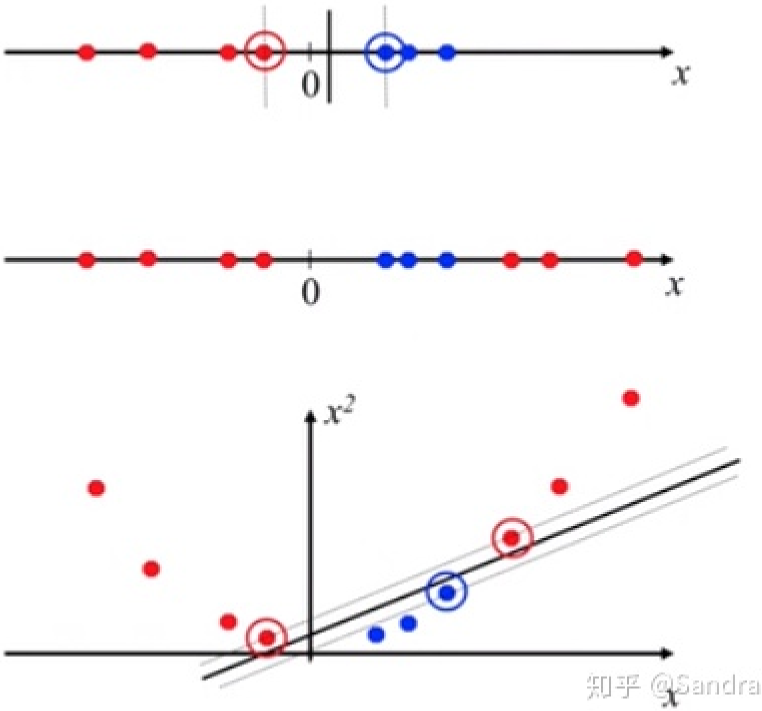
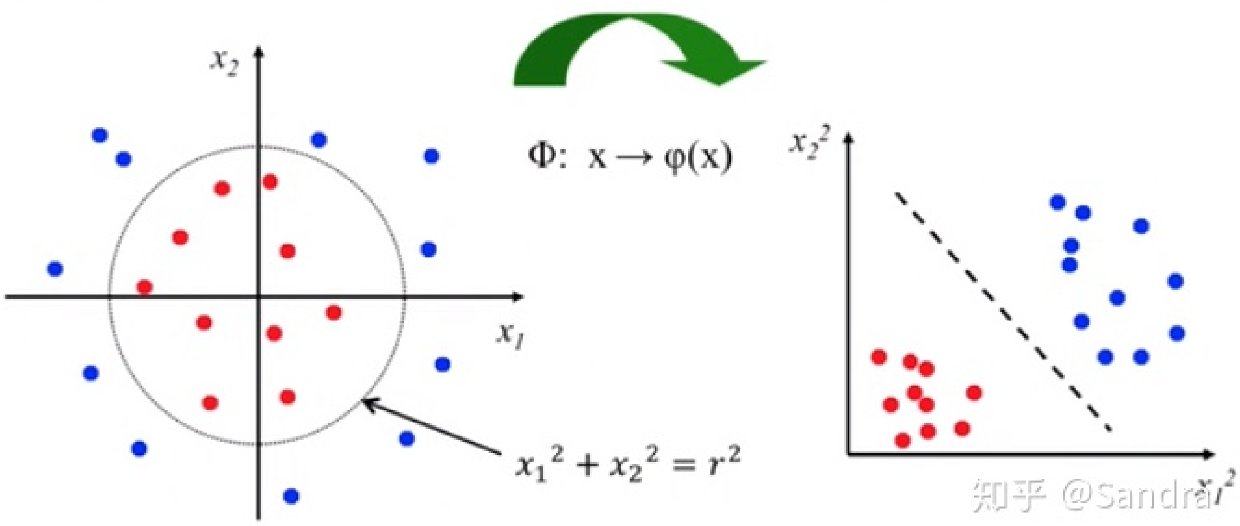
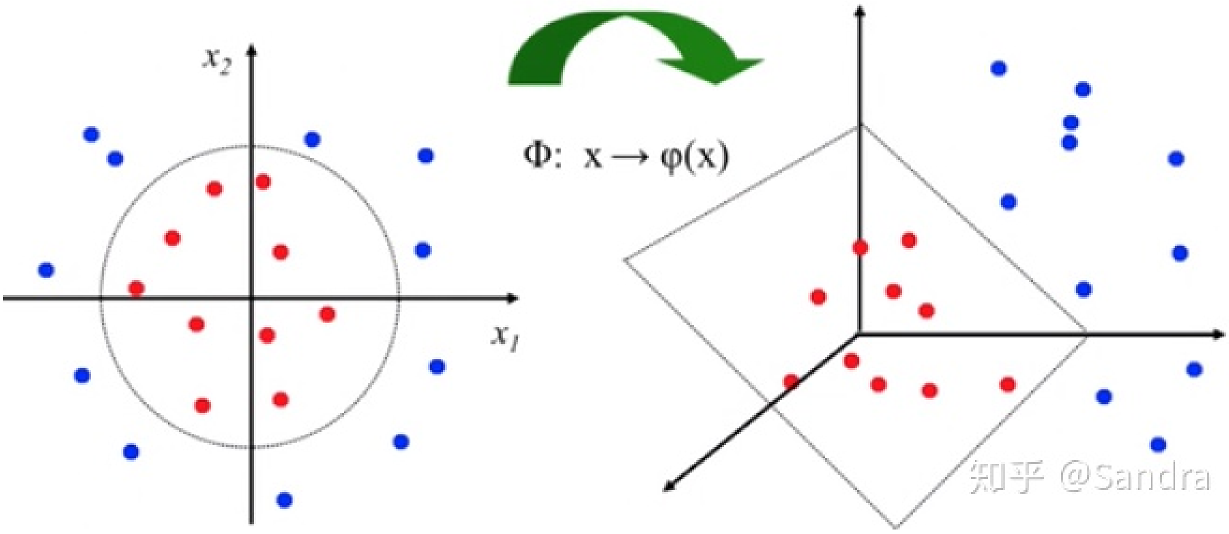
线性不可分问题下的SVM：



