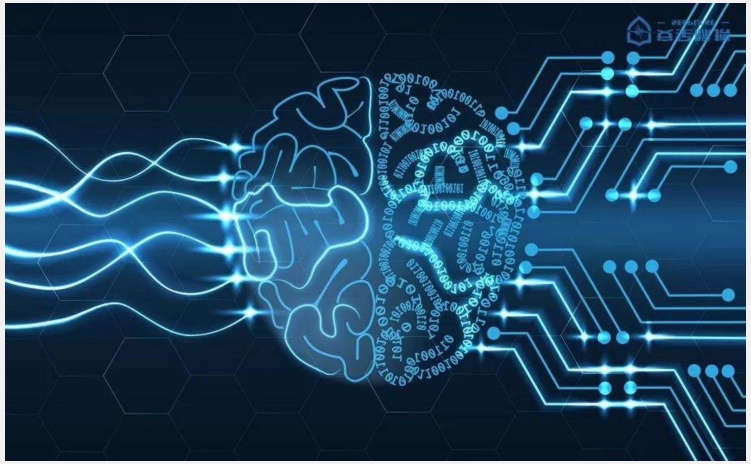


机器学习

Machine Learning

授课老师：谭毅华
电　　话：13886021197
办 公 室：科技楼1102
邮　　箱：yhtan@hust.edu.cn

第十一章、半监督学习

目录
CONTENTS

为人师表

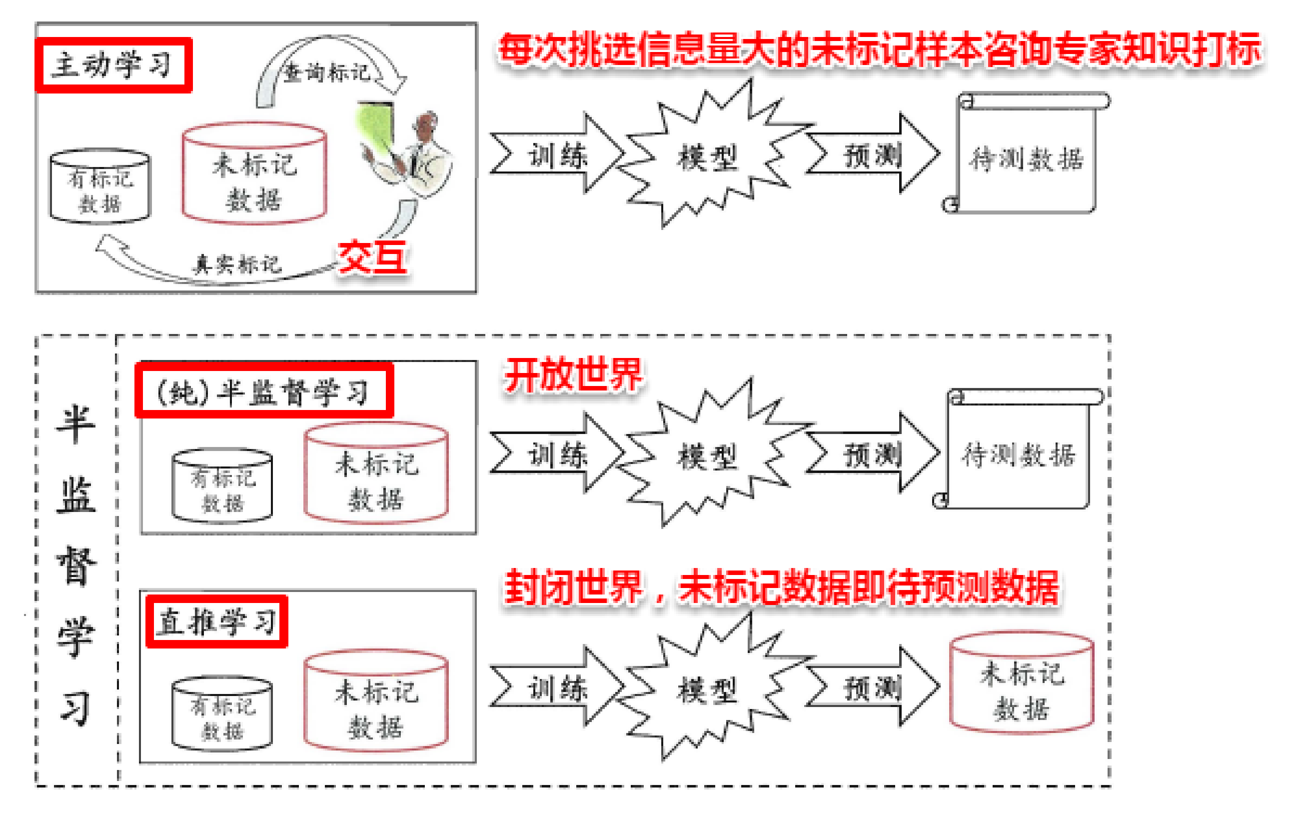
1.半监督学习问题

□ **问题的提出**：在有标签样本较少时，如何利用无标签样本提升学习性能

有监督的学习：学习器通过对大量有标记的训练例进行学习，从而建立模型用于预测未见示例的标记(label)。很难获得大量的标记样本。

无监督的学习：无训练样本，仅根据测试样本的在特征空间分布情况来进行标记，准确性差。

半监督的学习：有少量训练样本，学习机以从训练样本获得的知识为基础，结合测试样本的分布情况逐步修正已有知识，并判断测试样本的类别。



□ **主动学习**：需要与外界进行交互/查询/打标，其本质上仍然属于一种监督学习

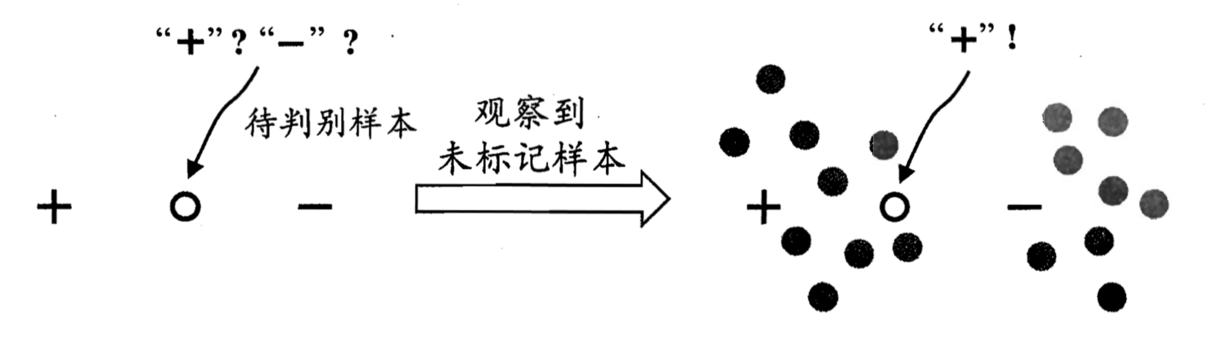
- **半监督学习分为两类**

➢ **纯半监督学习**：假定训练数据集中的未标记数据并非待预测数据

➢ **直推式学习**：学习过程中的未标记数据就是待预测数据

1.半监督学习问题

训练集含有标记样本$D_l=\{(x_1^l,y_1^l),(x_2^l,y_2^l),\dots,(x_n^l,y_n^l)\}$和未标记样本$D_u=\{x_1^u,x_2^u,\dots,x_m^u\}$



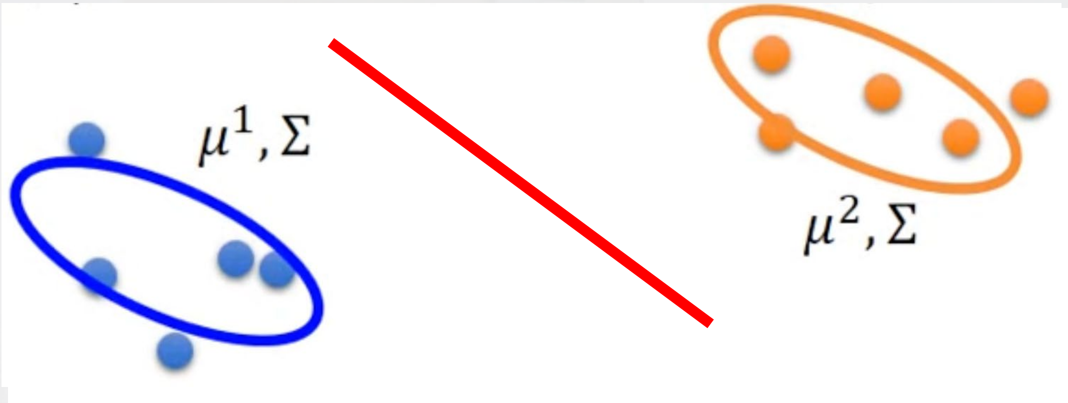
标注数据少，无法判断！ 加入无标签数据后，能够判断！

◘ 半监督学习算法的成立依赖于以下假设：

- **聚类假设**：同一聚类中的样本点很可能具有同样的类别标记。关注样本空间的整体特征，探测样本分布稠密和稀疏的区域，从而更好地约束决策边界穿过稀疏区域
- **流形假设**：高维中的数据存在着低维的特性。利用大量的无标签样增加样本空间的密度，从而更准确地获取样本的局部近邻关系，是聚类假设的推广
- **平滑假设**：相似的样本具有相同的标签。

