1. kernel 这个东西在 SVM 中真的只是一少部分，为了面试的话大概了解一下就可以了，面试官也很少有懂的。并不是说大家不爱学习，而是做 kernel 的人根本不关心 SVM 是什么，做 SVM 的人也根本不用关心 kernel 是个什么鬼。

2. **kernel 和 SVM 是两个完全没有关系的概念。**实际上在 SVM 提出以前，人们就提出了再生核希尔伯特空间（reproducing kernel Hilbert space，RKHS）这个概念，并且把它应用在信号处理中。如：在信号检测（signal detection）问题中，对于一条时间序列（time series），我如何知道它是一个随机步行（random walk）的噪音序列呢？还是有一个特定的模式（pattern）在里面呢？在这个情景下，RKHS 理论就给出了一个通过求解似然率（likelihood ratio）的假设检验方案，其中的 kernel 是某个随机过程在两个不同时间点的相关性（correlation）。

   另外，核方法可以用在 逻辑斯谛回归（logistic regression）、最小二乘法（least square）、降维（dimension reduction）等多处地方，也不是只和 SVM 这个概念绑定的。

在SVM中的应用

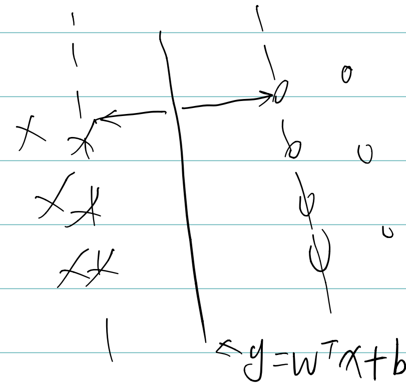
简单说几句，公式太难写了（笑）。

我们在使用原始数据 x 的时候发现数据并不可分，所以就寄希望于一个映射 $\Phi(x)$ ，这个映射把低维空间上的数据映射到高维空间，这样数据集就有可能变得可分了。

但是在考虑优化问题的对偶问题时，需要计算 $\langle x_i, x_j \rangle$ ，请注意到，我们已经把所有的 x 换成了 $\Phi(x)$ ，所以就变成需要计算 $\langle \Phi(x_i), \Phi(x_j) \rangle$ 。
为了不让计算变得很困难，我们就可以找到一个核函数 K ，满足 K 可以生成 Φ 所形成的高维空间，这样 $\langle \Phi(x_i), \Phi(x_j) \rangle$ 就可以简单的用 $K(x_i, x_j)$ 代替了。而 K 往往定义成和 x 的内积有关的式子，这样在低维空间中计算内积就很简单。

如：径向基函数里有 $\|x-y\|^2$ ，展开以后其实就含有两个范数项（注意范数就是内积）和一个内积项。

SVM



$$\begin{cases} w^T x_1 + b = 1 \\ w^T x_2 + b = -1 \end{cases}$$

$$w^T(x_1 - x_2) = 2$$

$$\underbrace{\frac{w^T}{\|w\|}}_{\text{投影}}(x_1 - x_2) = \frac{2}{\|w\|} \leftarrow 间隔$$

求 $\max \frac{2}{\|w\|}$

问题 $\min \frac{1}{2} \cdot w^T w$

s.t. $y_i(w x_i + b) \geq 1$

约束化为等式 $1 - y_i(w^T x_i + b) + a_i^2 = 0$

对偶问题: $\max_{\alpha} W(\alpha)$

其中: $W(\alpha) = \min(\frac{1}{2} w^T w) + \sum \alpha_i [1 - y_i(w^T x_i + b) + a_i^2]$

对于 $L = \frac{1}{2} w^T w + \sum \alpha_i [1 - y_i(w^T x_i + b) + a_i^2]$

$$\begin{cases} \frac{\partial L}{\partial w} = w + \sum(\alpha_i y_i x_i) = 0 \Rightarrow w = \sum \alpha_i y_i x_i \\ \frac{\partial L}{\partial a_i} = 2\alpha_i a_i = 0 \Rightarrow \alpha_i a_i = 0 \\ \frac{\partial L}{\partial b} = \sum \alpha_i (-y_i) = 0 \Rightarrow \sum \alpha_i y_i = 0 \end{cases}$$

$w^T w = (\sum \alpha_i y_i x_i)^T (\sum \alpha_j y_j x_j) = \sum \alpha_i \alpha_j y_i y_j x_i^T x_j$

$w^T x_i = \sum \alpha_j y_j x_j^T x_i$

$W(\alpha) = \frac{1}{2} \sum \sum \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum \alpha_i - \sum \alpha_i y_i \sum \alpha_j y_j x_j^T x_i + \underbrace{\sum \alpha_i b y_i}_{=0} + \underbrace{\sum \alpha_i a_i^2}_{=0}$

$= \sum \alpha_i - \frac{1}{2} \sum \sum \alpha_i \alpha_j y_i y_j x_i^T x_j$

核SVM

现在问题为 $\max_{\alpha} W(\alpha) = \sum \alpha_i - \frac{1}{2}\sum\sum \alpha_i\alpha_j y_i y_j x_i x_j$

s.t. $\sum \alpha_i y_i = 0$

$\alpha_i \geq 0$

$\downarrow$ 核化

$$\max W(\alpha) = \sum \alpha_i - \frac{1}{2}\sum\sum \alpha_i\alpha_j y_i y_j K(x_i, x_j)$$

SV • 许多 α_i^* 为0,

支持向量:

$$w = \sum y_i \alpha_i x_i$$

$$y = w^T x + b = \sum y_i \alpha_i x_i^T x + b$$

核化: $y = \sum y_i \alpha_i K(x_i, x) + b$

b • 在支持向量有

$$w^T x + b = y$$

$$\sum y_i \alpha_i x_i^T x + b = y \Rightarrow b^* = y - \sum y_i \alpha_i x_i^T x$$

Prediction.

- $y_{new} = sign(\Sigma \alpha_i y_i K(x_i, x_{new}) + b)$