基本思想：一维不可分映射为高维度（feature space），映射不唯一。





一般都是用几种固定的应设方法：

$$\Phi(x)=\left[\begin{array}{c} 1 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ \vdots \\ \sqrt{2}x_m \\ x_1^2 \\ x_2^2 \\ \vdots \\ x_m^2 \\ \sqrt{2}x_1x_2 \\ \sqrt{2}x_1x_3 \\ \vdots \\ \sqrt{2}x_1x_m \\ \sqrt{2}x_2x_3 \\ \vdots \\ \sqrt{2}x_2x_m \\ \vdots \\ \sqrt{2}x_{m-1}x_m \end{array}\right]
\begin{array}{l}
\left.\vphantom{1}\right\}\text{Constant Terms} \\
\left.\vphantom{\begin{array}{c}a\\a\\a\\a\end{array}}\right\}\text{Linear Terms} \\
\left.\vphantom{\begin{array}{c}a\\a\\a\\a\end{array}}\right\}\text{Pure Quadratic Terms} \\
\left.\vphantom{\begin{array}{c}a\\a\\a\\a\\a\\a\\a\\a\\a\end{array}}\right\}\text{Quadratic Cross-Terms}
\end{array}$$

从低维映射至高维，通过公式，大概含义例如有一个100维的数据，映射至5000维。

Number of terms

$$C_{m+2}^{2}=\frac{(m+2)(m+1)}{2}\approx\frac{m^2}{2}$$

SVM中大多为向量做内积：

$$\Phi(a)\cdot\Phi(b)=\begin{bmatrix}1\\ \sqrt{2}a_1\\ \sqrt{2}a_2\\ \vdots\\ \sqrt{2}a_m\\ a_1^2\\ a_2^2\\ \vdots\\ a_m^2\\ \sqrt{2}a_1a_2\\ \sqrt{2}a_1a_3\\ \vdots\\ \sqrt{2}a_1a_m\\ \sqrt{2}a_2a_3\\ \vdots\\ \sqrt{2}a_2a_m\\ \vdots\\ \sqrt{2}a_{m-1}a_m\end{bmatrix}\cdot\begin{bmatrix}1\\ \sqrt{2}b_1\\ \sqrt{2}b_2\\ \vdots\\ \sqrt{2}b_m\\ b_1^2\\ b_2^2\\ \vdots\\ b_m^2\\ \sqrt{2}b_1b_2\\ \sqrt{2}b_1b_3\\ \vdots\\ \sqrt{2}b_1b_m\\ \sqrt{2}b_2b_3\\ \vdots\\ \sqrt{2}b_2b_m\\ \vdots\\ \sqrt{2}b_{m-1}b_m\end{bmatrix}$$

$$1$$

$$\sum_{i=1}^{m}2a_ib_i$$

$$\sum_{i=1}^{m}a_i^2b_i^2$$

$$\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}2a_ia_jb_ib_j$$

$$\boxed{x_i\cdot x_j\Rightarrow\Phi(x_i)\cdot\Phi(x_j)}$$

22

Kernel Trick：

高维做内积复杂度是很高的，但是存在一个表达式（核函数，Kernel）：

$$(a\cdot b+1)^2=\phi(a)\cdot\phi(b)$$

，说明低维度该计算公式可以得到与高维度计算得出相同的结果，既保证的提高维度，又降低了计算复杂度。

$$\Phi(a)\cdot\Phi(b)=1+2\sum_{i=1}^{m}a_ib_i+\sum_{i=1}^{m}a_i^2b_i^2+\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}2a_ia_jb_ib_j$$

