支持向量机

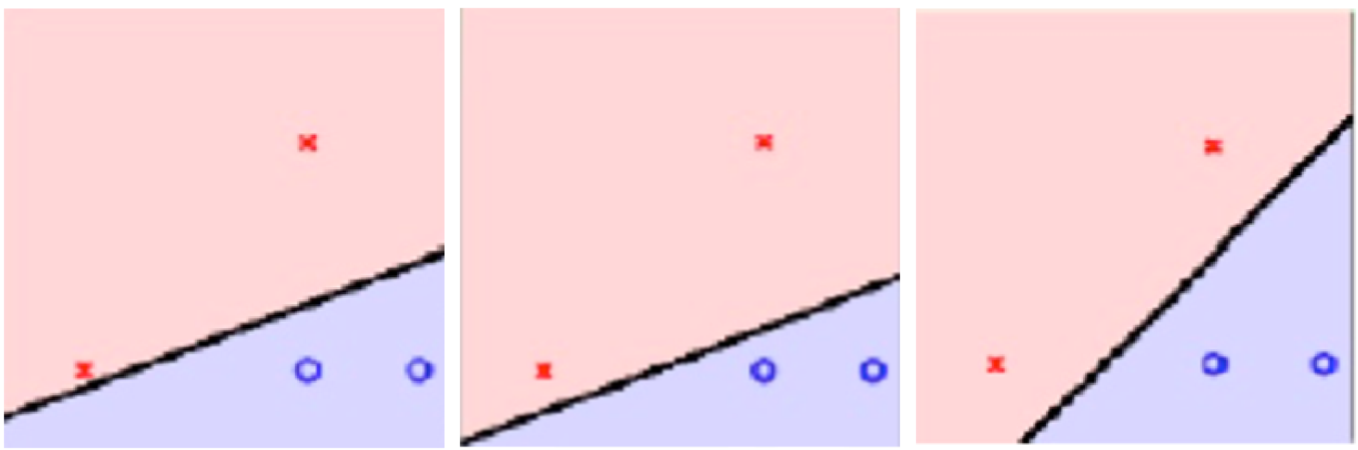
支持向量机

- 三个特征：间隔、对偶、核技巧
- 三类问题：
  - hard-margin SVM: 线性可分问题
  - soft-margin SVM: 线性不可分问题，可以允许一些分类错误
  - kernal SVM: 非线性问题
- 支持向量机的模型：