2.1.生成式方法

- **有监督GMM**：假定样本的概率由多个高斯分布组合形成，从而一个子高斯分布就代表一个类别。（假设各子高斯分布协方差相同）



$$p(x|C_i)=N(x|\mu_u,\Sigma)$$

$$p(C_i|x)=\frac{p(x|C_i)p(C_i)}{\sum_{i=1}^{N}p(x|C_i)p(C_i)}$$

$$C(x)=argmax\ \ p(C_i|x)$$

有标签数据的对数似然：

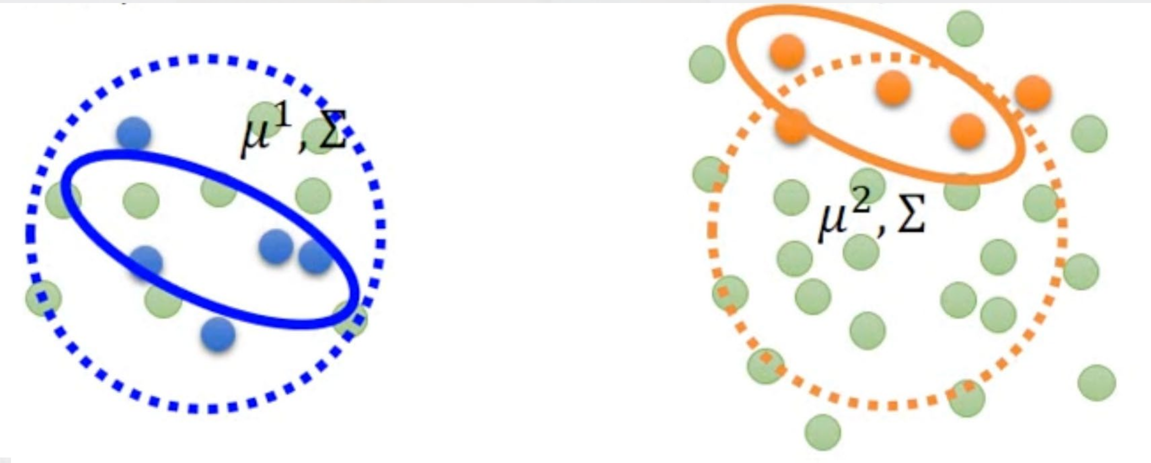
$$logL(\theta)=\sum_{x^r}logP_\theta(x^r,\hat{y}^r)$$

$$P_\theta(x^r,\hat{y}^r)=P_\theta(x^r|\hat{y}^r)P(\hat{y}^r)$$

使用EM算法最大化对数似然

2.1.生成式方法

□ 半监督GMM： 无标签数据使我们需要重新估计高斯分布的参数



有标签数据与无标签数据的对数似然：

$$logL(\theta) = \sum_{x^r} logP_\theta(x^r, \hat{y}^r) + \sum_{x^u} logP_\theta(x^u)$$

$$P_\theta(x^r, \hat{y}^r) = P_\theta(x^r|\hat{y}^r)P(\hat{y}^r)$$

$$P_\theta(x^u) = P_\theta(x^u|C_1)P(C_1) + P_\theta(x^u|C_2)P(C_2)$$

使用EM算法最大化对数似然

2.1.生成式方法

半监督GMM：参数估计迭代计算步骤

- 初始化参数： $\theta = \{P(C_1), P(C_2), \mu^1, \mu^2, \Sigma\}$

E

- 计算无标签数据的后验概率： $P_\theta(C_1|x^u)$

M

- 更新模型参数：

$$P(C_1) = \frac{N_1 + \sum_{x^u} P(C_1|x^u)}{N}$$

N_1：有标签数据中属于C_1的样本数量
N：总样本数量
$\sum_{x^u} P(C_1|x^u)$：无标签样本属于C_1的概率之和

$$\mu^1 = \frac{1}{N_1}\sum_{x^r \in C_1} x^r + \frac{1}{\sum_{x^u} P(C_1|x^u)}\sum_{x^u} P(C_1|x^u)x^u \quad \ldots\ldots$$

(更新μ_i, Σ_i, $P(C_i)$)

2.2.Self-Training

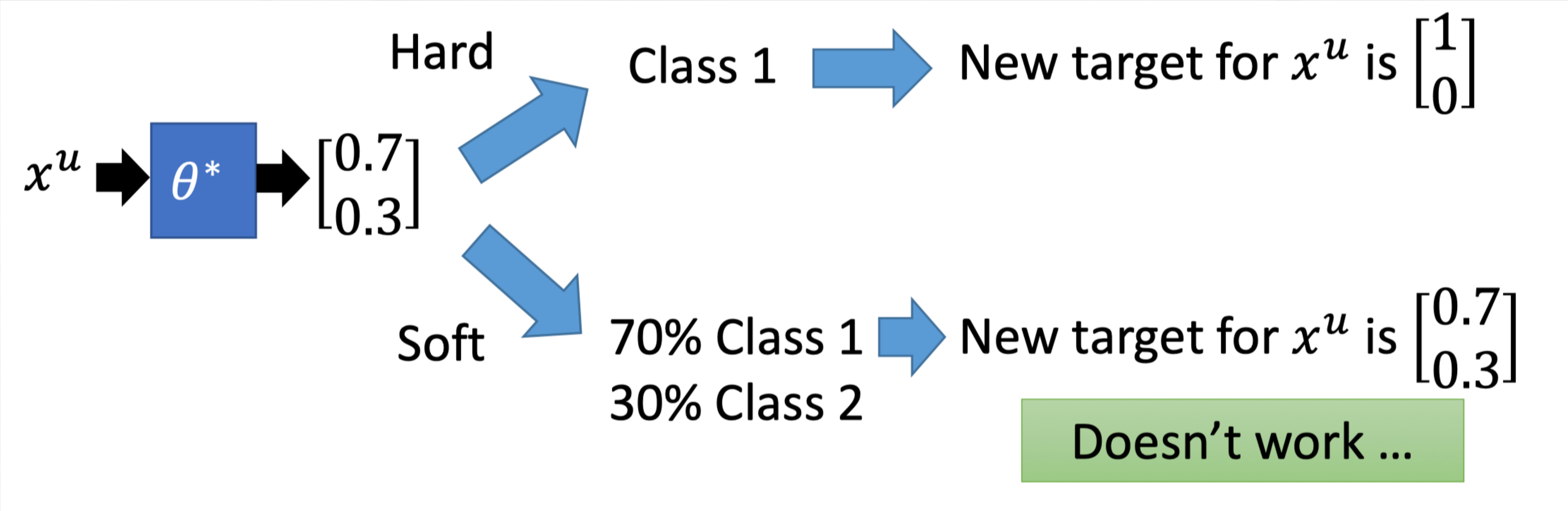
- 给定有标签集$\{(x^l, y^l)\}$，无标签集$\{x^u\}$
- Repeat：
  - 在有标签集上训练模型f^* （方法与模型无关）
  - 在无标签集上应用模型，计算无标签样本的伪标签y^u
  - 从无标签集中移除一部分样本，将其加入有标签集
- 如何选择将哪些样本加入有标签集，哪些样本仍留在无标签集?

