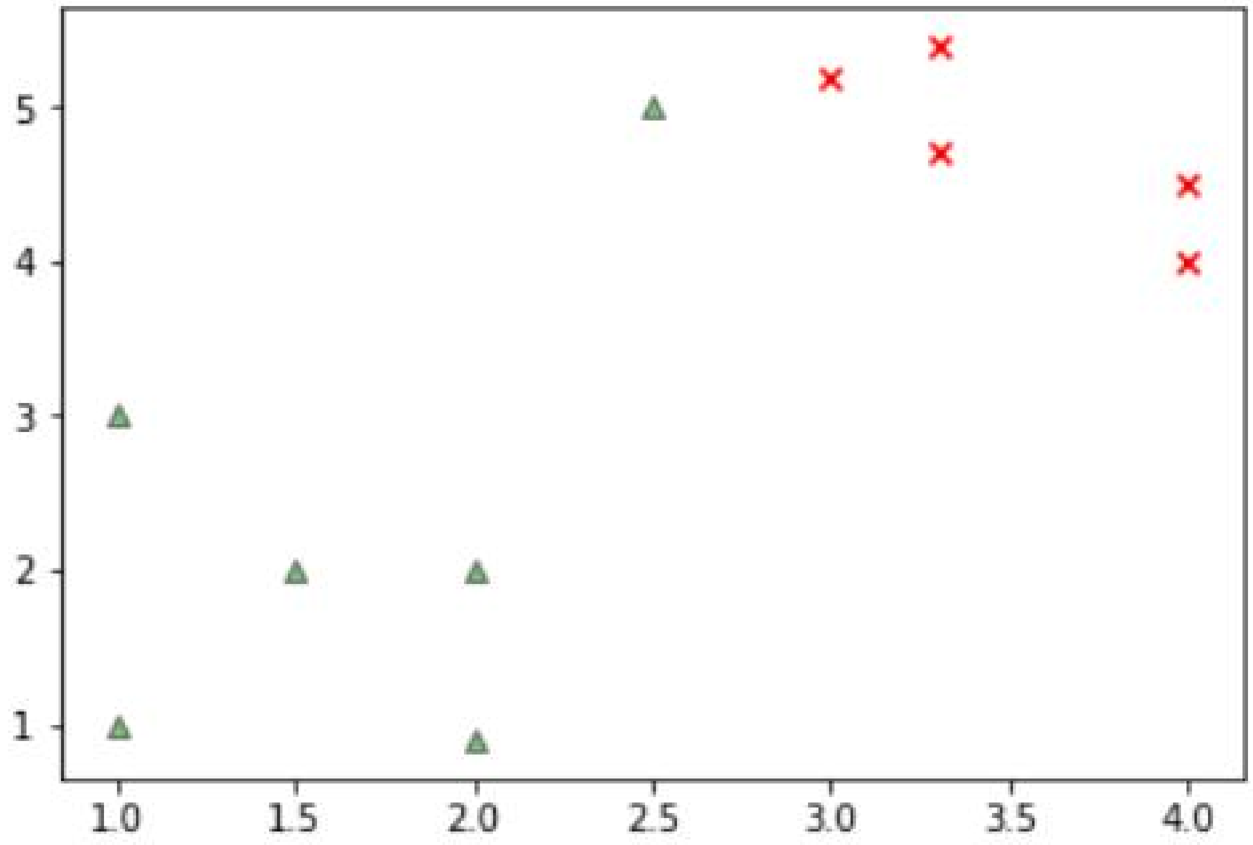
sklearn.svm.LinearSVC

关键参数：

- C = 1.0，误差惩罚力度，**C 越大，模型越倾向于 Hard SVM**

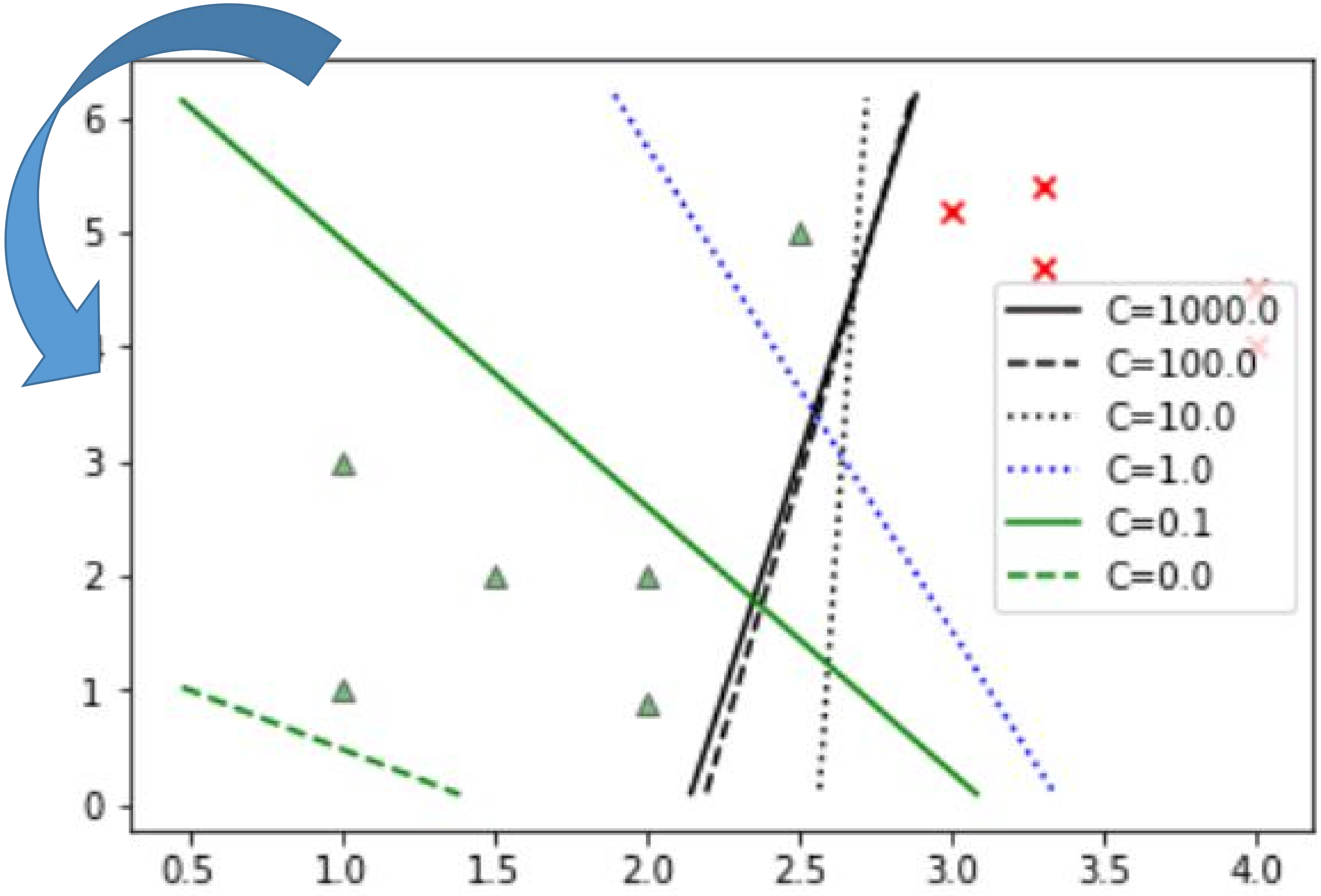
LinearSVC

C = 1000, 100, 10, 1, 0.1, 0.01



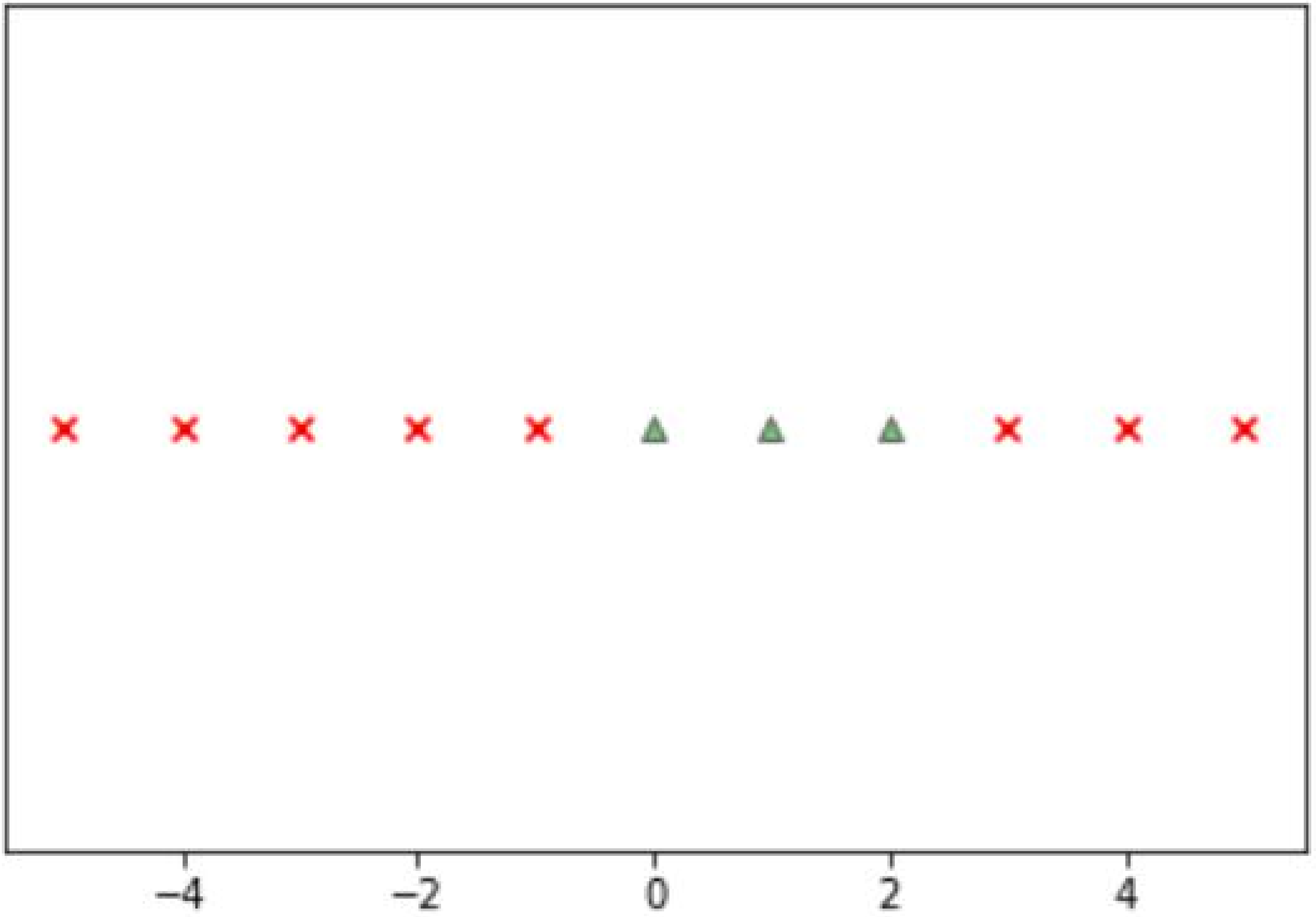
LinearSVC

C: large -> small
overfitting -> underfitting

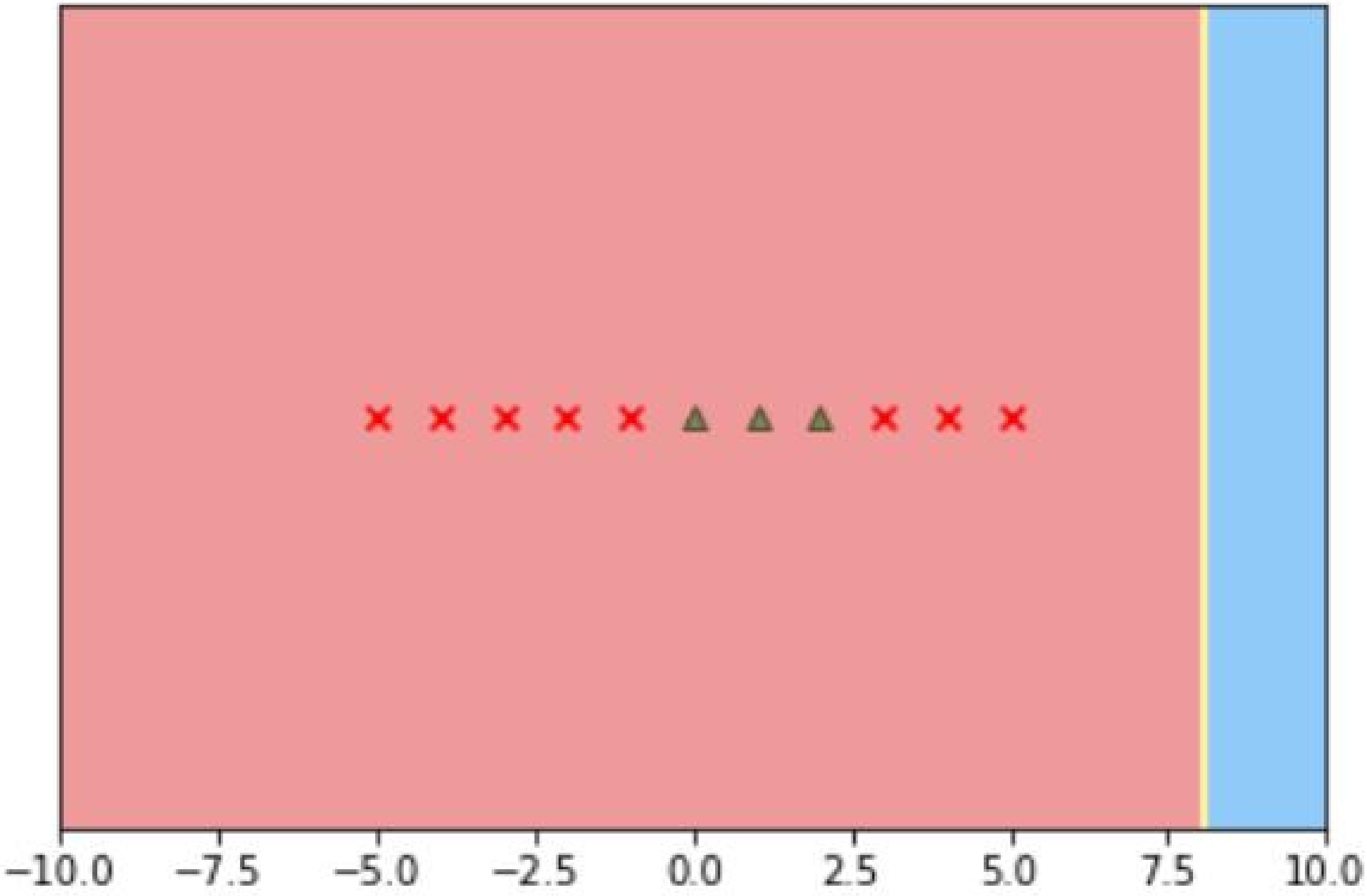