$$(a\cdot b+1)^2=(a\cdot b)^2+2a\cdot b+1=(\sum_{i=1}^{m}a_ib_i)^2+2\sum_{i=1}^{m}a_ib_i+1$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{m}a_ib_ia_jb_j+2\sum_{i=1}^{m}a_ib_i+1$$

$$=\sum_{i=1}^{m}(a_ib_i)^2+2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}a_ib_ia_jb_j+2\sum_{i=1}^{m}a_ib_i+1$$

$$K(a,b)=(a\cdot b+1)^2=\Phi(a)\cdot\Phi(b)$$

$O(m)$ $O(m^2)$

高维度操作=低维度操作

- The linear classifier relies on dot products $x_i\cdot x_j$ between vectors.
- If every data point is mapped into a high-dimensional space via some transformation Φ: $x\rightarrow\varphi(x)$, the dot product becomes: $\varphi(x_i)\cdot\varphi(x_j)$
- A *kernel function* is some function that corresponds to an inner product in some expanded feature space: $K(x_i,x_j)=\varphi(x_i)\cdot\varphi(x_j)$
- Example: $x=[x_1,x_2]$; $K(x_i,x_j)=(1+x_i\cdot x_j)^2$

$$K(x_i,x_j)=(1+x_i\cdot x_j)^2=1+x_{i1}^2x_{j1}^2+2x_{i1}x_{j1}x_{i2}x_{j2}+x_{i2}^2x_{j2}^2+2x_{i1}x_{j1}+2x_{i2}x_{j2}$$

$$=[1,x_{i1}^2,\sqrt{2}x_{i1}x_{i2},x_{i2}^2,\sqrt{2}x_{i1},\sqrt{2}x_{i2}]\cdot[1,x_{j1}^2,\sqrt{2}x_{j1}x_{j2},x_{j2}^2,\sqrt{2}x_{j1},\sqrt{2}x_{j2}]$$

$$=\Phi(\mathrm{x}_i)\cdot\Phi(\mathrm{x}_j),\quad \text{where }\ \Phi(x)=[1,x_1^2,\sqrt{2}x_1x_2,x_2^2,\sqrt{2}x_1,\sqrt{2}x_2]$$

非常巧妙的一点！

求解w和b：（转到高维空间了，但
$\phi(x)$
其实是不需要求得）

$$w = \sum_{i=1}^{l} \alpha_i y_i \Phi(x_i)$$

$$w \cdot \Phi(x_j) = \sum_{i=1}^{l} \alpha_i y_i \Phi(x_i) \cdot \Phi(x_j) = \sum_{i=1}^{l} \alpha_i y_i K(x_i, x_j)$$

$$b = \frac{1}{N_s} \sum_{s \in S} (y_s - \sum_{m \in S} \alpha_m y_m \Phi(x_m) \cdot \Phi(x_s)) = \frac{1}{N_s} \sum_{s \in S} (y_s - \sum_{m \in S} \alpha_m y_m K(x_m, x_s))$$

$$\boxed{g(x) = \sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b} \Longleftrightarrow g(x) = w \cdot x + b = \sum_{i=1}^{l} \alpha_i y_i x_i \cdot x + b$$

与之前的对照

发现转为高维的求解函数，仍与低维的基本一致。

核函数的强悍！

$$Polynomial: K(x_i, x_j) = (x_i \cdot x_j + 1)^d$$

$$Gaussian: K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

$$Hyperbolic\ Tangent: K(x_i, x_j) = \tanh(\kappa x_i \cdot x_j + c)$$

第一个是我们上面讲过的例子；第二个是高斯核函数：可以展开无数多项，所以可以映射至无穷维。

String Kernel：

文本转数值，转到高维空间

- Calculate the similarity between text strings.
- The substring ‘c-a-r’ is present in both Car and Custard.
- Each substring corresponds to a dimension of the feature space.

	c-a	c-t	a-t	b-a	b-t	c-r	a-r	b-r
$\phi(\text{cat})$	λ^2	λ^3	λ^2	0	0	0	0	0
$\phi(\text{car})$	λ^2	0	0	0	0	λ^3	λ^2	0
$\phi(\text{bat})$	0	0	λ^2	λ^2	λ^3	0	0	0
$\phi(\text{bar})$	0	0	0	λ^2	0	0	λ^2	λ^3

$$K(car, cat) = \lambda^4$$

$$K(car, car) = K(cat, cat) = 2\lambda^4 + \lambda^6$$

包含c-t，但是因为一共三个字母所以是三次方