Dataset有一半的数据有标签1，另一半无标签

1.忽视label，直接对dataset聚类，得到的聚类模型用在evalset上

缺点：对于聚类，无法衡量其准确度

2.忽视label，直接对dataset聚类得到的1，0标签添加给dataset。用分类器训练已经得到标签的trainset，作用分类器于evalset

缺点：对于聚类，无法衡量其准确度

3.对dataset中无标签的数据进行聚类，得到label。使用分类器训练有dataset。应用分类器到evalset上

1.对无标签数据进行聚类（2聚类），得到两个的聚类中心

2.计算label=1的聚类中心。

3.将无标签的两个聚类中心与label=1的聚类中心计算距离，

4.距离label=1的聚类中心进的cluster中的数据点与label=1的聚类合并

5.距离label=1的聚类中心进的cluster中的数据点标记label=0

检验聚类器好坏：

用得到的聚类器去对原本label=1的数据聚类，理想情况这些数据应该被聚为1类。

4.Positive-Unlabel-learning two step

Step 1. identifying reliable negative examples

1.1计算label=1的聚类中心

1.2计算所有点到label=1的聚类中心的距离

1.3选出距离label=1聚类中心最远的那些点 //个数未定，1000？？

1.4把这些距离label=1聚类中心最远的点标记为label=0，这些点就是初始的reliable negative examples

Step 2: training a classifier

2.1对已经有label的数据（label=1和reliable negative examples）使用classfier\_i训练，把得到的classfier\_i应用到没有label的数据，**找出新的reliable negative examples（只找reliable negative examples不用**classfier\_i找label=1的sample**）**

2.2反复迭代步骤2.1直到：

2.2.1达到最大迭代次数

2.2.2 model的w不再变化 //如果是树，没有w，那么就是数据点的label不再变化

2.3将最终得到的classifier\_n应用到evalset上

**4-2变种.Positive-Unlabel-learning two step**

**Step 1. identifying reliable negative examples**

**1.1将未分类的数据全部默认为0**

**1.2使用一个Isolation Forset训练出classifer**

**1.3用得到的classifier对未标记数据进行predict\_prob**

**1.4 对未标记数据的predict\_prob排序，选出predict\_prob最小的init\_num个作为solid negative**

**Step 2: training a classifier**

**2.1对已经有label的数据（label=1和reliable negative examples）使用classfier\_i训练，把得到的classfier\_i应用到没有label的数据，找出新的reliable negative examples（只找reliable negative examples不用classfier\_i找label=1的sample）**

**2.2反复迭代步骤2.1直到：**

**2.2.1达到最大迭代次数**

**2.2.2 model的w不再变化 //如果是树，没有w，那么就是数据点的label不再变化**

**2.3将最终得到的classifier\_n应用到evalset上**

//舍弃的方法

//3.随机对dataset中无标签的分配标签，使用分类器训练dataset。应用分类器到evalset上

Semi-supervised learning

//4.默认dataset中无标签的为负类，使用分类器训练dataset。应用分类器到evalset上

4-4 to 4-7

猜特征，Featureshold

猜测分布，初始solid negative

4-7 to 4-10