Kernel SVM

考虑以下一维度数据的分类问题

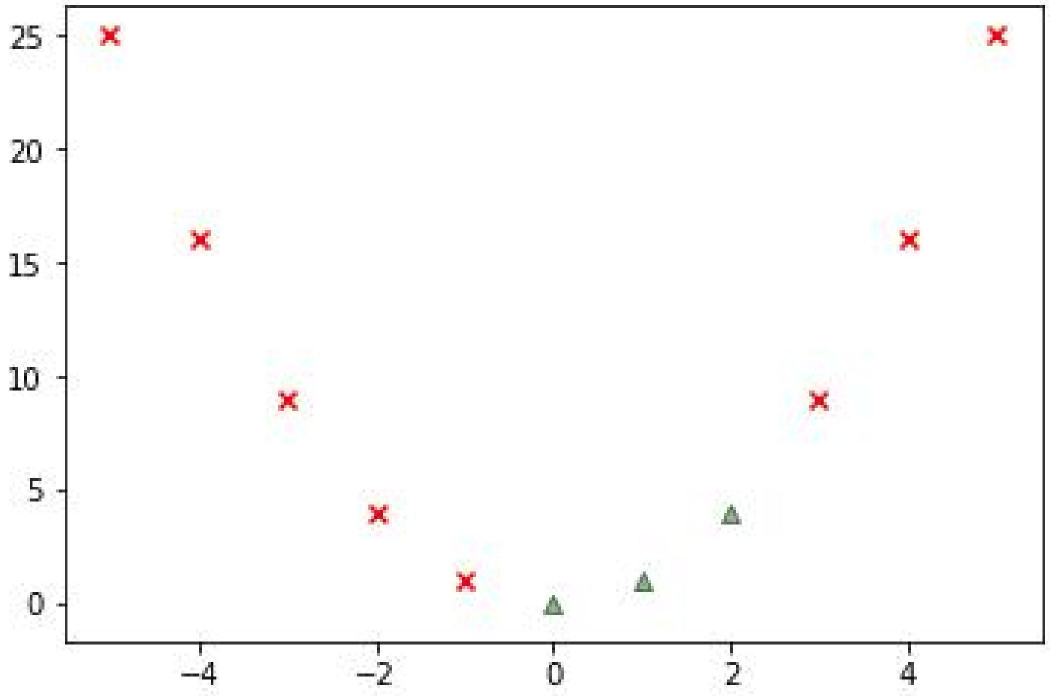


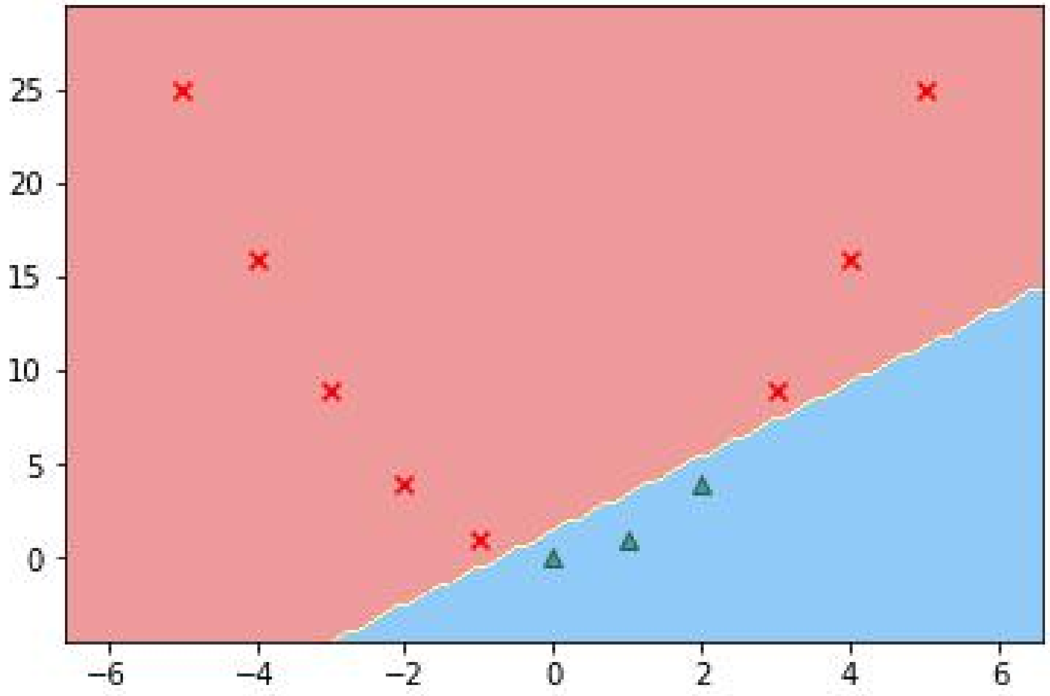
强行分类



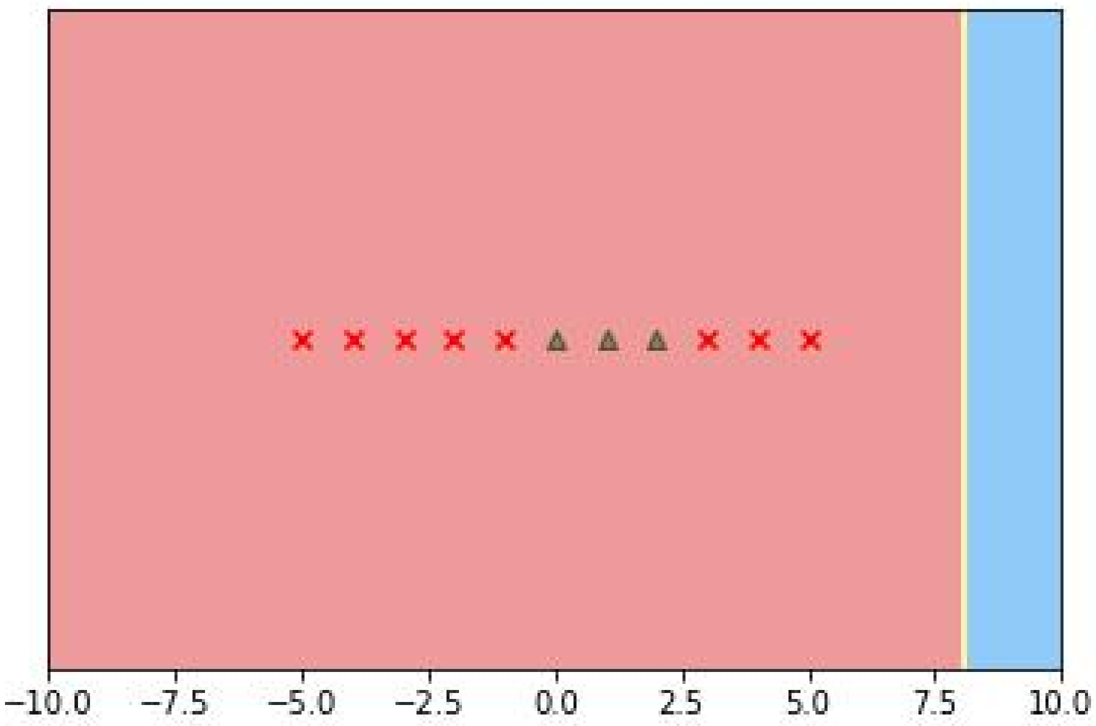
升维处理

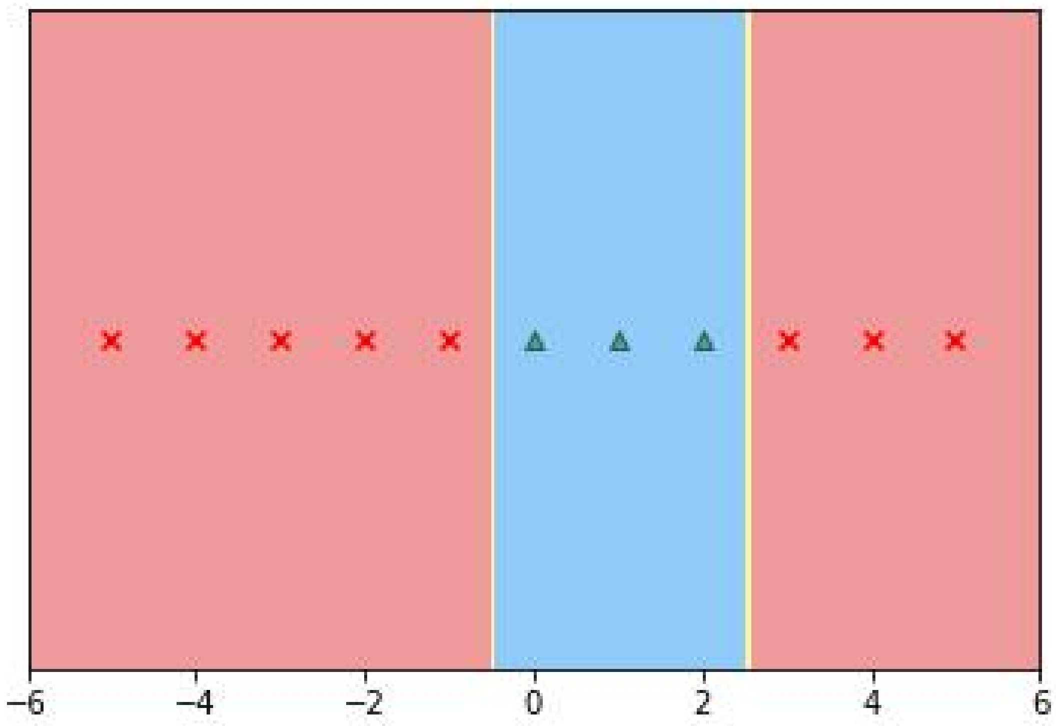
增加第二个维度：x2 = x1 * x1



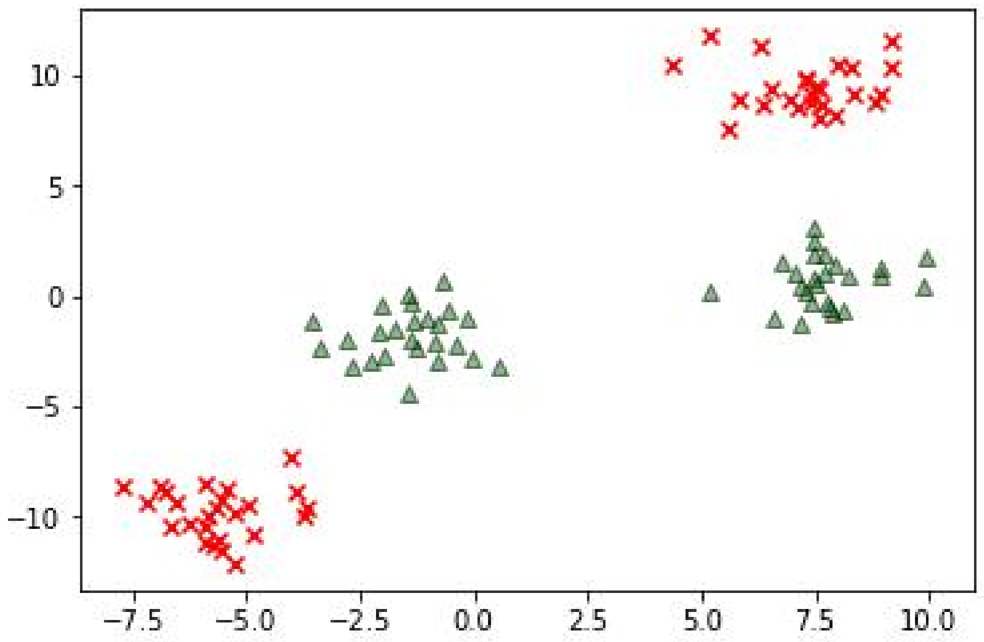


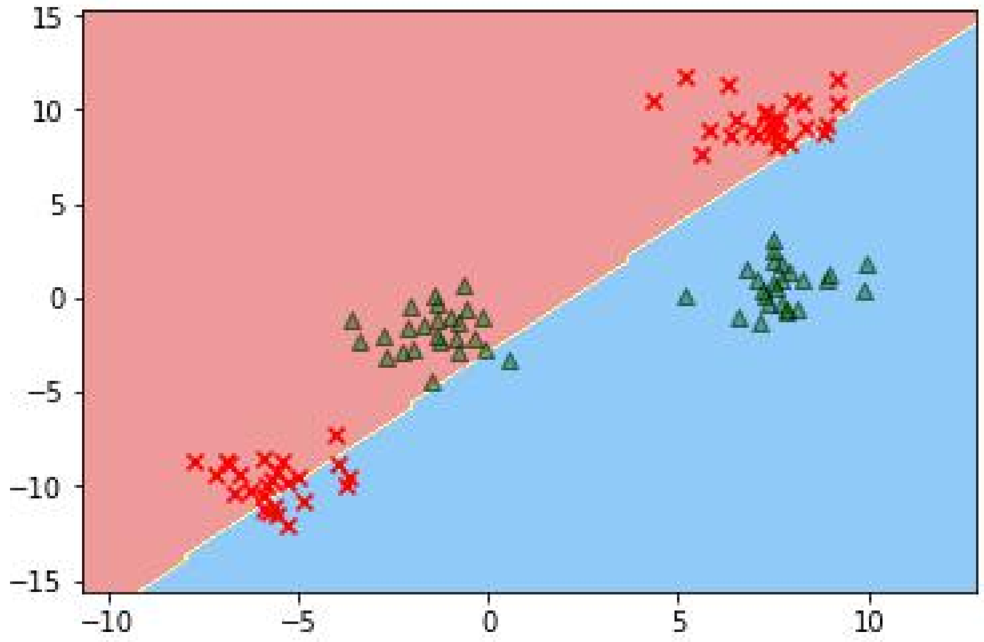