  ——为每个样本提供一个权重

- Self-Training 无法用于回归任务

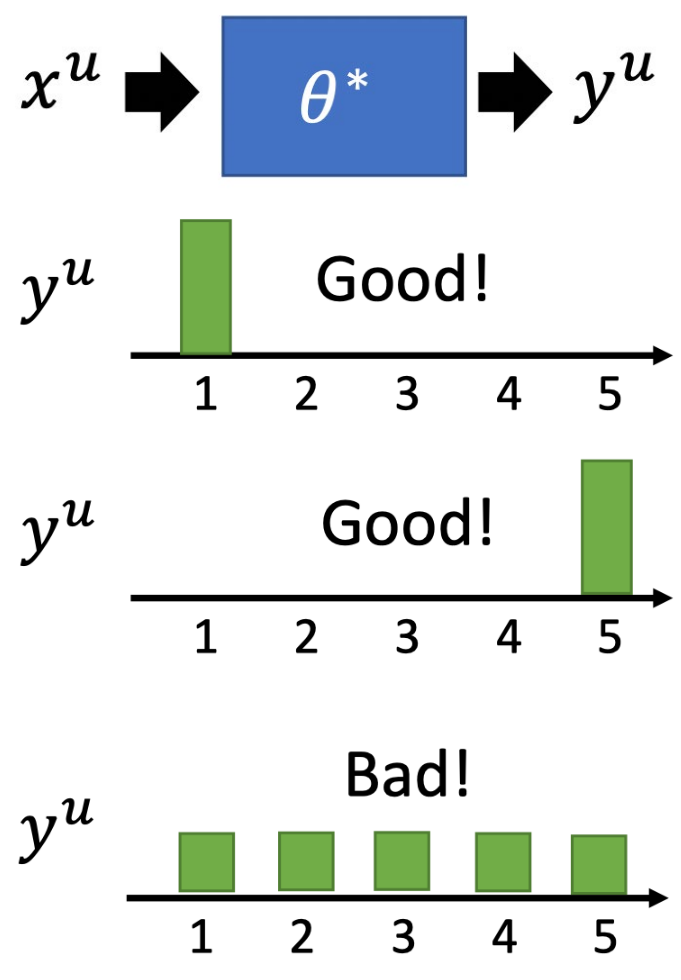
2.2.Self-Training

- 思想类似于生成式方法
- 思考：伪标签 Soft Label v.s. Hard Label
- 考虑神经网络模型，网络参数θ^*



2.2. Self-Training——熵正则化

- 限制y^u的分布，使其更集中



$E(y^u) = 0$

$E(y^u) = 0$

$E(y^u) = \log 5$

$$E(y^u) = -\sum_{C=1}^{5} y_c^u \log y_c^u$$ 期望E越小越好

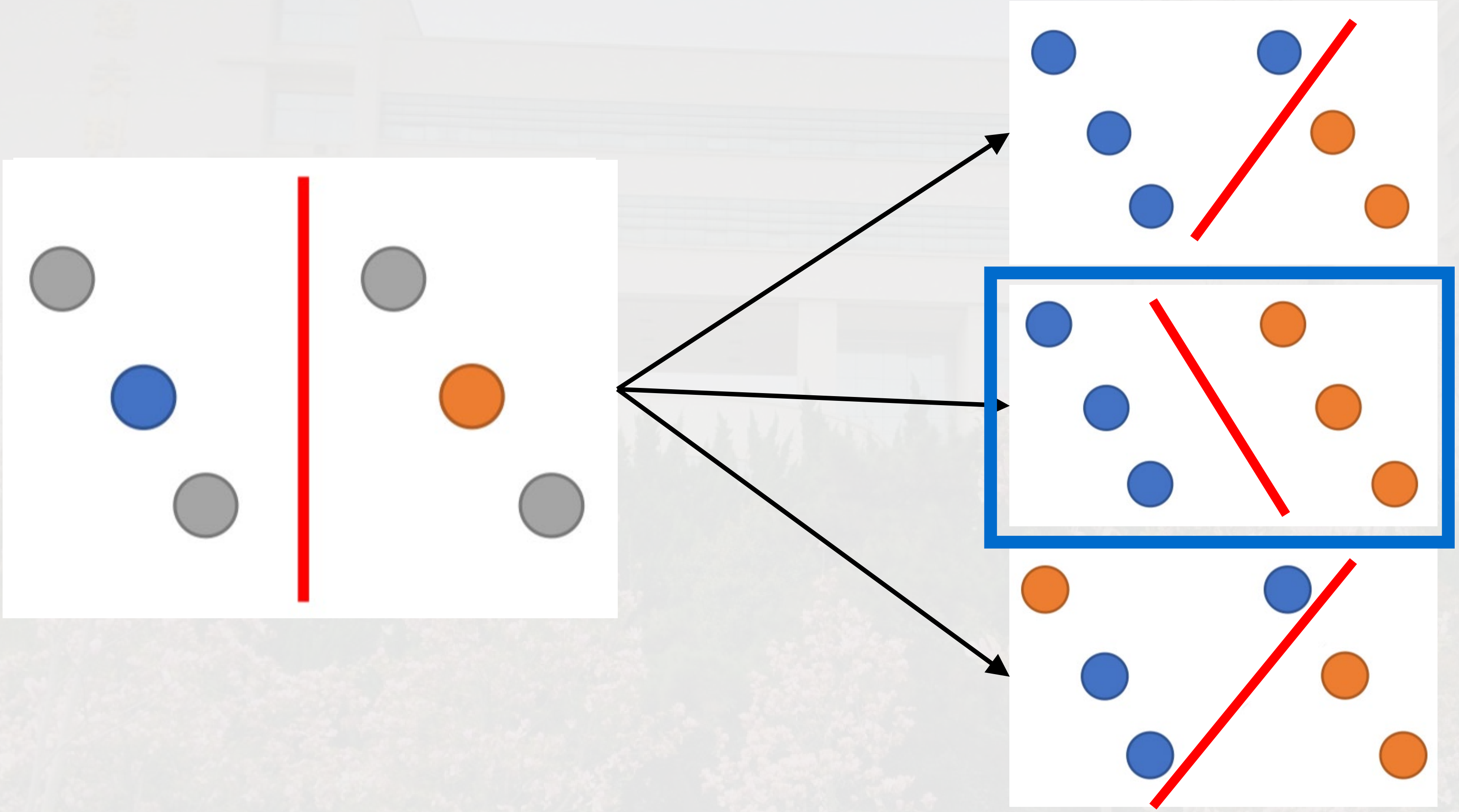
$$Loss = \sum_{x^l} CE(y^l, \hat{y}^l)$$ 有标签数据

$$+\lambda \sum_{x^u} E(y^u)$$ 无标签数据

or 选择熵较小的无标签数据加入有标签集进行Self-Training

2.3.半监督SVM

- **TSVM(Transductive SVM)**：尝试为未标记样本找到合适的标记指派，使得超平面划分后的间隔最大化（低密度分割假设）



2.3. 半监督SVM

$D=\{(\boldsymbol{x}_1,y_1),\ldots,(\boldsymbol{x}_l,y_l),\boldsymbol{x}_{l+1},\ldots,\boldsymbol{x}_m\}$ $\{D_l, D_u\}$

- **TSVM目标**：为D_u中的样本给出预测标记$\hat{y}=(\hat{y}_{l+1},\hat{y}_{l+2},\cdots,\hat{y}_m)$，$\hat{y}_i\in\{-1, +1\}$,满足：

$$\min_{\boldsymbol{w},b,\hat{\boldsymbol{y}},\boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C_l\sum_{i=1}^{l}\xi_i + C_u\sum_{i=l+1}^{m}\xi_i$$

松弛变量 hinge损失

$$\text{s.t. } y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geqslant 1-\boxed{\xi_i,}\ i=1,2,\ldots,l,$$

$$\hat{y}_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geqslant 1-\boxed{\xi_i,}\ i=l+1,l+2,\ldots,m,$$

$$\xi_i\geqslant 0,\ i=1,2,\ldots,m,$$

对应标记样本

对应未标记样本

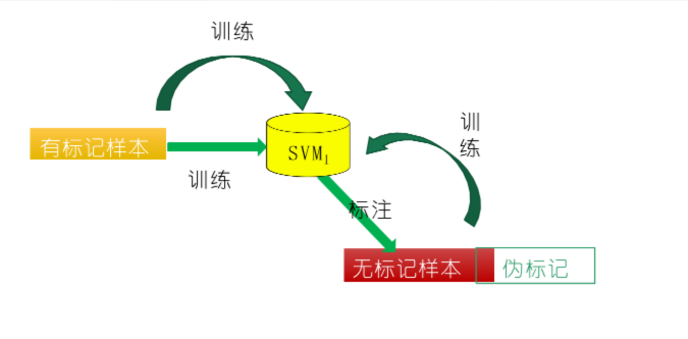
C_l与C_u用于平衡模型复杂度、有标记样本与未标记样本重要程度的折中参数

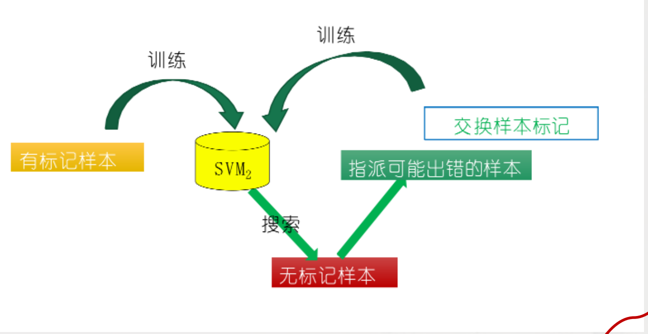
穷举每个无标签样本的预测标记计算复杂度过高，TSVM采用迭代策略

2.3. 半监督SVM

- **TSVM求解**：迭代地进行局部搜索。分两个环节：分配伪标记并获得初始SVM、交换错分的未标记样本更新SVM

- **1. 计算初始SVM**
  - **有标记样本训练SVM**
  - **获得未标记样本的伪标签后，再训练SVM**





- **2. 交换标记重复训练SVM**
  - **挑选两个可能错分的异类未标记样本**
  - **交换其标签，重新训练SVM**

输入： 有标记样本集 $D_l = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_l, y_l)\}$;
未标记样本集 $D_u = \{\boldsymbol{x}_{l+1}, \boldsymbol{x}_{l+2}, \ldots, \boldsymbol{x}_{l+u}\}$;
折中参数 C_l, C_u.

过程：

```
用 D_l 训练一个 SVM_l;  初始SVM
                                                   伪标记
用 SVM_l 对 D_u 中样本进行预测, 得到 ŷ = (ŷ_{l+1}, ŷ_{l+2}, ..., ŷ_{l+u});
初始化 C_u ≪ C_l;
while C_u < C_l do
    基于 D_l, D_u, ŷ, C_l, C_u 求解式(13.9), 得到 (w, b), ξ;
    while ∃{i, j | (ŷ_i ŷ_j < 0) ∧ (ξ_i > 0) ∧ (ξ_j > 0) ∧ (ξ_i + ξ_j > 2)} do
        ŷ_i = −ŷ_i;        松弛变量越大表示离超平面越近，越容易分错
        ŷ_j = −ŷ_j;
        基于 D_l, D_u, ŷ, C_l, C_u 重新求解式(13.9), 得到 (w, b), ξ
    end while
    C_u = min{2C_u, C_l}  逐渐增大Cu
end while
```

输出： 未标记样本的预测结果：$\hat{\boldsymbol{y}} = (\hat{y}_{l+1}, \hat{y}_{l+2}, \ldots, \hat{y}_{l+u})$ 最终调整后的结果

挑选准则

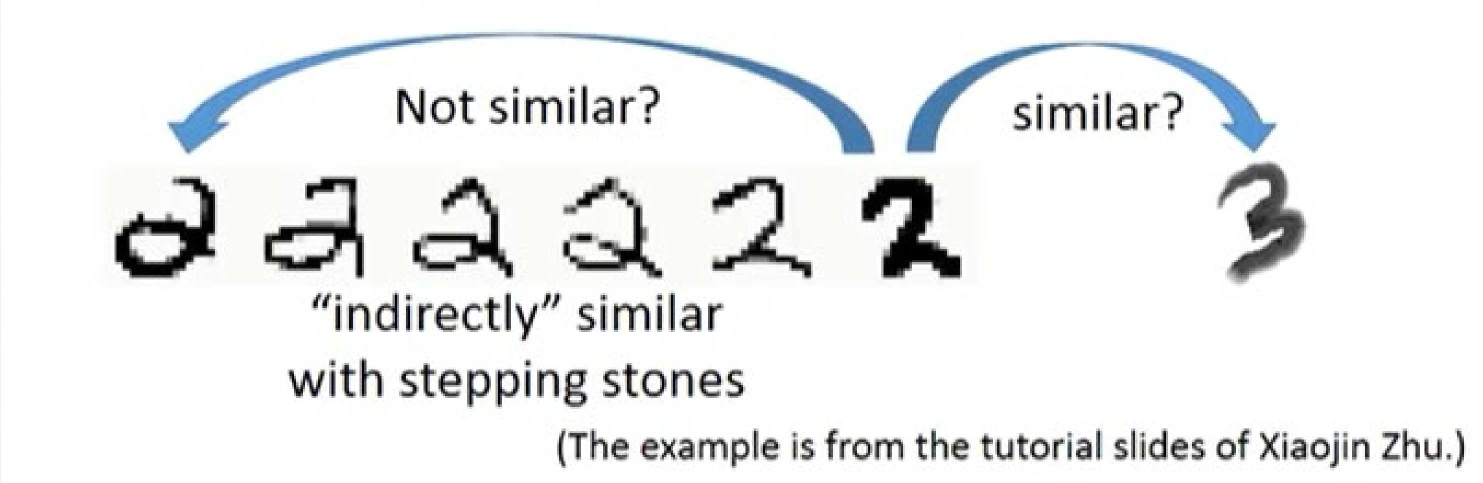
- **减轻类别不平衡的影响，基于伪标记而当作正反例的未标记样本数，将C_u拆为C_u^+、 C_u^-**