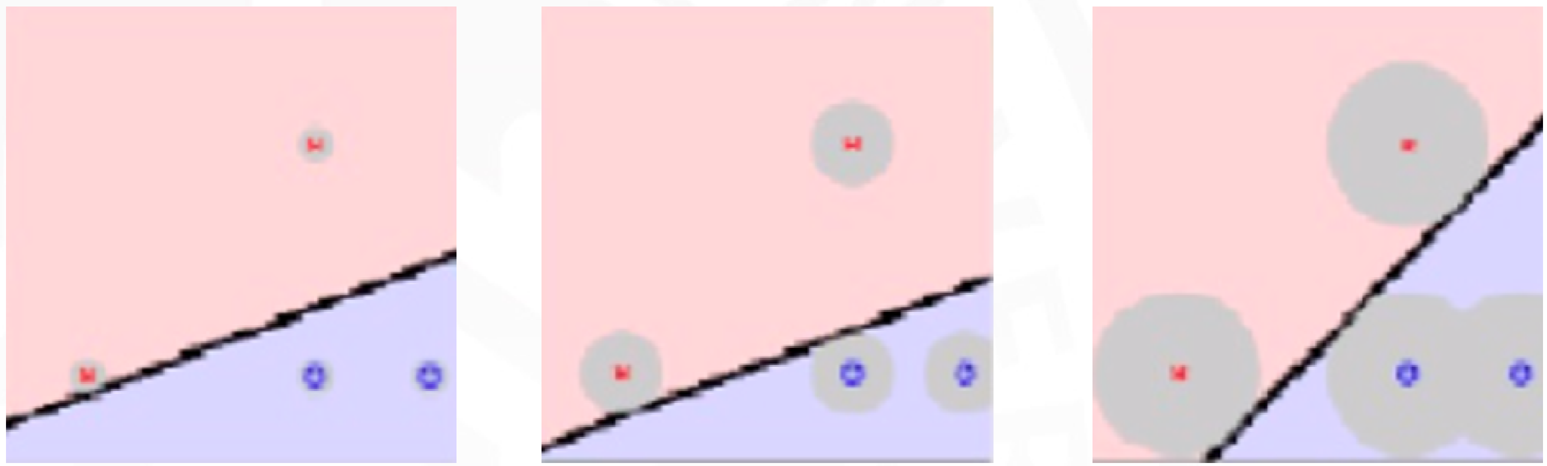
$$y = \mathbf{w}^T \mathbf{x} + b$$

支持向量机的基本思想

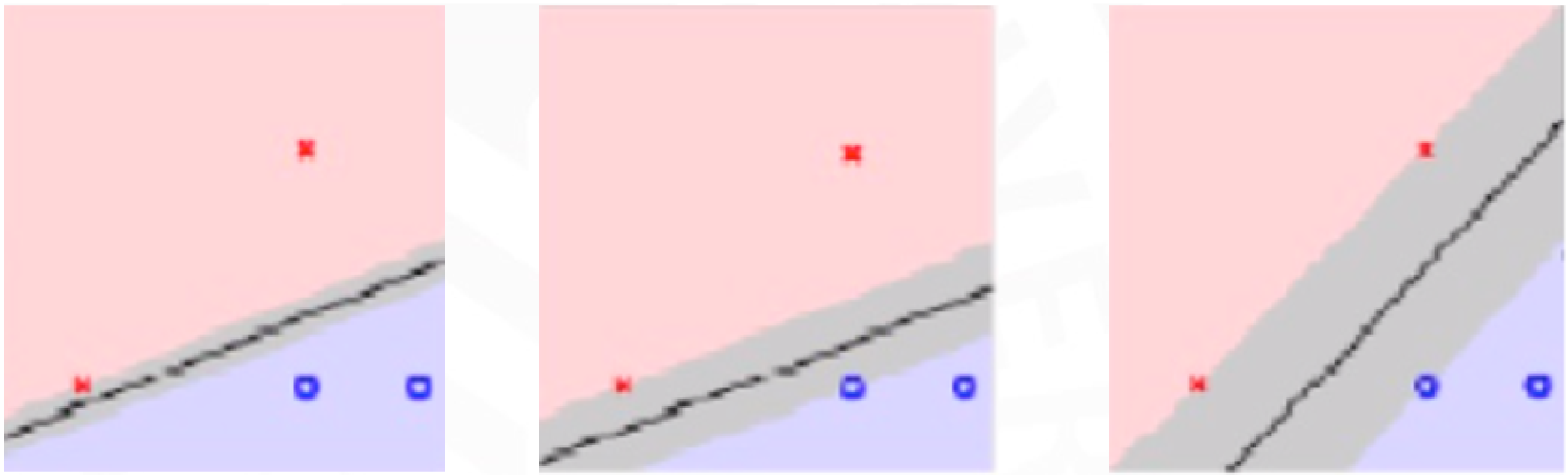
- 感知机算法得到的三条直线:



- 样本点附近的圆形区域越大，表示分类直线对测量数据误差的容忍性越高



- 用圆形区域表示分类线能够容忍多少误差，相当于计算点到直线的距离。



- 基本思想：找到一条最好的直线，离样本点距离足够的大
- 间隔(margin)越大越好

SVM模型建立

- 数据集可以描述为:

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

- 原模型1:

$$\begin{aligned}
\max_{\mathbf{w},b}\quad & margin(\mathbf{w},b)\\
s.t.\quad & \mathbf{w}^T\mathbf{x}_i+b>0,\quad y_i=1\\
& \mathbf{w}^T\mathbf{x}_i+b<0,\quad y_i=-1\qquad (i=1,2,\ldots,N)
\end{aligned}$$

- 原模型2:

$$\begin{aligned}
\max_{\mathbf{w},b}\quad & margin(\mathbf{w},b)\\
s.t.\quad & y_i(\mathbf{w}^T\mathbf{x}_i+b)>0\qquad (i=1,2,\ldots,N)
\end{aligned}$$

- 定义间隔:

$$margin(\mathbf{w}, b) = \min_i \frac{|\mathbf{w}^T \mathbf{x}_i + b|}{\|w\|}$$

- 原模型3:

$$\begin{aligned} &\max_{\mathbf{w},b} \quad \min_i \frac{|\mathbf{w}^T \mathbf{x}_i + b|}{\|w\|} \\ &s.t. \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) > 0 \quad (i = 1, 2, \ldots, N) \end{aligned}$$

- 引理1. 若$(\mathbf{w}^*, b^*)$是原模型3的解，那么对任意的$r > 0$, $(r\mathbf{w}^*, rb^*)$仍是该模型的解