在原始数据上的分类效果





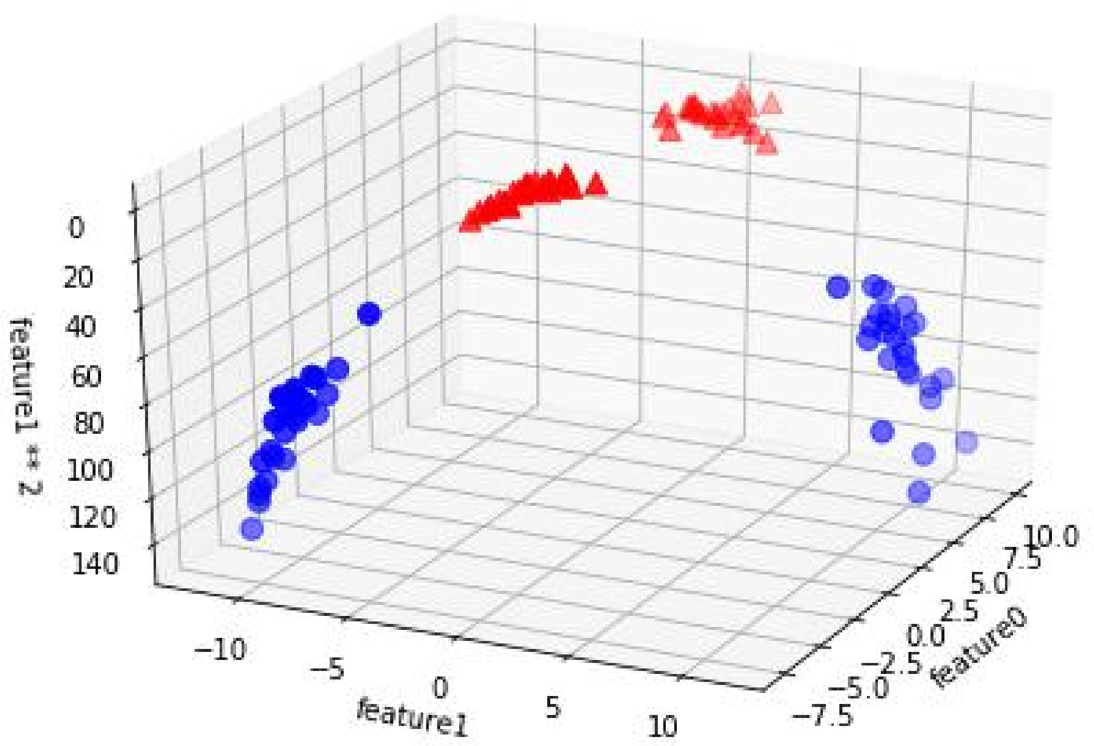
考虑以下二维度数据的分类问题

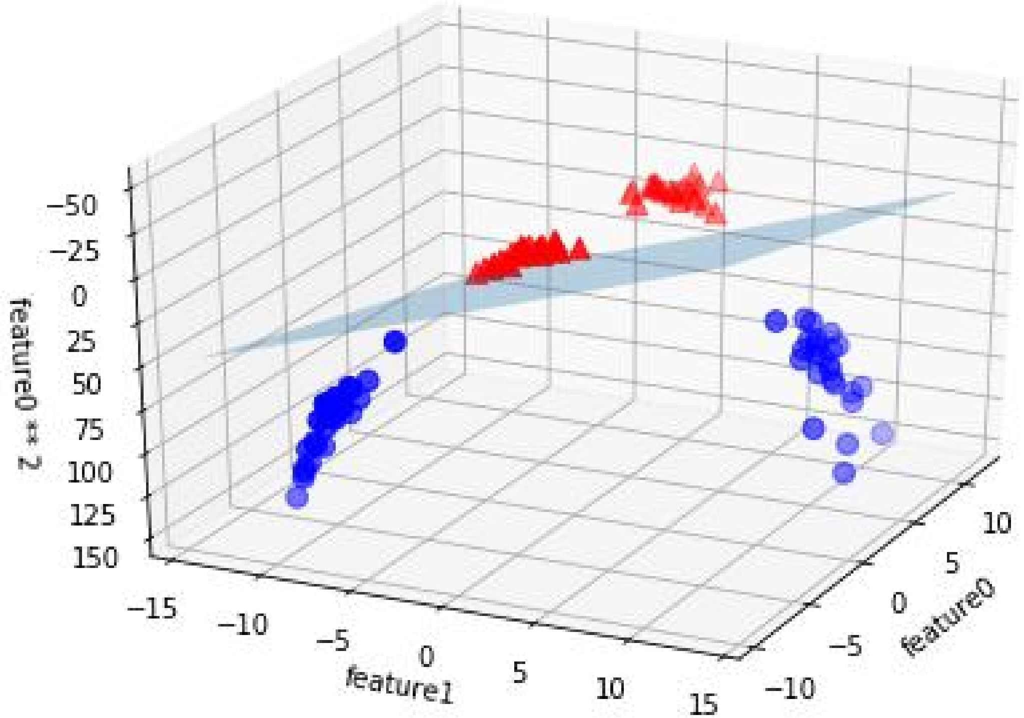




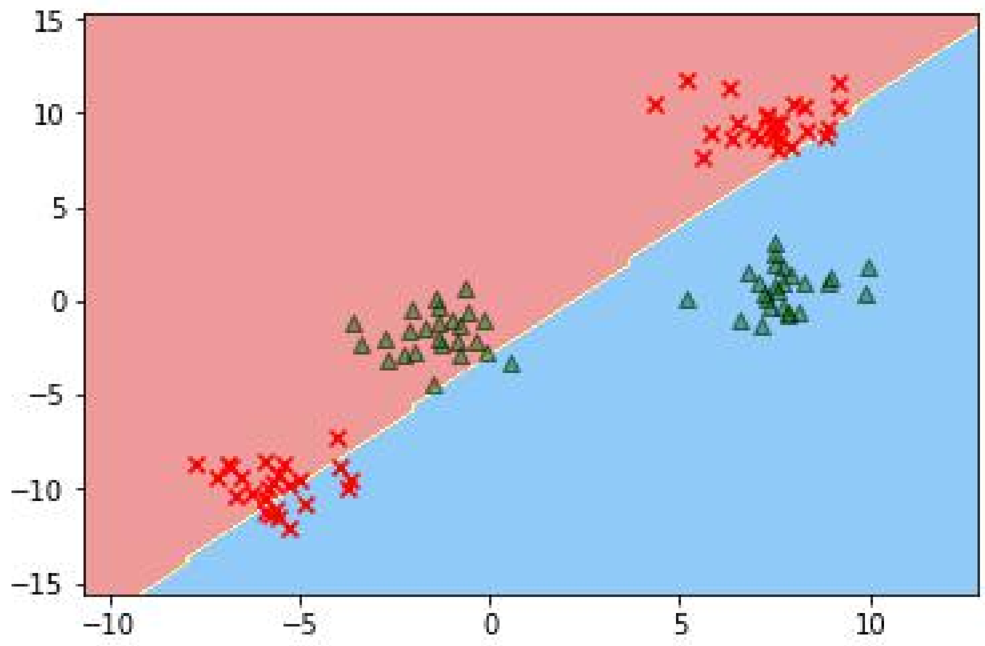
升维处理

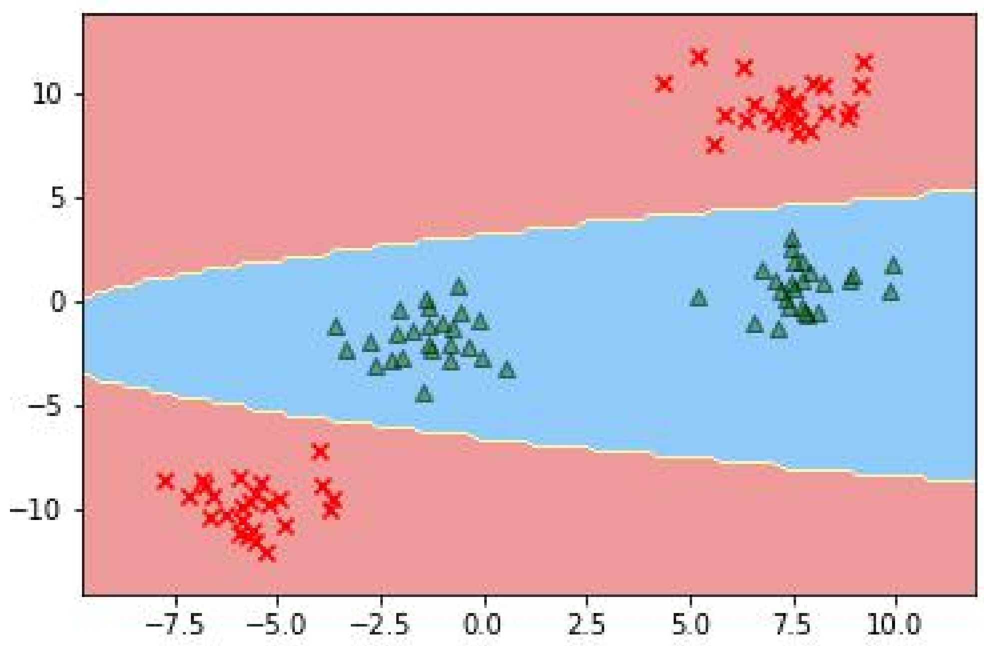
增加第三个维度：x3 = x2 * x2





在原始数据上的分类效果





总结

$$x \xrightarrow{\text{升维}} \phi(x) \xrightarrow{\text{SVM}} w^T\phi(x)+b$$

原空间　　　　特征空间

特征空间上线性分类器 = 原空间上的非线性分类器

如果原空间维度有限（有限特征），则一定存在一个高维特征空间使得样本可分

Soft SVM with Hinge Loss Function (Optional)

min
(w,b)

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\max\left(0, 1 - y_i\left(w^T x_i + b\right)\right)$$

$$\min_{w,b,\varepsilon_i} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\varepsilon_i$$

$$\text{s.t} \quad y_i\left(w^T x_i + b\right) \geq 1 - \varepsilon_i$$

$$\varepsilon_i \geq 0, i = 1,2,\ldots,m$$

对偶问题

$$\max_{\alpha} \quad \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$\text{s.t} \quad \sum_{i=1}^{m}\alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C, i = 1,2,\ldots,m$$

