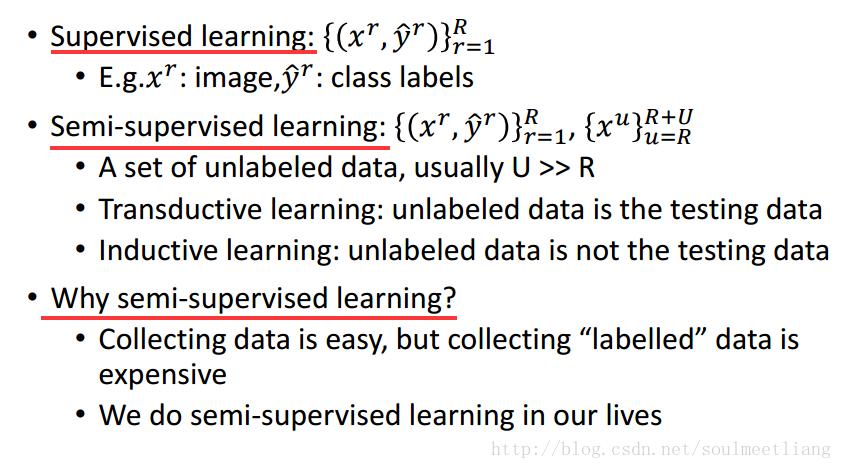
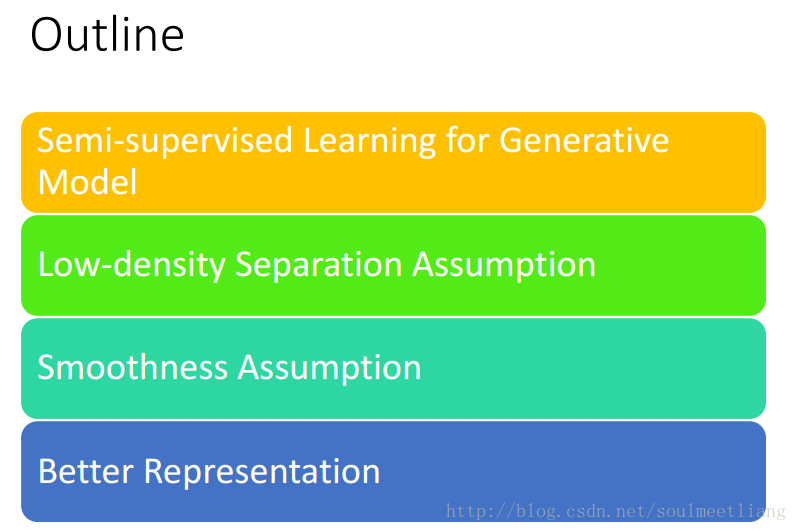
Semi-supervised Learning



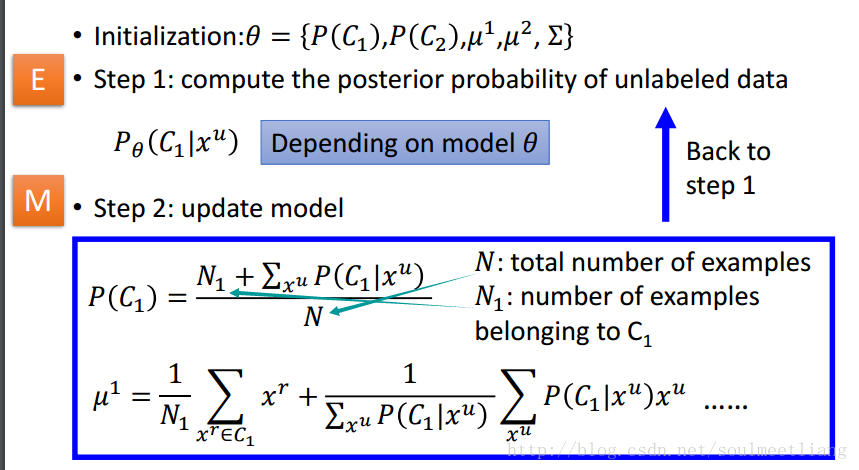
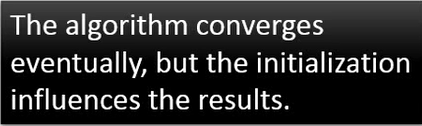
Transductive learning :unlabeled data is the testing data，但并不是cheat，因为用的是testing set的feature，而不是label.