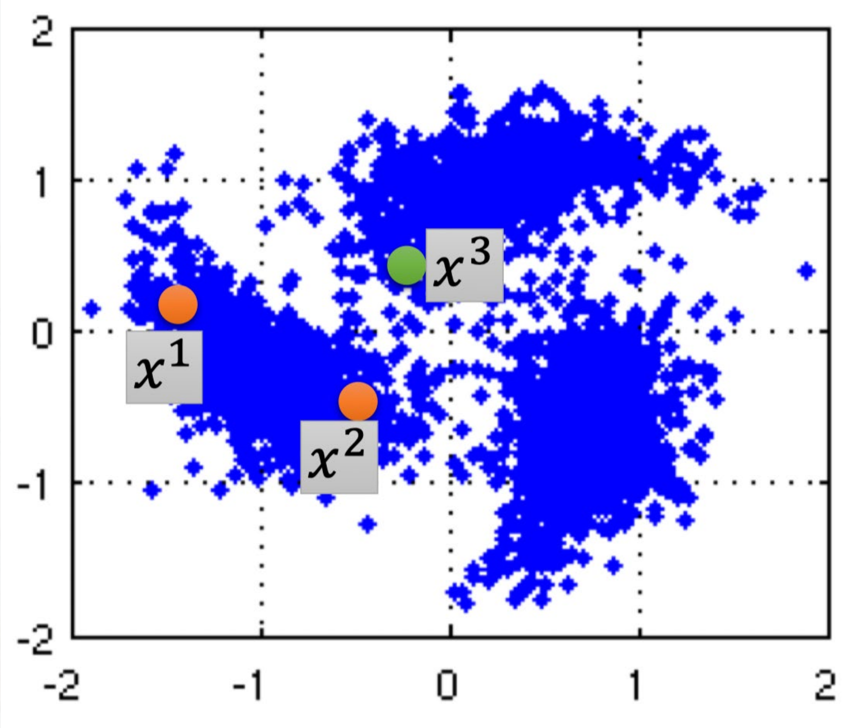
$$C_u^+ = \frac{u_-}{u_+} C_u^-$$

2.4.图半监督学习——平滑假设

- **平滑假设**：相似的样本具有相同的标签

- 样本x的分布非均匀
- 若x^1和x^2在高密度区域相邻，则具有相同的标签





x^1 x^2 具有相同标签
x^2 x^3 具有不同标签

x^1 x^2 are connected by a high density path

2.4.图半监督学习

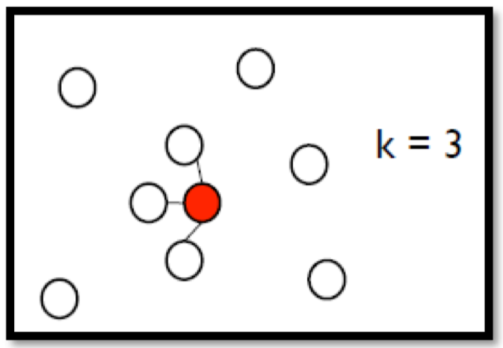
□ 如何判断样本在高密度区域相邻?

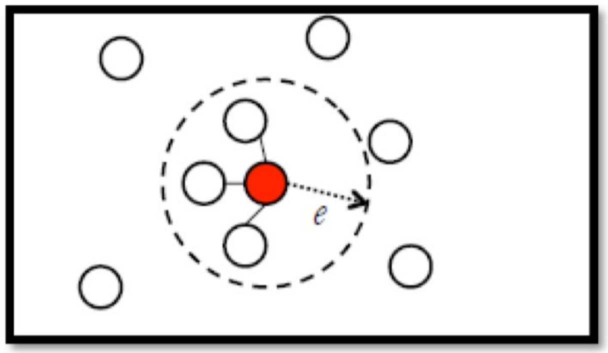
——使用图结构描述样本点

□ 如何构建图?