维度问题 (Optional)

原空间

$$x$$

特征空间

$$\phi(x)$$

$$\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boxed{x_i^T x_j}$$

$$\begin{pmatrix} x_1[1] \\ x_1[2] \\ \dots \\ x_1[d] \end{pmatrix} \bullet \begin{pmatrix} x_2[1] \\ x_2[2] \\ \dots \\ x_2[d] \end{pmatrix}$$

$$\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boxed{\phi(x_i)^T\phi(x_j)}$$

$$\begin{pmatrix} \phi(x_1)[1] \\ \phi(x_1)[2] \\ \dots \\ \dots \\ \phi(x_1)[d] \end{pmatrix} \bullet \begin{pmatrix} \phi(x_2)[1] \\ \phi(x_2)[2] \\ \dots \\ \dots \\ \phi(x_2)[d] \end{pmatrix}$$

Kernel Trick (Optional)

找到一个函数 k 使得以下等式成立

$$\phi(x_i)^T \phi(x_j) = \kappa(x_i, x_j)$$

Kernel Example (Optional)

$$\phi\text{:}\ \begin{pmatrix} x[1] \\ x[2] \end{pmatrix} \rightarrow \begin{pmatrix} x[1]^2 \\ \sqrt{2}x[1]x[2] \\ x[2]^2 \end{pmatrix}$$

$$\kappa\left(x_i, x_j\right) = \left(x_i{}^T x_j\right)^2$$

可以证明： $$\phi\left(x_i\right)^T \phi\left(x_j\right) = \kappa\left(x_i, x_j\right)$$

Kernel Example 2 (Optional)

如果$\kappa(x_i, x_j) = x_i^T x_j$，则$\phi(x) = ?$

$$\phi(x_i)^{\mathrm{T}} \phi(x_j) = \kappa(x_i, x_j) = x_i^T x_j$$

$$\Rightarrow \quad \phi(x) = x$$

问题 (Optional)

$$\text{升维} \rightleftarrows \phi \;\underset{?}{\rightleftarrows}\; \kappa$$

Mercer's Theorem (Optional)

当 κ 满足下列条件时，一定有一个 ϕ 与之对应

即核函数**隐式地定义了特征空间**

核矩阵 **K =** $\begin{bmatrix} \kappa(x_1, x_1) & \cdots & \kappa(x_1, x_j) & \cdots & \kappa(x_1, x_m) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \kappa(x_i, x_1) & \cdots & \kappa(x_i, x_j) & \cdots & \kappa(x_i, x_m) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \kappa(x_m, x_1) & \cdots & \kappa(x_m, x_j) & \cdots & \kappa(x_m, x_m) \end{bmatrix}$ 对称，且半正定

常用（和不常用）的核函数

线性核 $\kappa(x_i, x_j) = x_i^T x_j$ → **不升维**

多项式核 $\kappa(x_i, x_j) = (x_i^T x_j)^d, \ \mathrm{d} \geq 1$ → **d=2, 所有两次项**

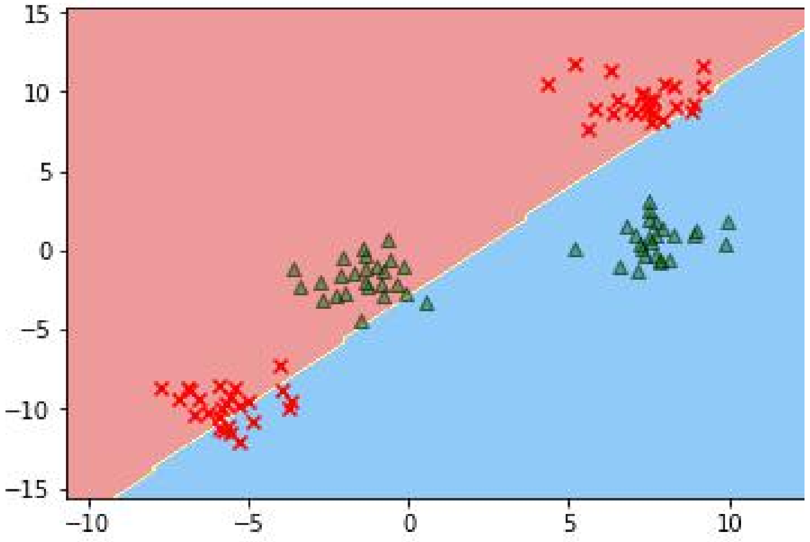
高斯核 $\kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \ \sigma > 0$

拉普拉斯核 $\kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{\sigma}\right), \ \sigma > 0$

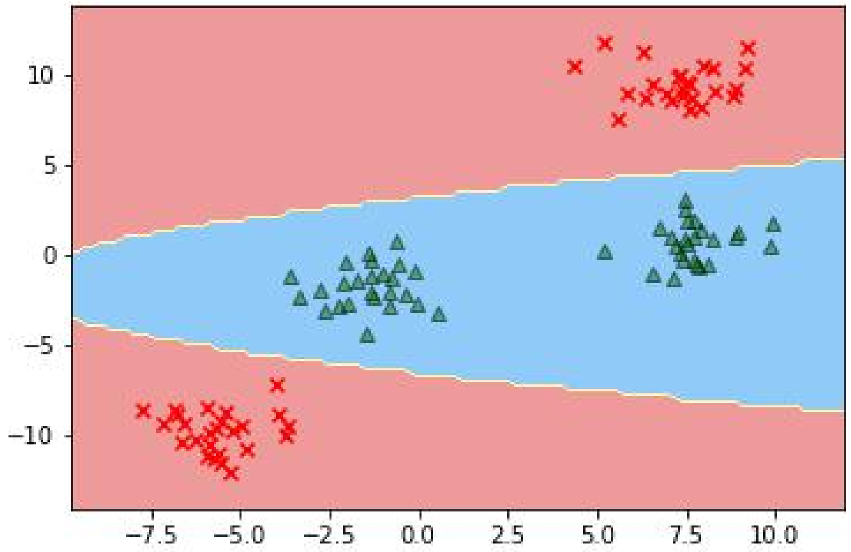
Sigmoid 核 $\kappa(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta), \ \beta > 0, \theta < 0$

