支持向量机：

与逻辑回归和神经网络相比，支持向量机，或者简称**SVM**，在学习复杂的非线性方程时提供了一种更为清晰，更加强大的方式。
SVM是通过构建超平面将数据分割开来。
对于可以直接线性分割的数据集，可以直接构建下图所示的超平面。

$$w^T x + b = 0$$

此时为了衡量分类器的好坏，我们采用函数间隔来衡量样本点距离超平面的距离。
函数间隔：

$$\gamma = y(w^T x + b) = yf(x)$$

其中只要分类正确就为正，而且只需要考虑离超平面最近的点，若w,b同时乘以2，虽然它的超平面没有变，但是它的函数间隔却变大了。
因此，我们引入几何间隔这个概念，即对w加一定的约束条件。
几何间隔：

$$\gamma' = \frac{\gamma}{||w||}$$

实际上，我们希望分类的效果尽可能的准确，也就是超平面距最近点的距离越远越好。

Hinge Loss:

$$L_i = \sum_{j \neq t_i} max(0, f(x_i, W)_j - (f(x_i, W)_{y_i} - \triangle))$$

- 对于训练集中的第i个数据xi
- 在W下会有一个得分结果向量f(xi,W)
- 第j类的得分为我们记作f(xi,W)j

$$f(x_i, W)_j$$

是错误分类的得分，后一项为正确得分减去一个Δ，我们把后一项称作警戒线，很明显我们希望分类差距越大越好，也就是说，当预测错误得分超过警戒线时，会得到一个惩罚值，这样只要循环训练，直到最小化损失函数，就找到了超平面。

目标函数(令函数间隔=1)：

$$max\frac{1}{||w||}s.t., y_i(w^Tx_i+b) \geq 1, i=1,...,n$$

一般来讲，我们对于极大值问题，都通过转化为其对偶形求解。

对偶问题：

$$min\frac{1}{2}||w||^2 s.t., y_i(w^Tx_i+b) \geq 1, i=1,...,n$$

此外，由于这个问题的特殊结构，还可以通过拉格朗日对偶性（Lagrange Duality）变换到对偶变量 (dual variable) 的优化问题，即通过求解与原问题等价的对偶问题（dual problem）得到原始问题的最优解，这就是线性可分条件下支持向量机的对偶算法，这样做的优点在于：一者对偶问题往往更容易求解；二者可以自然的引入核函数，进而推广到非线性分类问题。

但是现实中数据往往是线性不可分的，我们需要找到非线性的超平面，核函数可以有效的帮我们把数据映射到高维空间，

核函数的本质是两个函数的内积，通过核函数将其隐射到高维空间，在高维空间非线性问题转化为线性问题, SVM得到超平面是高维空间的线性分类平面。其分类结果也视为低维空间的非线性分类结果, 因而带核的SVM就能分类非线性问题。

核函数的选择：

- 如果特征的数量大到和样本数量差不多，则选用LR或者线性核的SVM；
- 如果特征的数量小，样本的数量正常，则选用SVM+高斯核函数；
- 如果特征的数量小，而样本的数量很大，则需要手工添加一些特征从而变成第一种情况。