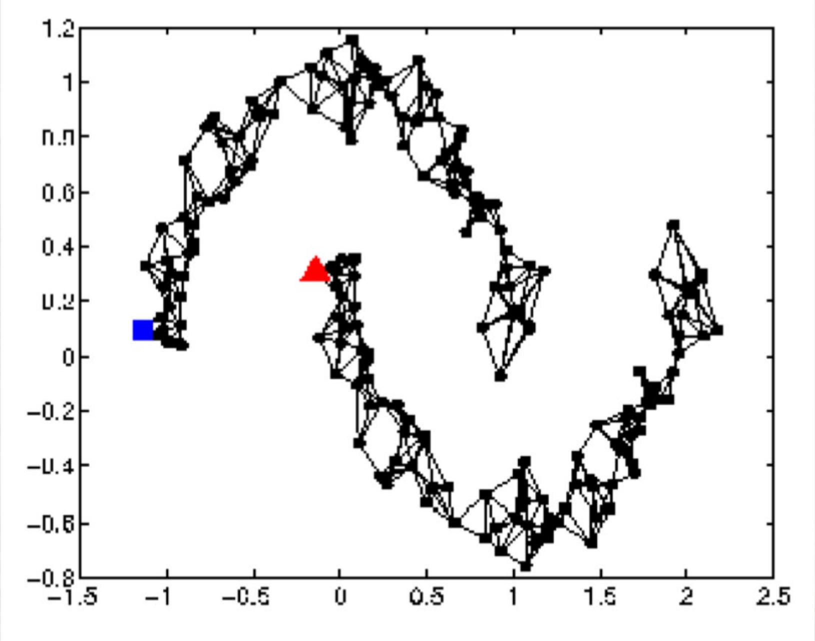
1. 具有自然的图结构：网页链接、论文引用……
2. 从数据构建图：

- 计算样本之间的相似度$s(x^i, x^j) = \exp(-\gamma\|x^i - x^j\|^2)$
- 将K近邻相连
- 将一定范围内的样本相连



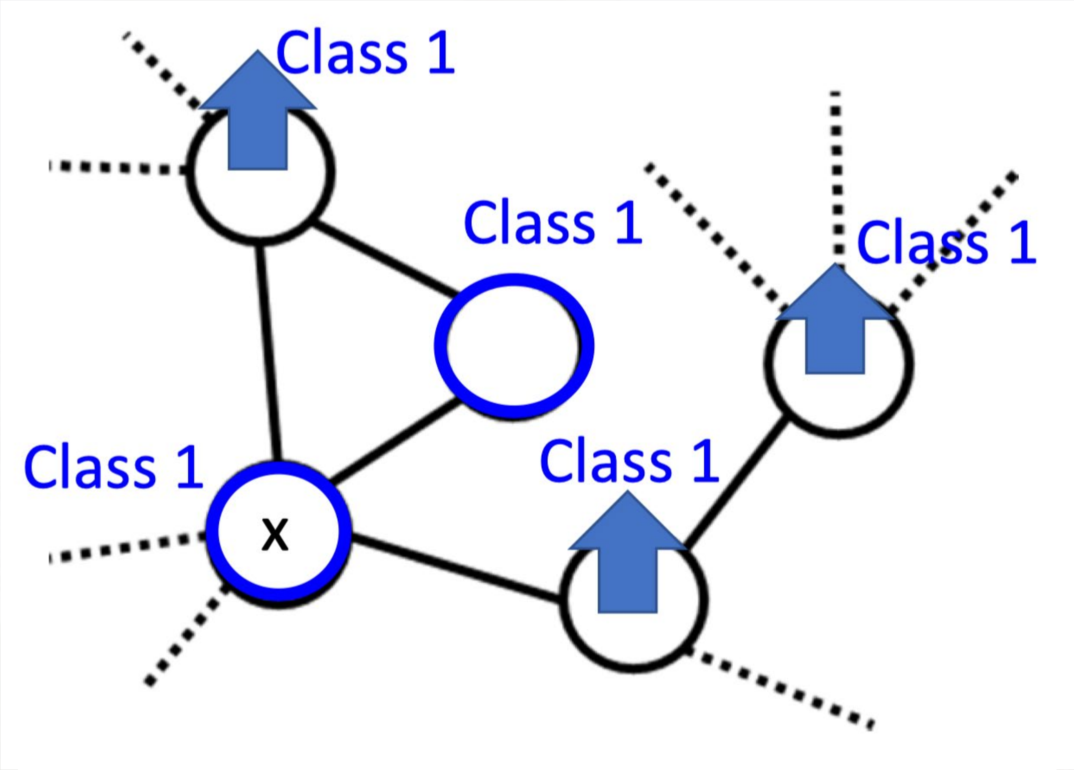


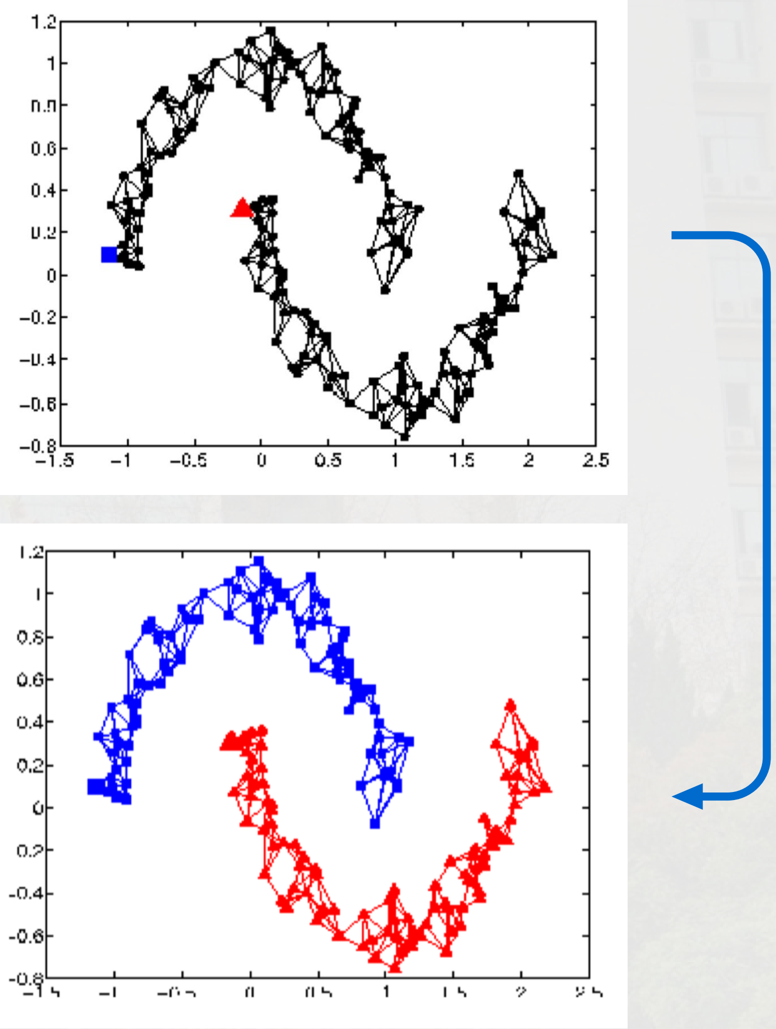
- 边的权重正比于$s(x^i, x^j)$



2.4.图半监督学习

- **节点标签扩散**：将已标记样本根据样本特征的相似性进行扩散，获得未标记样本的标记。





◻ 能量函数：定性评价图的平滑性

$$E=\frac{1}{2}\sum_{i,j} w_{i,j}\left(y^i-y^j\right)^2=\boldsymbol{y}^T L\boldsymbol{y}$$

对于所有样本进行计算（包括有标签数据和无标签数据），越小越好

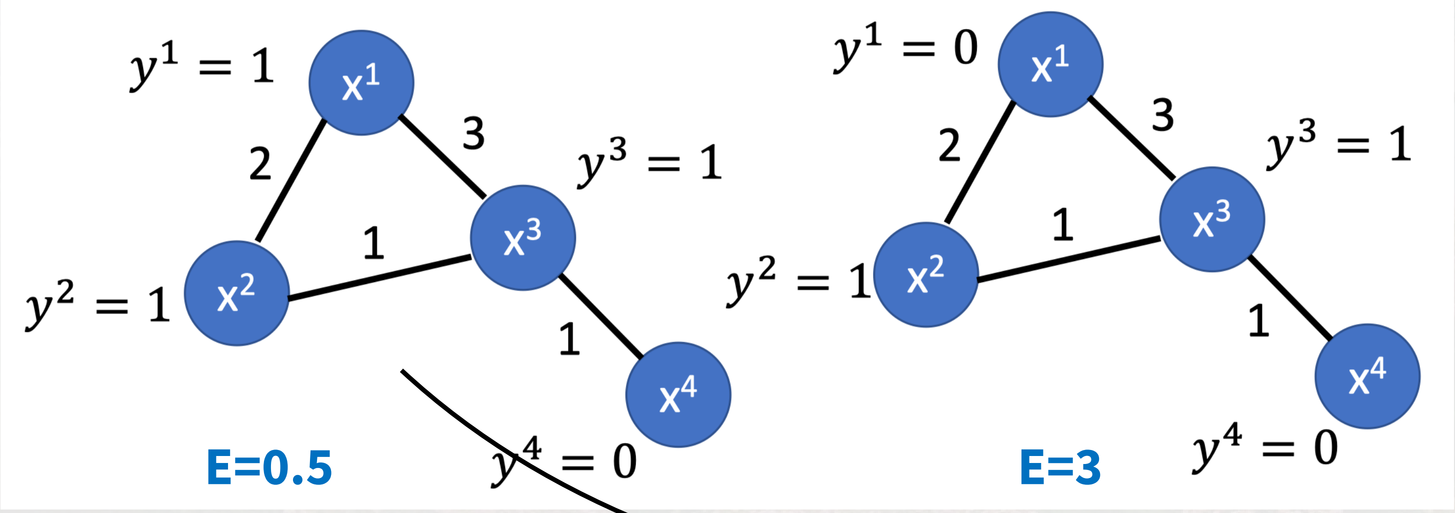
$\boldsymbol{y}$：所有样本的标预测标签 $(R+U)$维

$\boldsymbol{y}=\left[\dots y^i\dots y^j\right]^T \qquad \boldsymbol{y}=(\boldsymbol{y}_l^T,\boldsymbol{y}_u^T)$

L：图拉普拉斯矩阵 $(R+U)\times(R+U)$维

$L=D-W$

$$(\mathbf{W})_{ij}=\begin{cases}\exp\left(\frac{-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|_2^2}{2\sigma^2}\right), & \text{if } i\neq j\,;\\ 0\,, & \text{otherwise}\,,\end{cases}$$



左图更加平滑！

$$W=\begin{bmatrix}0&2&3&0\\2&0&1&0\\3&1&0&1\\0&0&1&0\end{bmatrix}\qquad D=\begin{bmatrix}5&0&0&0\\0&3&0&0\\0&0&5&0\\0&0&0&1\end{bmatrix}$$

2.4.图半监督学习

- **求解**：1. 作为正则项，使用梯度下降求解

$$E=\frac{1}{2}\sum_{i,j} w_{i,j}\left(y^i-y^j\right)^2=\boldsymbol{y}^T L\boldsymbol{y} \qquad Loss=\sum_{x^l} CE(y^l,\hat{y}^l)+\lambda E$$

- **求解**：2. 解析解

设$\boldsymbol{y}$由判别函数$\boldsymbol{f}$求得，即$y_i=sign\left(f(x_i)\right)$，$\boldsymbol{f}=(\boldsymbol{f}_l^T,\ \boldsymbol{f}_u^T)$，可令$E=\boldsymbol{f}^T L\boldsymbol{f}$

写成分块矩阵形式：$E(f)=(\boldsymbol{f}_l^{\mathrm{T}}\ \boldsymbol{f}_u^{\mathrm{T}})\left(\begin{bmatrix}\mathbf{D}_{ll} & \mathbf{0}_{lu}\\ \mathbf{0}_{ul} & \mathbf{D}_{uu}\end{bmatrix}-\begin{bmatrix}\mathbf{W}_{ll} & \mathbf{W}_{lu}\\ \mathbf{W}_{ul} & \mathbf{W}_{uu}\end{bmatrix}\right)\begin{bmatrix}\boldsymbol{f}_l\\ \boldsymbol{f}_u\end{bmatrix}$ $(L=D-W)$

最小化能量函数：$\frac{\partial E(f)}{\partial \boldsymbol{f}_u}=\mathbf{0}$ 解得：$\boldsymbol{f}_u=(\mathbf{D}_{uu}-\mathbf{W}_{uu})^{-1}\mathbf{W}_{ul}\boldsymbol{f}_l$.

令 $\mathbf{P}_{uu}=\mathbf{D}_{uu}^{-1}\mathbf{W}_{uu},\ \mathbf{P}_{ul}=\mathbf{D}_{uu}^{-1}\mathbf{W}_{ul}$ 则 $\boldsymbol{f}_u=(\mathbf{I}-\mathbf{P}_{uu})^{-1}\mathbf{P}_{ul}\boldsymbol{f}_l$

代入D、W、$\boldsymbol{f}_l$即可求得$\boldsymbol{f}_u$

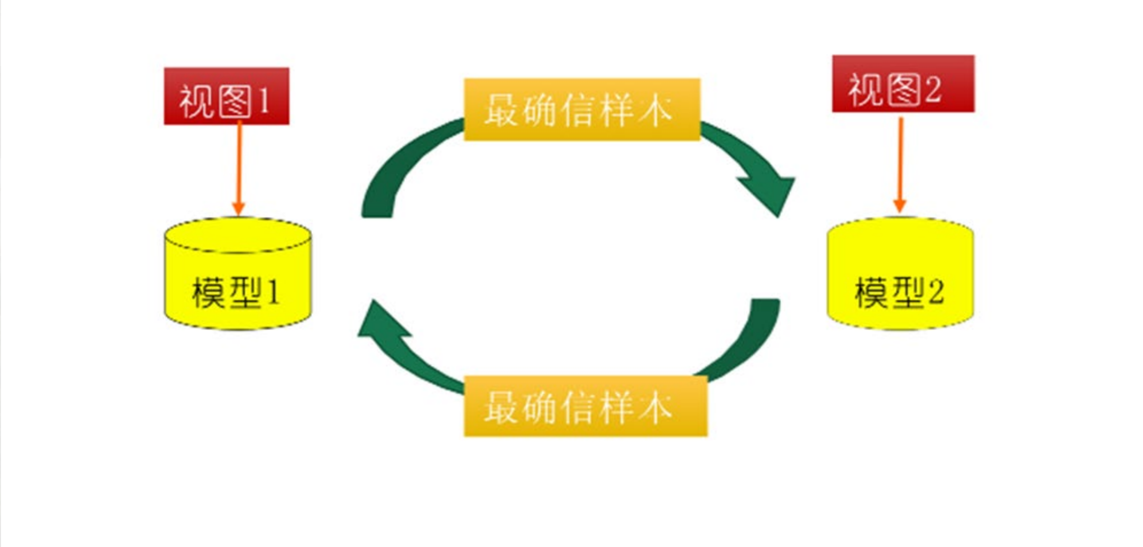
2.5.基于分歧的方法

- 多视图数据：一个数据对象有多个属性集，每个属性集构成了视图。
  - 样本（$\langle x^1, x^2\rangle, y$），其中$x^i$为样本在视图i中的示例，y为标记
  - 例如电影中的声音和视频分别对应一个视图，类型则为“动作片”、“爱情片”等
- 多视图具有相容性，进而具有互补性
  - 不同视图输出空间是一致的，以电影为例，类别均应为{爱情片、动作片}
  - 故可利用多视图的互补性加强分类的准确性