Inductive learning:只用labeled data,因为事先不知道testing data的feature.



**Semi-supervised learning的好坏是基于假设的正确性的**

## Generative Mode



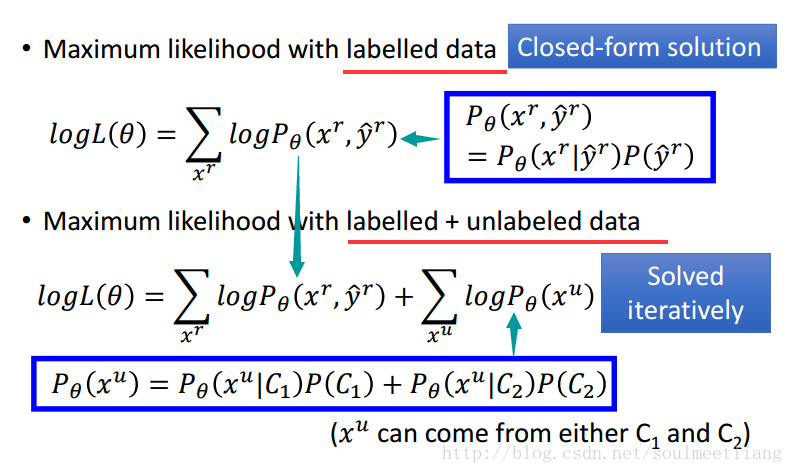
平均值

第二步，通过初始化的参数算出unlabeled data中c1的概率。

第一步，Random得到或者通过labeled data初始化

Posterior probability

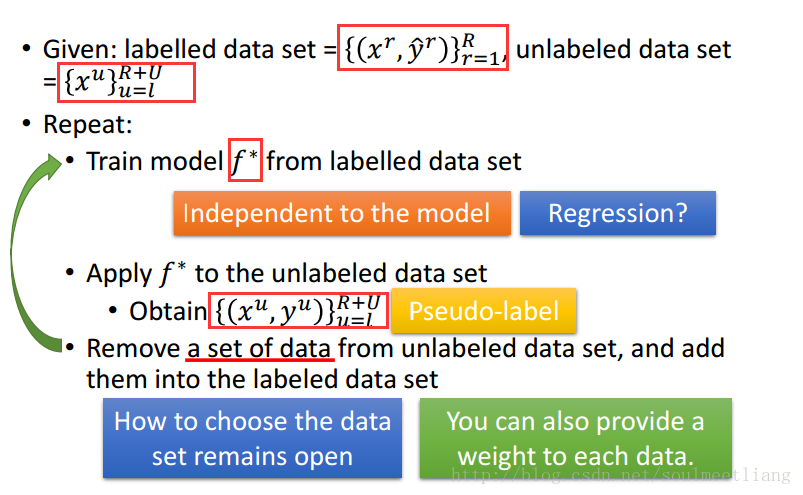
在相同条件下进行*n*次重复试验，如果[随机事件](https://baike.baidu.com/item/%E9%9A%8F%E6%9C%BA%E4%BA%8B%E4%BB%B6)***A***发生的次数为*m*，那么*m/n*称为随机事件***A***的频率**（frequency）**；当试验重复数*n*逐渐增大时，随机事件***A***的频率越来越稳定地接近某一数值*p*，那么就把*p*称为随机事件***A***的概率。这样定义的概率称为统计概率（**statistics probability**），或者称[后验概率](https://baike.baidu.com/item/%E5%90%8E%E9%AA%8C%E6%A6%82%E7%8E%87)**（posterior probability）**。