Gaussian Kernel

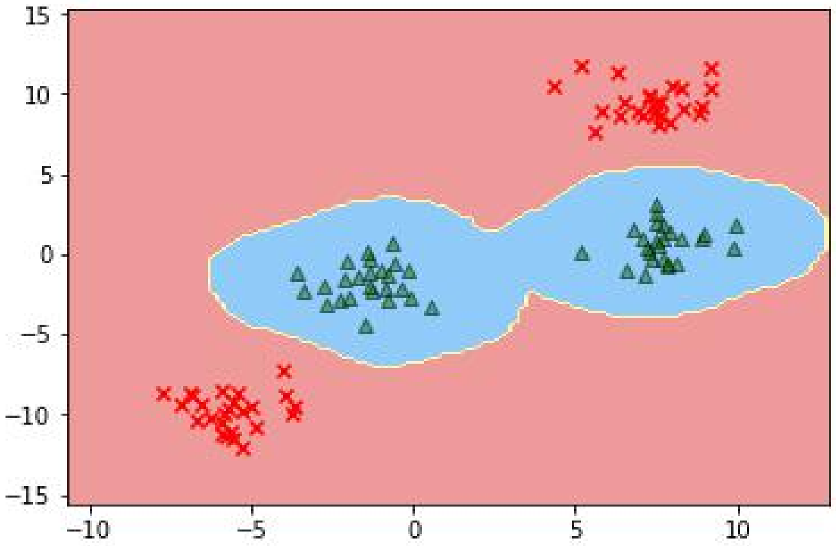
Linear Kernel



手动生成特征空间



Gaussian Kernel



Kernel SVM 调参

SVC

sklearn.svm.SVC

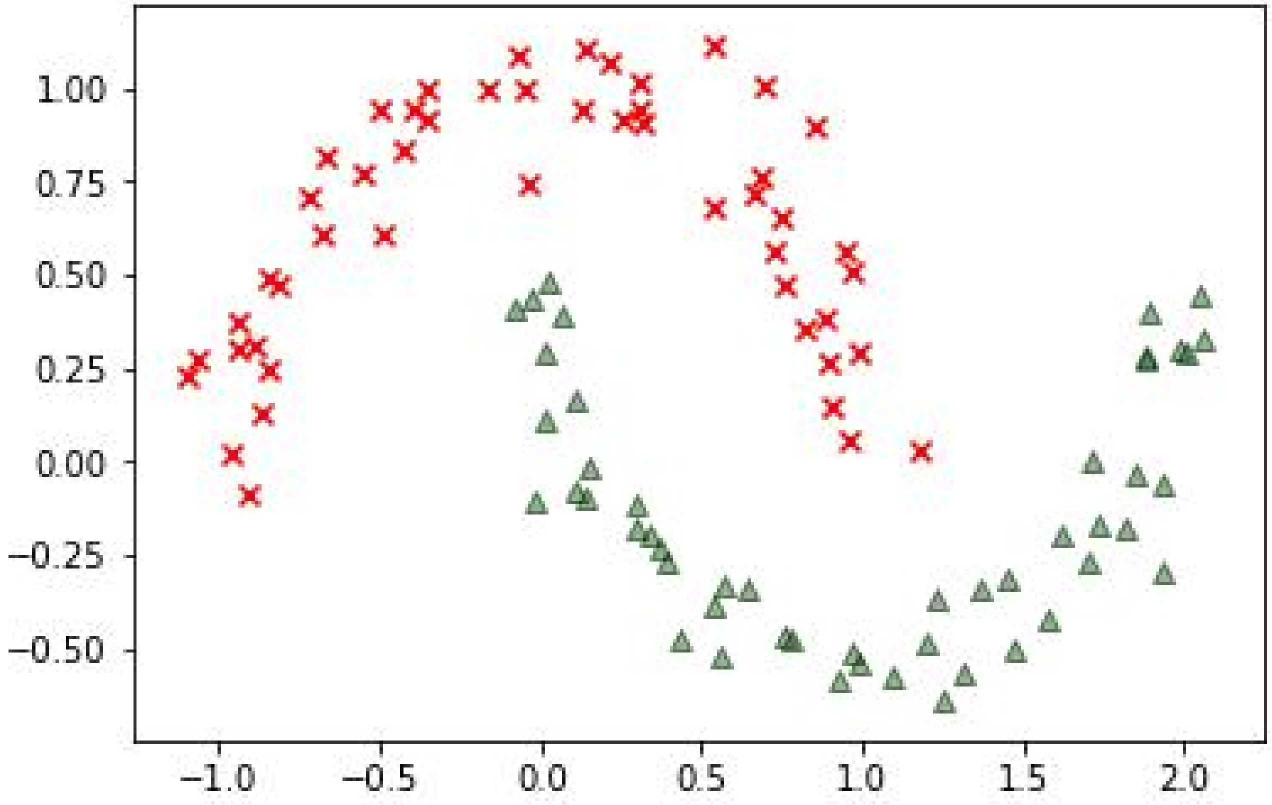
关键参数：

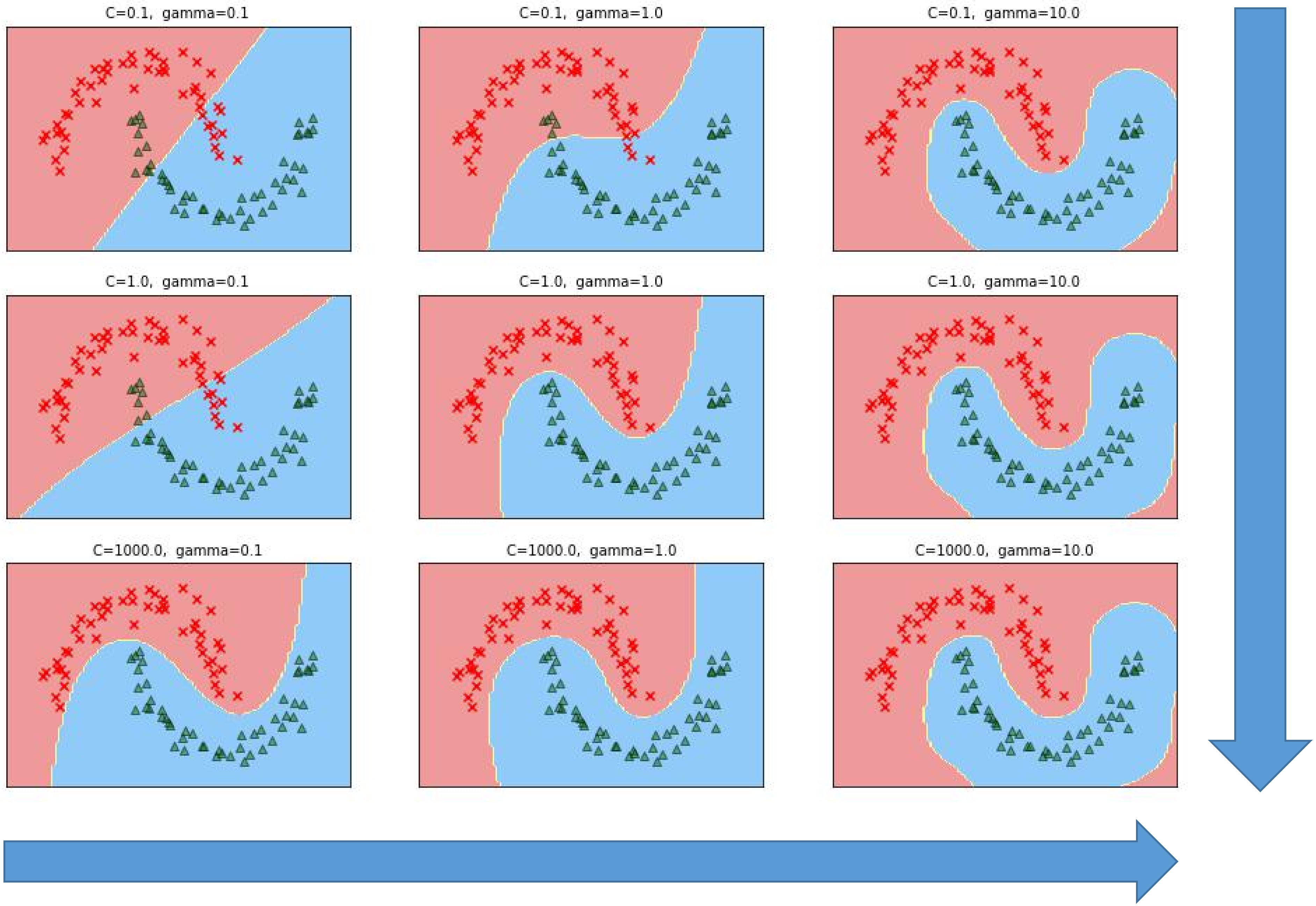
- kernal = ‘rbf', **'rbf' - 高斯核，'linear' - 线性核, 'ploy' - 多项式核**
- C = 1.0
- gamma = 1 / n_features