2.5.基于分歧的方法

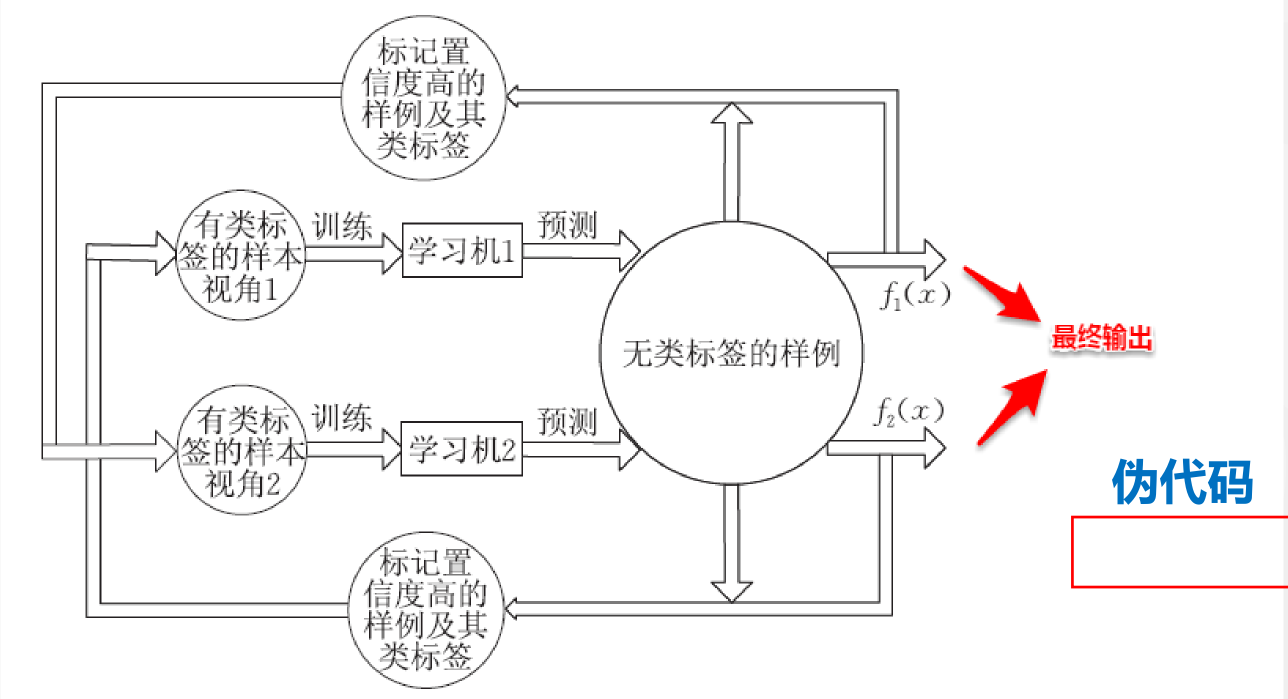
- 协同训练：基于两个充分且条件独立的视图，利用未标记数据相互促进。
  - 充分：每个视图均包含产生最优学习器的信息
  - 条件独立：给定类别标记条件下，两个视图独立

- 基于标记样本训练视图模型1
- 以模型1挑选出该视图最确信的未标记样本赋予其伪标记，并将以上伪标记样本作为新的标记样本加至视图模型2的训练集
- 对视图2模型进行训练，进而挑出该视图模型最确信的未标记样本赋予其伪标记，将其加至视图模型1的训练集
- 重复以上两步，直到两个分类器不变



2.5.基于分歧的方法

□ 协同训练算法



伪代码

输入: 有标记样本集 $D_l = \{(\langle \boldsymbol{x}_1^1, \boldsymbol{x}_1^2\rangle, y_1), \ldots, (\langle \boldsymbol{x}_l^1, \boldsymbol{x}_l^2\rangle, y_l)\}$;
未标记样本集 $D_u = \{\langle \boldsymbol{x}_{l+1}^1, \boldsymbol{x}_{l+1}^2\rangle, \ldots, \langle \boldsymbol{x}_{l+u}^1, \boldsymbol{x}_{l+u}^2\rangle\}$;
缓冲池大小 s;
每轮挑选的正例数 p;
每轮挑选的反例数 n;
基学习算法 $\mathfrak{L}$;
学习轮数 T.

设置缓冲池，减少了每轮计算置信度的次数

过程:

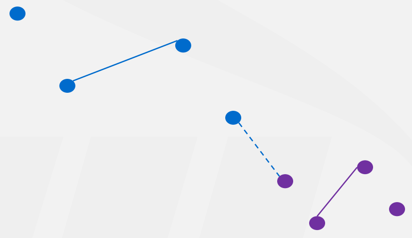
```
从 D_u 中随机抽取 s 个样本构成缓冲池 D_s
D_u = D_u \ D_s;
for j = 1,2 do
    D_l^j = {(x_i^j, y_i) | (⟨x_i^j, x_i^{3-j}⟩, y_i) ∈ D_l};   各视图的有标记样本
end for
for t = 1,2,...,T do
    for j = 1,2 do
        h_j ← 𝔏(D_l^j);   基于每个视图训练初始学习器
        考察 h_j 在 D_s^j = {x_i^j | ⟨x_i^j, x_i^{3-j}⟩ ∈ D_s} 上的分类置信度，挑选 p 个正例
        置信度最高的样本 D_p ⊂ D_s、n 个反例置信度最高的样本 D_n ⊂ D_s;
        由 D_p^j 生成伪标记正例 D̃_p^{3-j} = {(x_i^{3-j}, +1) | x_i^j ∈ D_p^j};
        由 D_n^j 生成伪标记反例 D̃_n^{3-j} = {(x_i^{3-j}, -1) | x_i^j ∈ D_n^j};
        D_s = D_s \ (D_p ∪ D_n);   两个学习器挑选的不会有重复
    end for
    if h_1, h_2 均未发生改变 then
        break
    else
        for j = 1,2 do
            D_l^j = D_l^j ∪ (D̃_p^j ∪ D̃_n^j);   加入打过伪标的未标记样本
        end for
        从 D_u 中随机抽取 2p + 2n 个样本加入 D_s   补充缓冲池
    end if
end for
```

输出: 分类器 h_1, h_2　最终输出两个分类器做集成

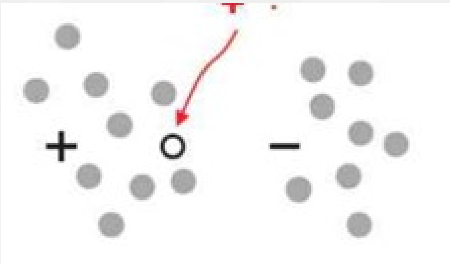
研究表明，协同训练无需多多视图数据，只要保证多个弱学习器之间有显著分歧即可

2.6.半监督聚类

- 拥有部分额外监督信息时：可利用监督信息改善聚类效果。
- 两种监督信息：
  - 约束：必连（实线）与勿连（虚线）



  - 少量有标签样本



2.6.半监督聚类

□ 约束K-均值算法

输入：样本集 $D=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_m\}$;
必连约束集合 $\mathcal{M}$;
勿连约束集合 $\mathcal{C}$;
聚类簇数 k.

过程:

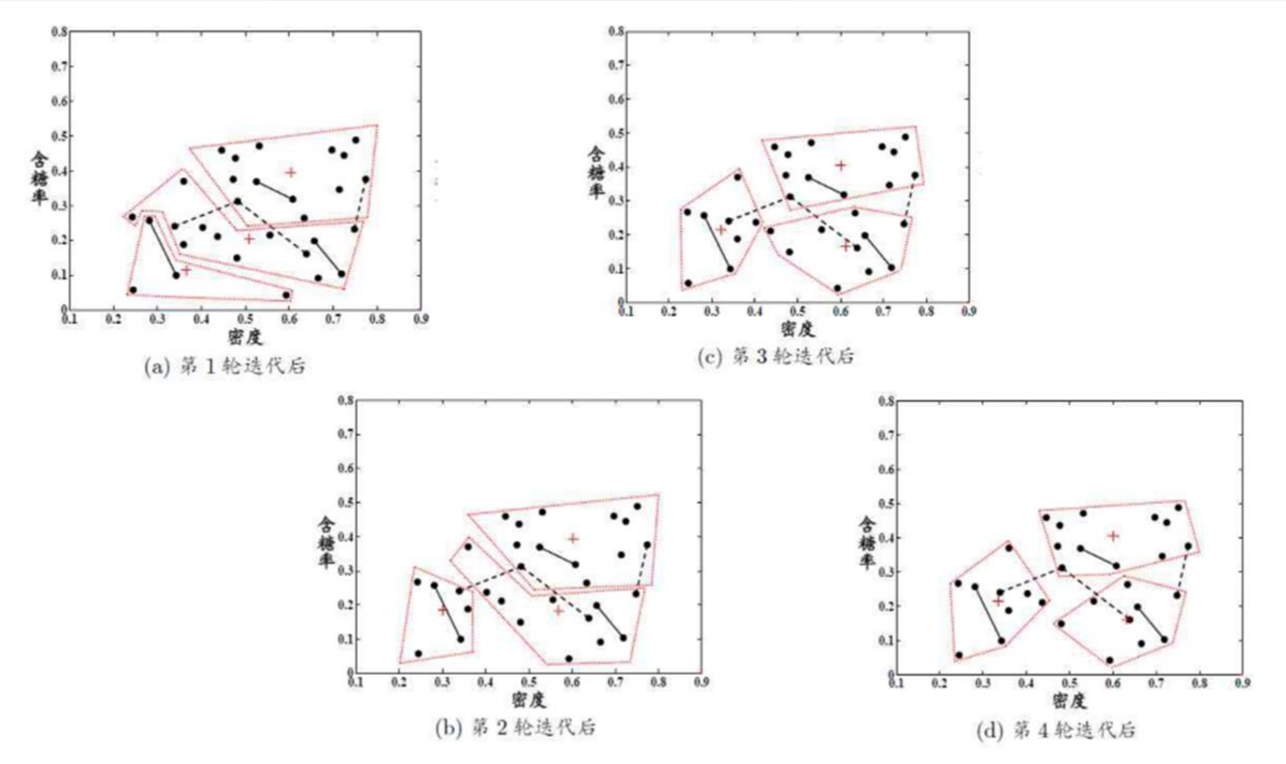
从 D 中随机选取 k 个样本作为初始均值向量$\{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\ldots,\boldsymbol{\mu}_k\}$;
repeat
  $C_j=\varnothing\ (1\leqslant j\leqslant k)$;
  for $i=1,2,\ldots,m$ **do**
    计算样本 $\boldsymbol{x}_i$ 与各均值向量 $\boldsymbol{\mu}_j\ (1\leqslant j\leqslant k)$ 的距离: $d_{ij}=\|\boldsymbol{x}_i-\boldsymbol{\mu}_j\|_2$;
    $\mathcal{K}=\{1,2,\ldots,k\}$;
    is_merged=false;
    while ¬ is_merged **do**
      基于 $\mathcal{K}$ 找出与样本 $\boldsymbol{x}_i$ 距离最近的簇: $r=\arg\min_{j\in\mathcal{K}} d_{ij}$;
      检测将 $\boldsymbol{x}_i$ 划入聚类簇 C_r 是否会违背 $\mathcal{M}$ 与 $\mathcal{C}$ 中的约束;
      if ¬ is_voilated **then**
        $C_r=C_r\bigcup\{\boldsymbol{x}_i\}$;
        is_merged=true
      else
        $\mathcal{K}=\mathcal{K}\setminus\{r\}$; 若不满足则寻找距离次小的类簇
        if $\mathcal{K}=\varnothing$ **then**
          break并返回错误提示
        end if
      end if
    end while
  end for
  for $j=1,2,\ldots,k$ **do**
    $\boldsymbol{\mu}_j=\frac{1}{|C_j|}\sum_{\boldsymbol{x}\in C_j}\boldsymbol{x}$;
  end for
until 均值向量均未更新

输出：簇划分 $\{C_1,C_2,\ldots,C_k\}$

对样本进行划分时，需检测是否满足约束关系，其它步骤均相同

在标准K-Means算法中加入对约束的检测，依次判断距离无标签样本最近的簇是否满足约束

- **西瓜数据集4.0,施加若干必连和勿连约束，聚类迭代过程**



(a) 第 1 轮迭代后
(c) 第 3 轮迭代后
(b) 第 2 轮迭代后
(d) 第 4 轮迭代后

2.6.半监督聚类

▢ 少量有标记样本的聚类算法

输入：样本集 $D = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$;
少量有标记样本 $S = \bigcup_{j=1}^{k} S_j$;
聚类簇数 k .

过程:

for $j = 1, 2, \ldots, k$ do
  $\boldsymbol{\mu}_j = \frac{1}{|S_j|} \sum_{\boldsymbol{x} \in S_j} \boldsymbol{x}$
end for

使用带标记样本各类别的均值向量作为初始类中心

repeat
  $C_j = \varnothing \ (1 \leqslant j \leqslant k)$;
  for $j = 1, 2, \ldots, k$ do
    for all $\boldsymbol{x} \in S_j$ do
      $C_j = C_j \bigcup \{\boldsymbol{x}\}$
    end for
  end for

带标记样本直接划入对应类簇

  for all $\boldsymbol{x}_i \in D \setminus S$ do
    计算样本 $\boldsymbol{x}_i$ 与各均值向量 $\boldsymbol{\mu}_j \ (1 \leqslant j \leqslant k)$ 的距离: $d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{\mu}_j\|_2$;
    找出与样本 $\boldsymbol{x}_i$ 距离最近的簇: $r = \arg\min_{j \in \{1,2,\ldots,k\}} d_{ij}$;
    将样本 $\boldsymbol{x}_i$ 划入相应的簇: $C_r = C_r \bigcup \{\boldsymbol{x}_i\}$

划分无标记样本

  end for
  for $j = 1, 2, \ldots, k$ do
    $\boldsymbol{\mu}_j = \frac{1}{|C_j|} \sum_{\boldsymbol{x} \in C_j} \boldsymbol{x}$;

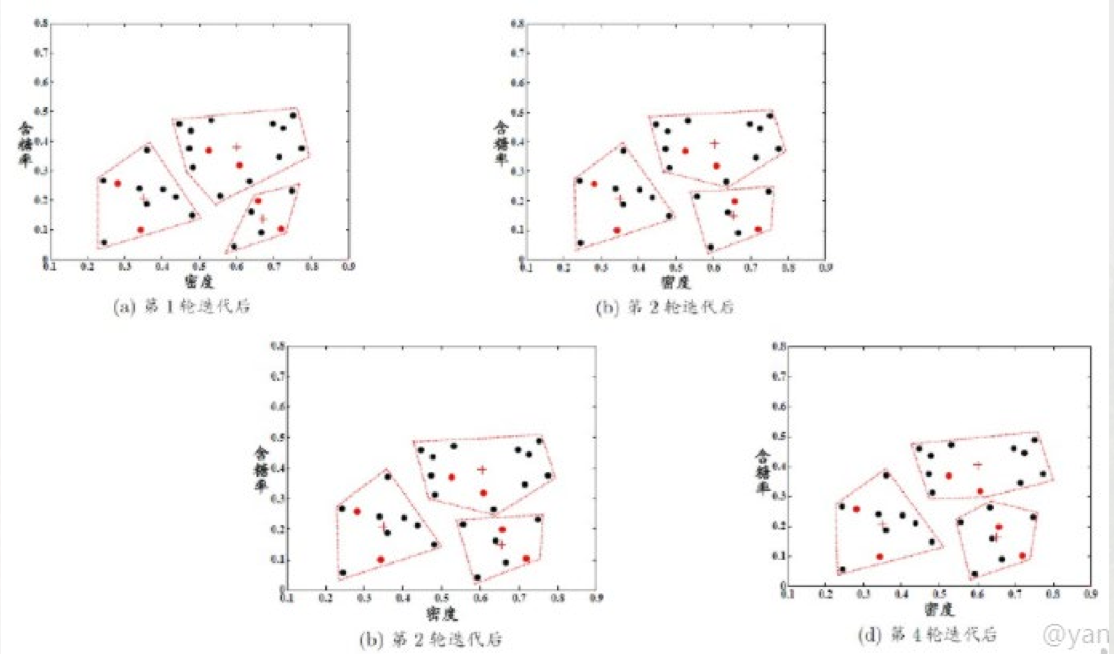
重新计算类中心

  end for
until 均值向量均未更新

输出：簇划分 $\{C_1, C_2, \ldots, C_k\}$

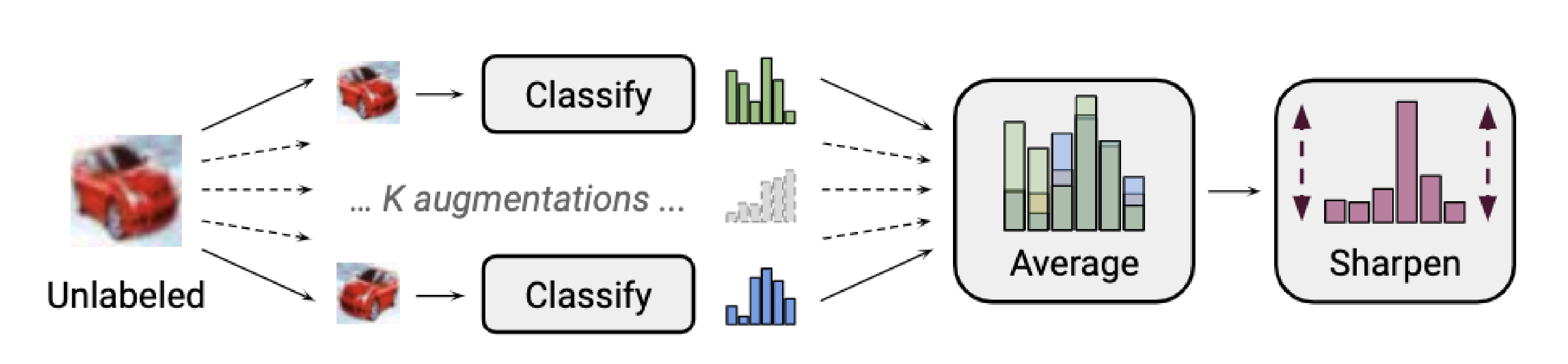
先根据有标记样本初始化聚类中心，再使用K-Means算法划分无标记样本

- **西瓜数据集4.0,假定部分样本类别已知**



XX-Match

- **一致性正则：**给输入加入扰动，模型的输出应该尽可能保持不变或者近似
  - 注入 noise 可以通过模型本身的随机性（如 dropout）或者直接加入噪声，也可以通过 data augmentation；计算一致性的方法，可以使用 L2，也可以使用 KL divergency、cross entropy。
- **熵最小化：**要求预测近似于one-hot标签