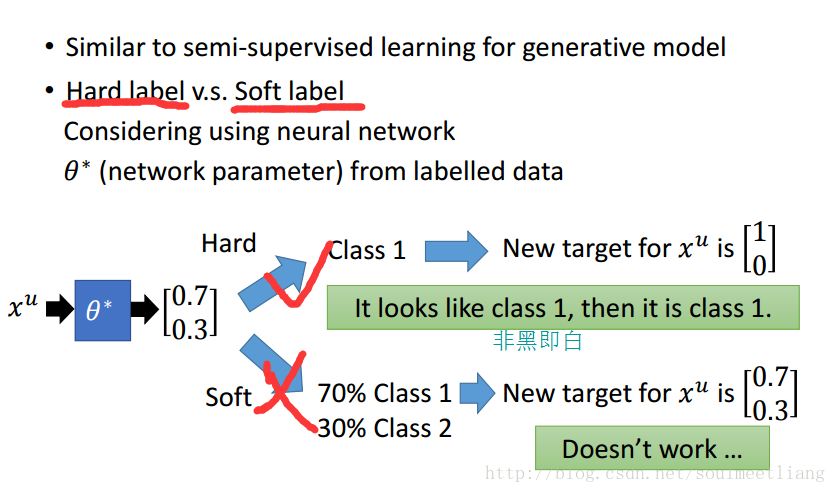
c1出现xu的几率，加上c2出现xu的几率

## Low-density separation assumption

Self-training



Self-training hard, generative model soft



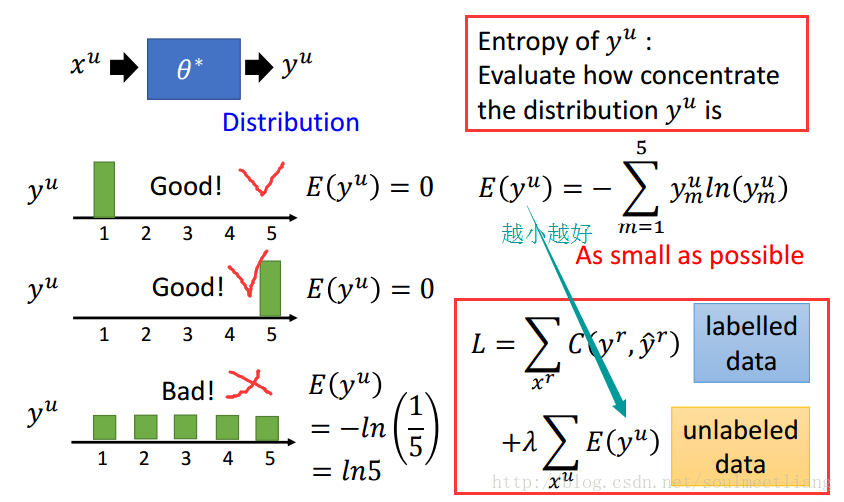
Generative model之所以是soft，是因为其在计算概率的时候的结果可能是0.7、0.3或者其他概率并不是非黑即白的，只不过人为的将P>0.5的归为class 1、class 2

Logistic regression则是y^=1/0

更好的方法：

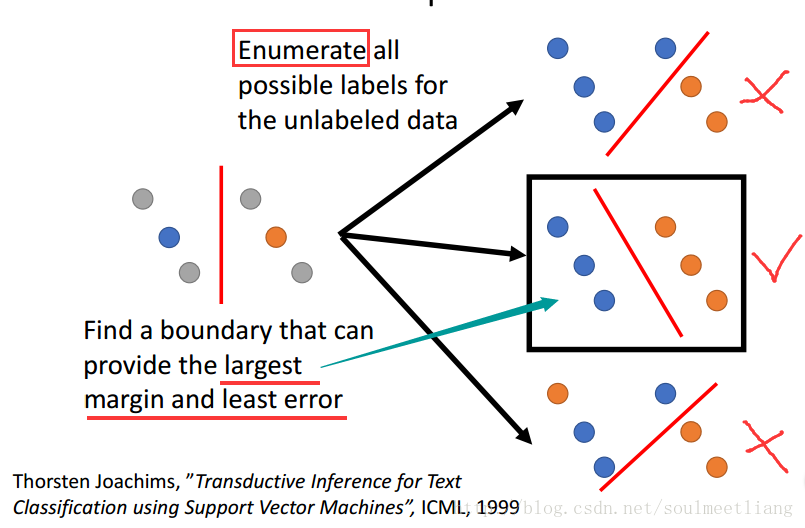
Entropy-based regularization（使class 1、与class 2 集中分布）

