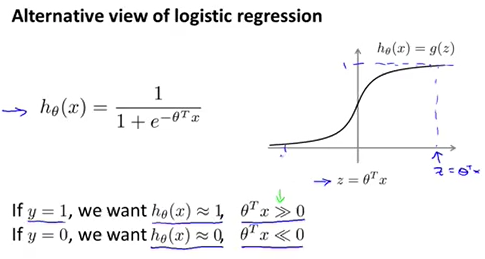
# 支持向量机

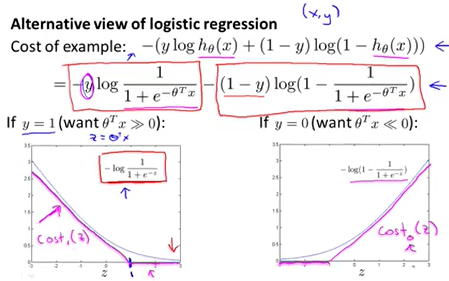
## 优化目标

从逻辑回归一点点修改来得到本质上的支持向量机。

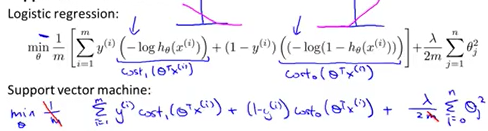
逻辑回归模型：



逻辑回归损失函数，为了构建支持向量机，用紫红色的损失函数来代替逻辑回归损失函数（一个样本）。



用最新的损失函数代替逻辑回归的损失函数得到支持向量机的损失函数，损失函数最小化得到模型参数θ：

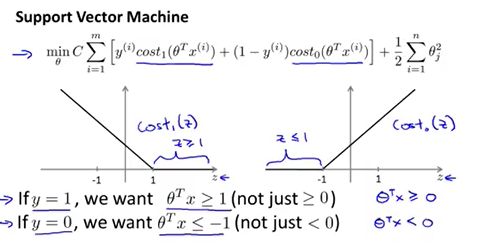


损失函数最小化可以等效为该损失函数：

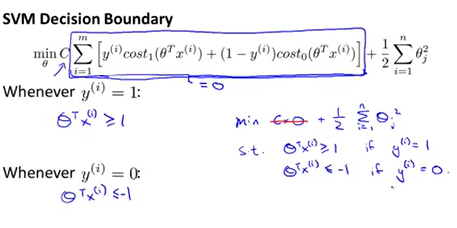


## 大间距分类器

支持向量机损失函数设定的条件为1和-1，即相当于增加了一个额外的安全因子。



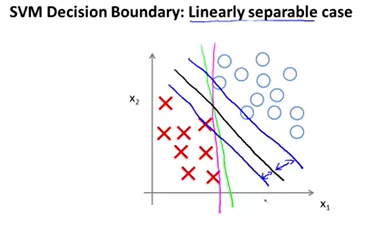
考虑C无穷大情况，SVM决策边界：



当求解上个优化问题，你会得到有趣的决策边界：

**SVM决策边界：线性可分**

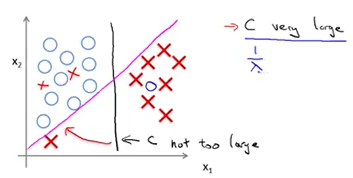
如下图，支持向量机的决策边界为一条黑色直线



从数学上来讲，黑色直线有更大的间距，这个距离叫做支持向量机的间距，这是支持向量机具有鲁棒性的原因，用最大间距分离训练样本。

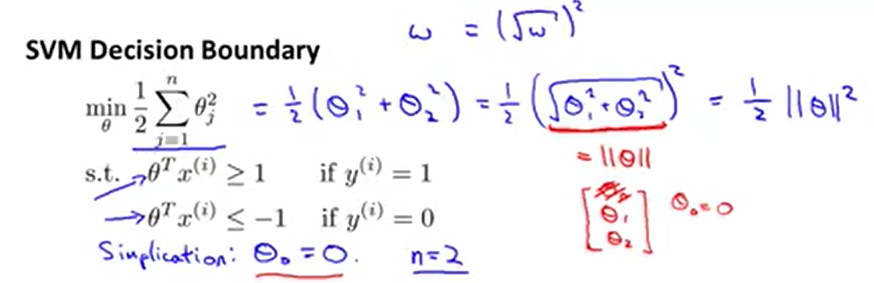
**SVM决策边界：存在异常点**

若C设置的非常大时，则会影响SVM的决策边界，若想把异常点对决策边界的影响降到最低，则C应设置的较小一点（下图，红色线为C很大的决策边界，黑色线为C较小的决策边界）