证明:

$$\frac{|r{\mathbf{w}^*}^T\mathbf{x}_i + rb^*|}{\|r\mathbf{w}^*\|} = \frac{|{\mathbf{w}^*}^T\mathbf{x}_i + b|}{\|\mathbf{w}^*\|}$$

$$y_i((r\mathbf{w}^*)^T\mathbf{x}_i + rb^*) > 0 \Leftrightarrow y_i({\mathbf{w}^*}^T\mathbf{x}_i + b^*) > 0 \qquad \square$$

- 则我们约束$(\mathbf{w}, b)$, 使得$\min_i |\mathbf{w}^T\mathbf{x}_i + b| = 1$

- 原模型4：

$$
\begin{aligned}
\min_{\mathbf{w},b} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} \\
s.t. \quad & y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \quad (i = 1, 2, \ldots, N)
\end{aligned}
$$

- 这是一个凸优化问题，也是一个二次规划问题(QP问题)，包含$p+1$个优化变量，N项约束

- 例1. 假设平面上有四个点，两个正类，两个负类：

$$X = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \\ 3 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \quad \begin{array}{cccccc} & & -b & \geq 1 & & (1) \\ -2w_1 & -2w_2 & -b & \geq 1 & & (2) \\ 2w_1 & & +b & \geq 1 & & (3) \\ 3w_1 & & +b & \geq 1 & & (4) \end{array}$$

$$\left\{ \begin{array}{ll} (i)\&(iii) & \Rightarrow w_1 \geq +1 \\ (i)\&(ii) & \Rightarrow w_2 \leq -1 \end{array} \right\} \Rightarrow \frac{1}{2}\mathbf{w}^T\mathbf{w} \geq 1$$

$$w_1 = 1, w_2 = -1, b = -1$$

$$g(\mathbf{x}) = sign(x_1 - x_2 - 1)$$

对偶

- 线性支持向量机的拉格朗日函数为：

$$L(\mathbf{w}, b, \lambda) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{i=1}^{N} \lambda_i (1 - y_i(\mathbf{w}^T x_i + b))$$

- 对偶问题：

$$\begin{aligned} \max_{\lambda} \min_{\mathbf{w},b} \quad & \tfrac{1}{2}\mathbf{w}^T\mathbf{w} + \textstyle\sum_{i=1}^{N} \lambda_i (1 - y_i(\mathbf{w}^T\mathbf{x}_i + b)) \\ s.t. \quad & \lambda_i \geq 0, i = 1, 2, \ldots, N. \end{aligned}$$

- 利用KKT条件中的稳定性条件可得：

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \lambda_i y_i = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i$$

代入对偶问题可得对偶问题2：

$$\begin{aligned} \min_{\lambda} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \textstyle\sum_{i=1}^{N} \lambda_i \\ s.t. \quad & \lambda_i \geq 0, i = 1, 2, \ldots, N. \\ & \sum_{i=1}^{N} \lambda_i y_i = 0 \end{aligned}$$

这也是一个凸优化问题，也是一个二次规划问题(QP问题)，包含N个优化变量，$N+2$项约束

- 其他KKT条件:
  - 主问题可行：$1 - y_i(\mathbf{w}^T\mathbf{x}_i + b) \leq 0$
  - 对偶问题可行：$\lambda_i \geq 0$
  - 互补松弛：$\lambda_i(1 - y_i(\mathbf{w}^T\mathbf{x}_i + b)) = 0$
- 支持向量： 对偶变量$\lambda_i > 0$对应的样本
- 支持向量集SV： 所有支持向量组成的集合
- 引理2. 线性支持向量机中，支持向量是距离划分超平面最近的样本，落在最大间隔边界上。
- 证明：由互补松弛条件可知：当$\lambda_i > 0$时，$1 - y_i(\mathbf{w}^T\mathbf{x}_i + b) = 0$ $\square$

仅利用支持向量$\lambda_i \in SV$, 可计算：$\mathbf{w} = \sum_{i \in SV} \lambda_i y_i \mathbf{x}_i, b = y_i - \mathbf{w}^T \mathbf{x}_i$.

- 线性支持向量机的假设函数可表示为：

$$h(\mathbf{x}) = sign(\sum_{i \in SV} \lambda_i y_i \mathbf{x}_i^T \mathbf{x} + b)$$

- 线性支持向量机的损失函数可表示为：

$$Loss(\mathbf{w}, b) = \sum_i \max(0, 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b))$$