像加了一个regularization item，所以叫regularization。



## Outlook: Semi-supervised SVM

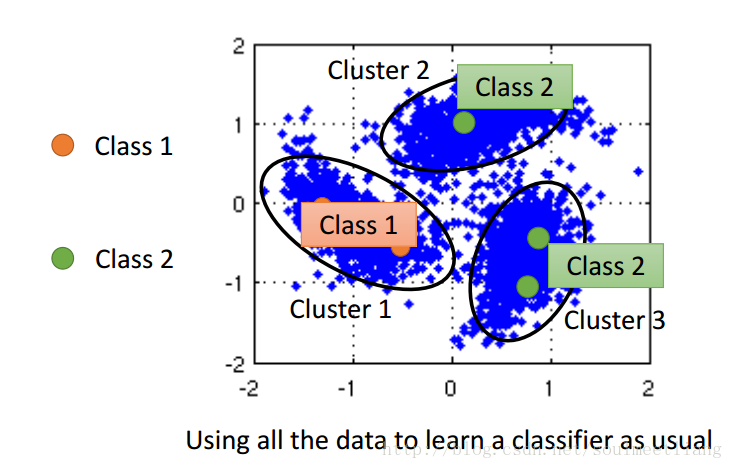
穷举所有unlabeled数据的分类的可能



Seems impossible，如果有一万笔unlabeled data，就有2^10000种可能。

深入学习，后续会讲。

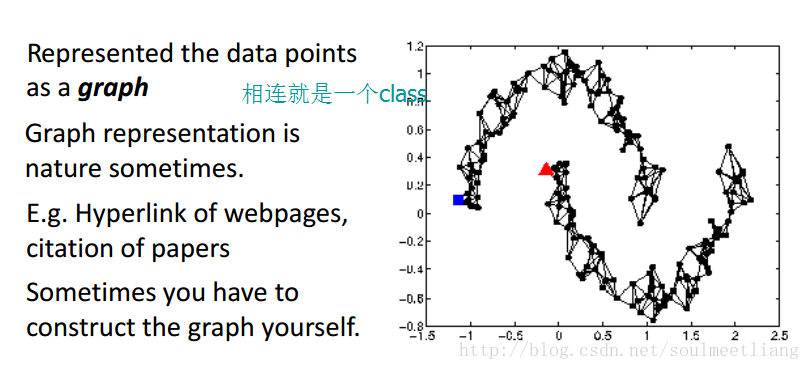
# Smoothness Assumption



这种方法不一定made sense ，需要class很强。   
But，How to know x1 and x2 are close in a high density region (connected by a high density path)   
还有另一种方法：

### Graph-based Approach

**Graph-construction**

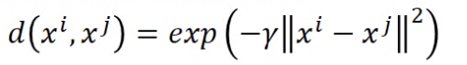


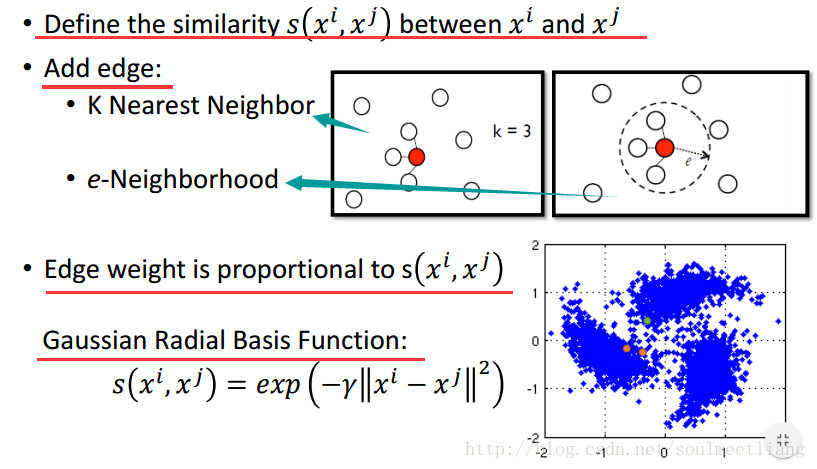
**Gaussian Radial basis function（高斯径向基函数）**