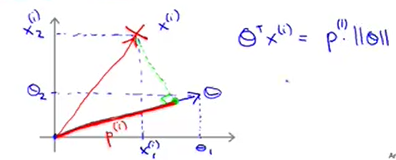


## 大间距分类器的数学推导

SVM决策边界：



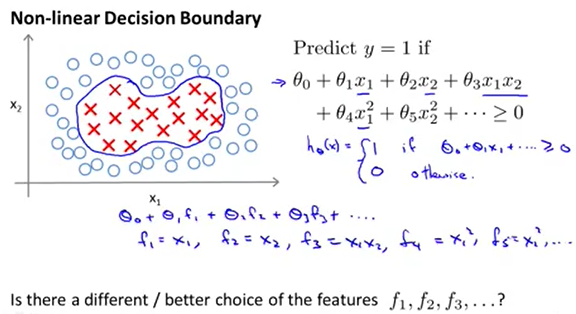
损失函数的向量内积意义：



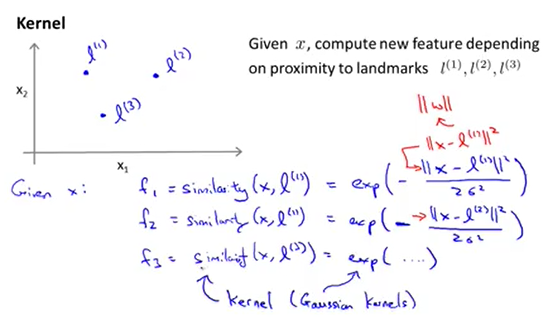
因此，最大决策边界的法向量是θ，损失函数最小化等于最大化样本在决策边界的投影，即间距。

## 核函数

核函数可以用来进行非线性分类，非线性决策边界：

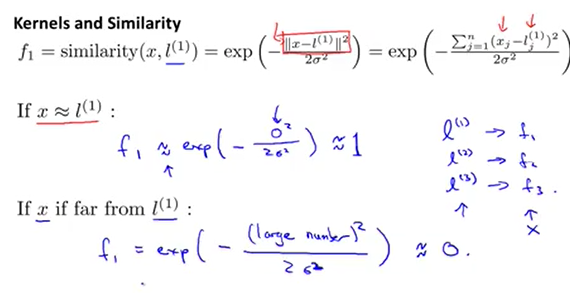


定义f1，f2，f3新变量的定义为与样本的相似度，相似度函数即核函数。

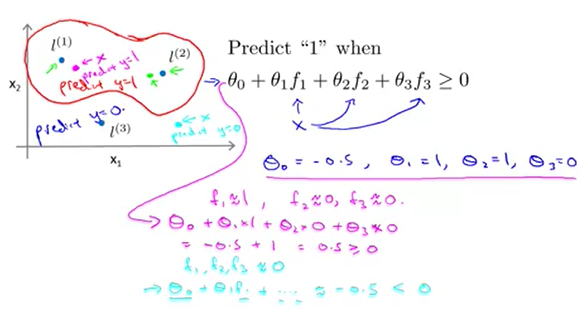


**核函数和相似度：**

给出一个训练样本能够计算与其他训练样本的核函数。

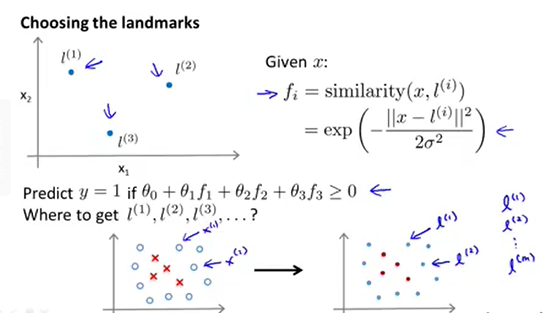


例：对每一个训练样本计算与三个标记点的相似度函数，θ参数给定的情况下，可以得到非决策边界。

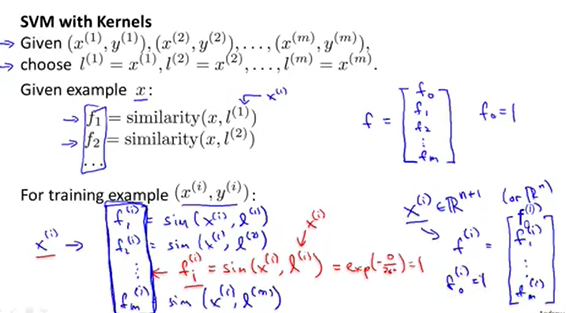


## 核函数（二）

上节核函数并没有给出怎么得到标记点，怎么选择标记点是机器学习中必须解决的问题，下图把训练样本都当成标记点，特征函数基本是在描述样本与样本间的距离。



因此，支持向量机的核函数：



上图x可以是训练样本，也可以是验证数据集样本和测试样本。

因此支持向量机的模型构建过程：