$$\exp(-\gamma \| x - x' \|^2)$$

SVC

C = 0.1, 1, 1000
gamma = 0.1, 1, 10



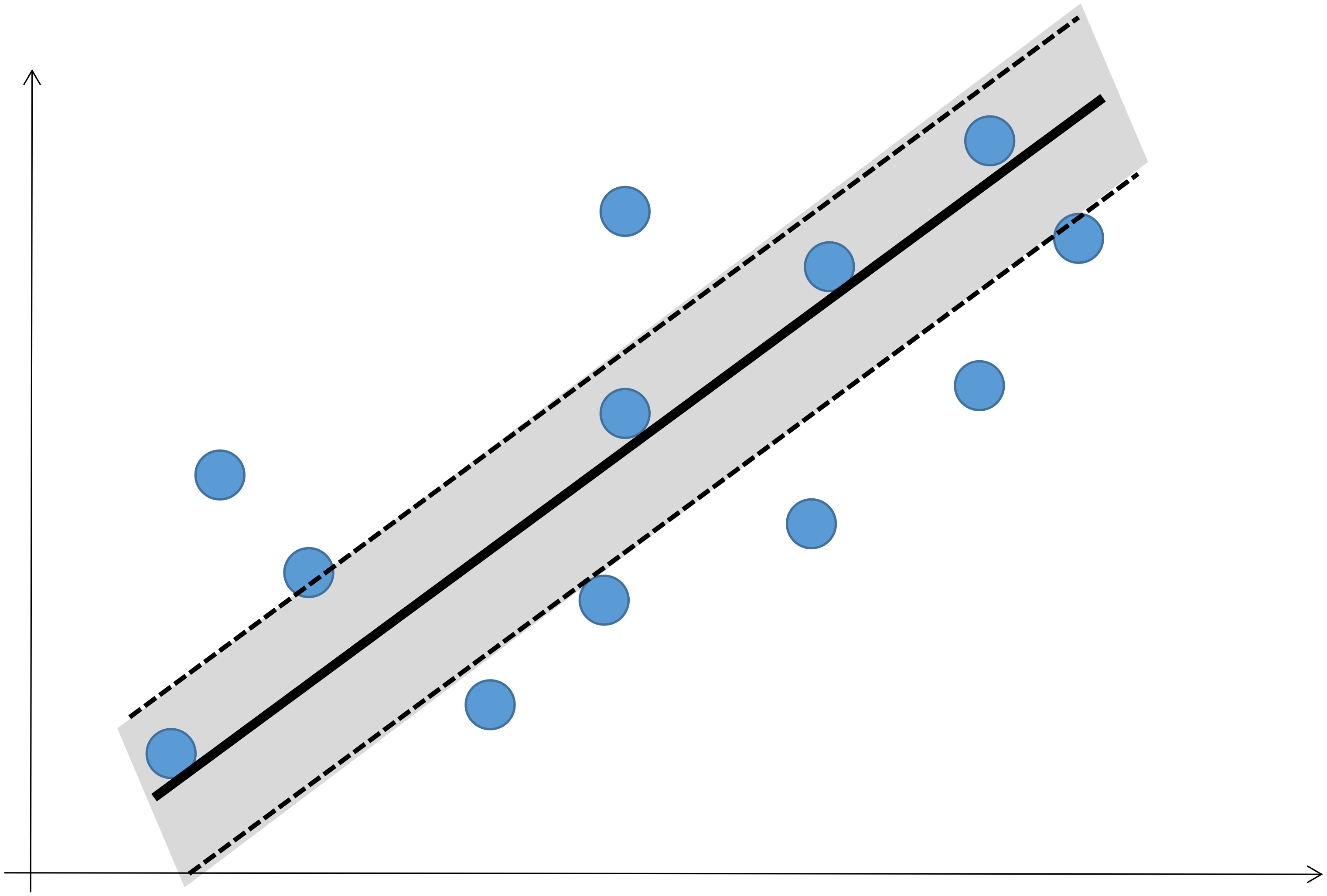


C 增加，
模型复杂度提高，
过拟合风险变大

gamma 增加， 模型复杂度提高， 过拟合风险变大

支持向量回归（SVR）

SVR



SVR 损失函数

$$\min_{w} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \boxed{l(e_i)}$$

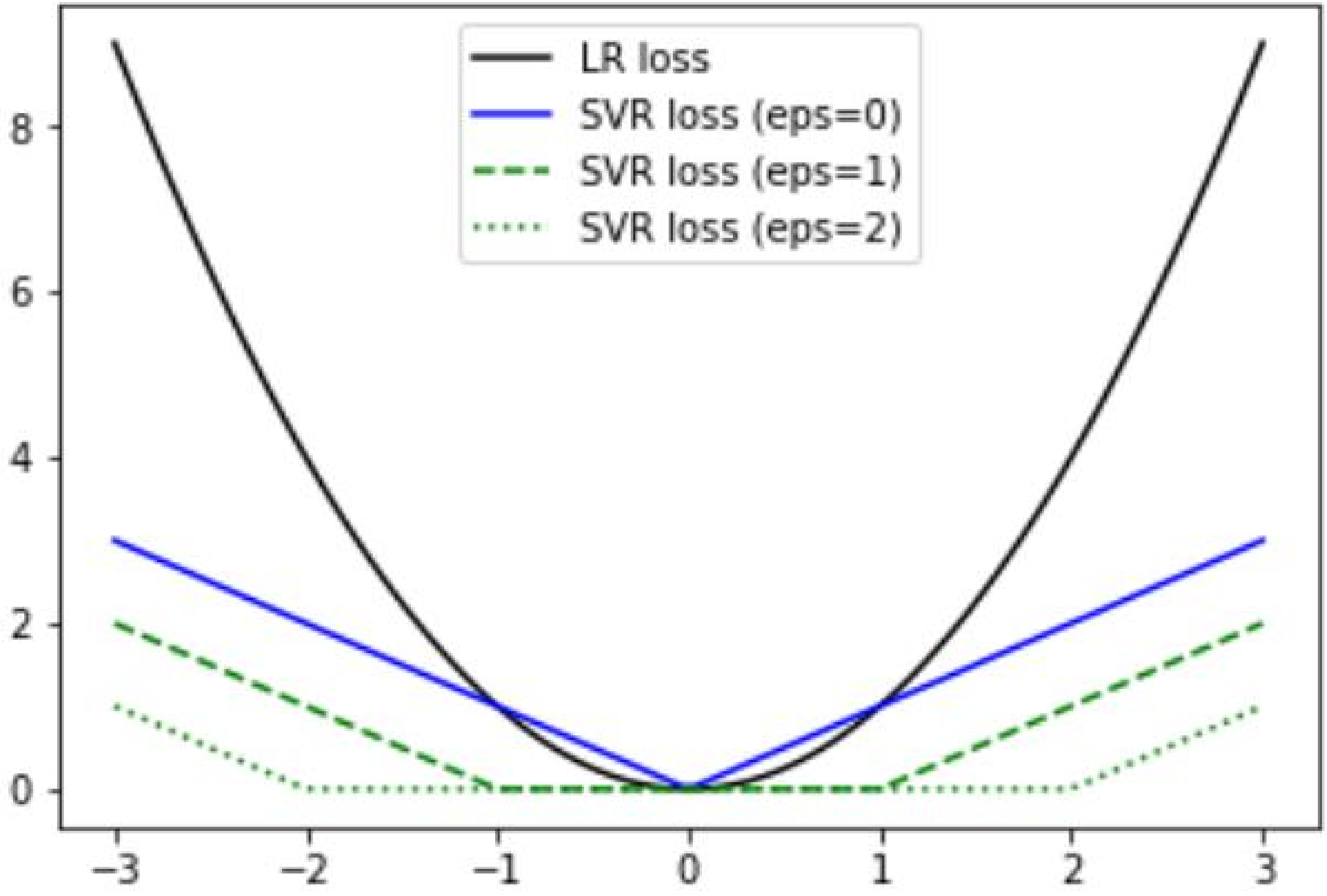
Linear Regression

$$\min_{w} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} e_i^{\ 2}$$

SVR

$$\min_{w} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \left(|e_i| - \varepsilon\right)^+$$

SVR 损失函数



LR loss
SVR loss (eps=0)
SVR loss (eps=1)
SVR loss (eps=2)
8
6
4
2
0
−3
−2
−1
0
1
2
3

LinearSVR

sklearn.svm.LinearSVR

关键参数：

- epsilon = 0
- C = 1.0：**C 越大，则 w 越接近 0**