1）radial：表示输入点与center点的距离

2）basis function：表示‘combined’

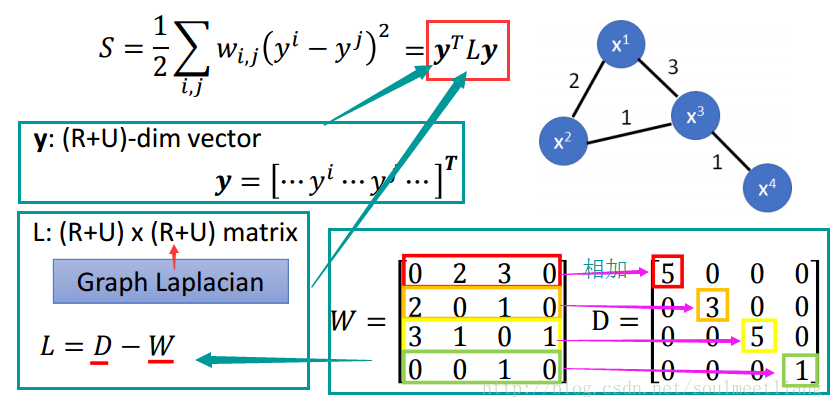
最常用的径向基函数是高斯核函数 ,形式为 k(||x-xc||)=exp{- ||x-xc||^2/(2\*σ^2) }

即，其中xc为核函数中心点,σ为函数的宽度参数 , 控制了函数的径向作用范围（||…||norm运算）。



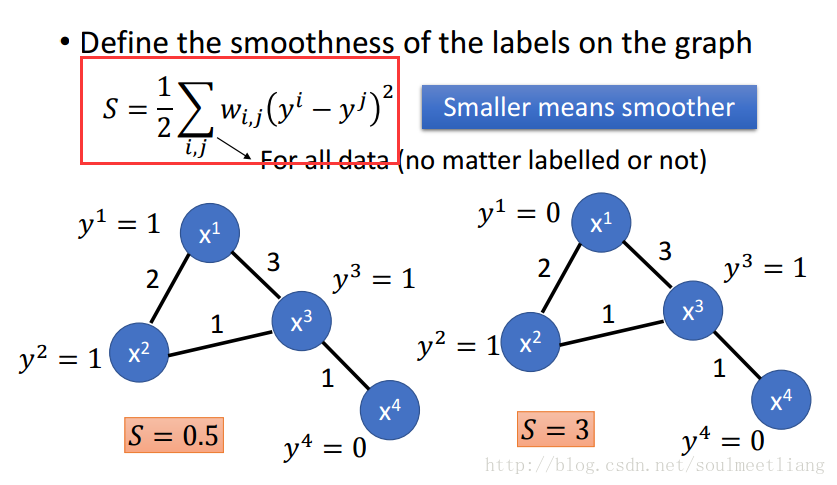
**下降速度会非常快，所以会有比较好的performance**

**怎样在Graph 中定量地表示平滑度**



所有点两两参与运算，W=∑wij，相似度

weight edge,代表相似性



Y是贴上的label

展开运算两式就会相等

**Graph Laplacian（图的拉普拉斯算子、拉普拉斯矩阵）**

其各个点之间的都有相应的边连接，我们用某个指标（这地方可以任意选择，比如欧氏距离、测地距离、或者高斯相似度等）来衡量两个点的相似度，表示为W=∑wij(高斯径向基函数)没有边连接的其相似度自然为零，W是个对称矩阵；某个点的与所有点的相似度之和，表示为D=diag(d);d=rowSum(W)；D是个对角阵；我们的拉普拉斯矩阵则是L=D−W