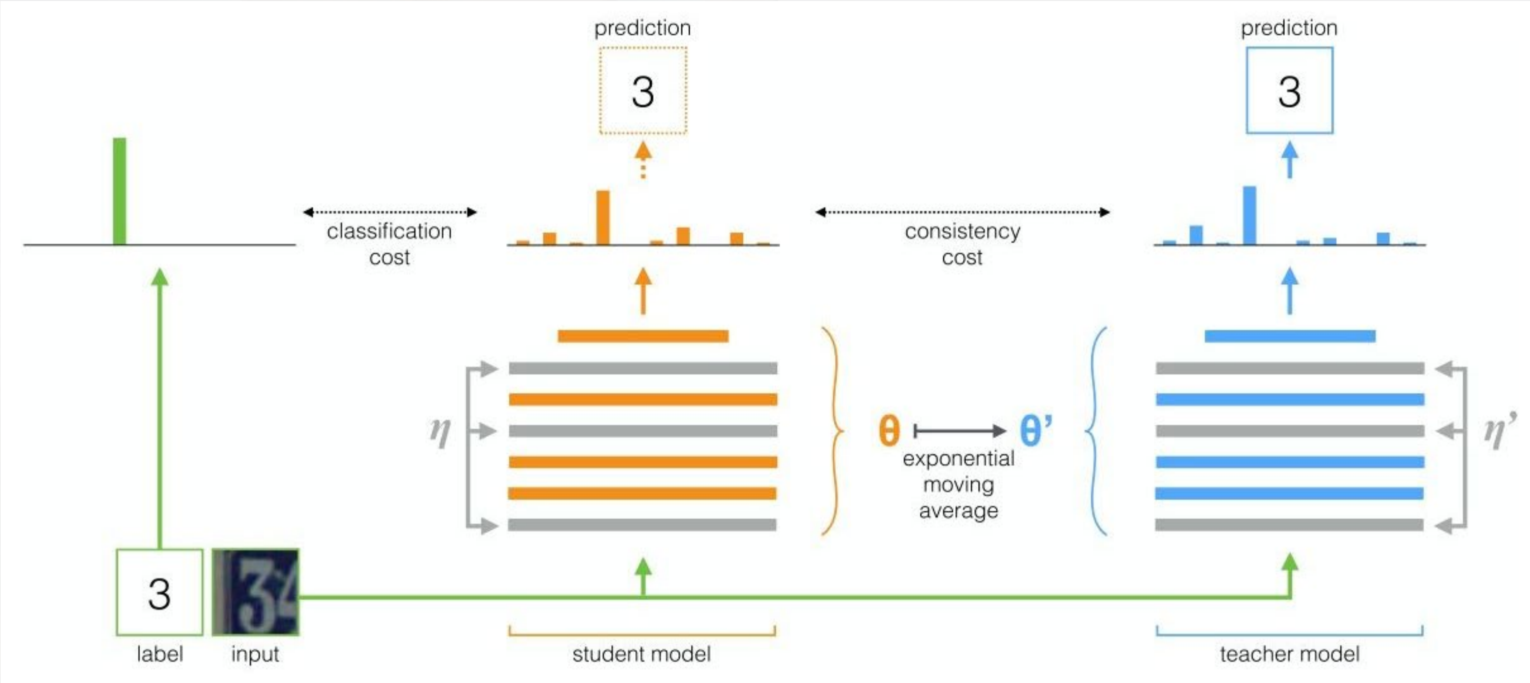
MixMatch:
多次数据增强后取平均预测作为伪标签

- MixMatch: A Holistic Approach to Semi-Supervised Learning
- ReMixMatch: Semi-Supervised Learning with Distribution Alignment and Augmentation Anchoring
- UDA: Unsupervised Data Augmentation for Consistency Training
- FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence

3.领域前沿

Teacher Student Model

➢ 包含两个model，一个student，一个teacher，teacher引导student从数据中学习“知识”



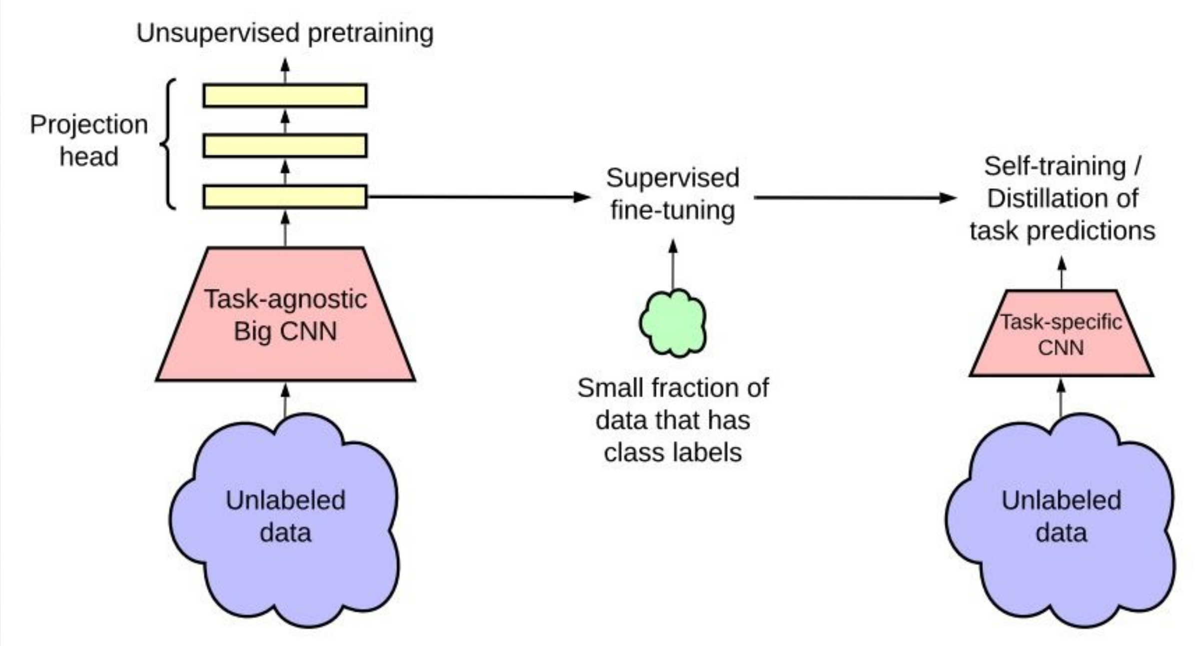
Mean Teachers:
对Student的预测计算分类损失
Student和Teacher之间计算一致性损失
Teacher使用指数滑动平均进行更新

- Temporal Ensembling for Semi-Supervised Learning
- Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results

3.领域前沿

☐ Self-supervised based SSL

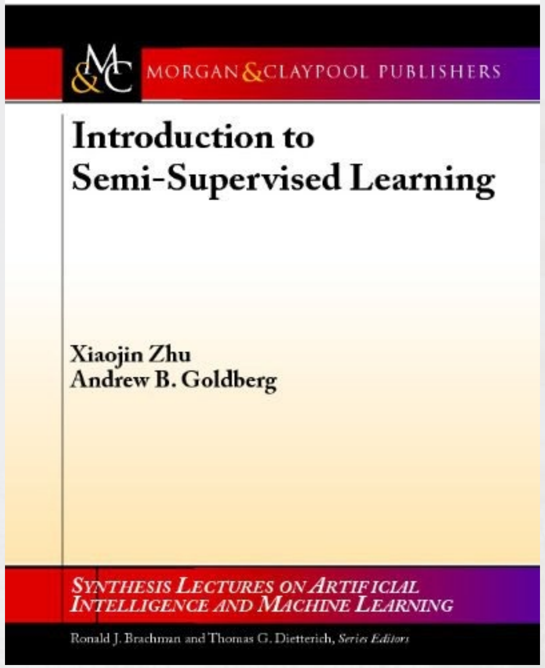
➢ 使用自监督方法进行预训练，再使用有标签数据进行微调



Big Self-Supervised Models:
先试用自监督学习进行预训练
再使用有监督学习进行微调
最后在半监督数据集上进行Self-Training

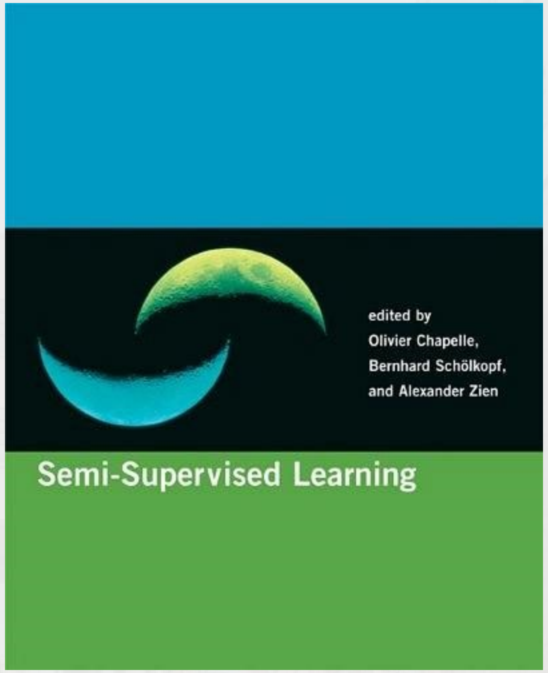
- S4L: Self-Supervised Semi-Supervised Learning
- Big Self-Supervised Models are Strong Semi-Supervised Learners

4.相关文献



Introduction to Semi-Supervised Learning

Xiaojin Zhu and Andrew B. Goldberg



Semi-Supervised Learning

Chapelle Olivier;
Scholkopf Bernhard;
Zien Alexander

视频教学：
https://www.youtube.com/watch?v=fX_guE7JNnY&list=PLJV_el3uVTsPy9oCRY30oBPNLCo89yu49&index=14

相关资料汇总：https://github.com/yassouali/awesome-semi-supervised-learning

• 本章小结

- □ 建立无标记样本利用的概念
- □ 高斯混合模型实现赋予未标记样本伪标签
- □ Self-Training：自我迭代确定伪标签
- □ TSVM法利用局部搜索异类错分未标记样本迭代求近似解
- □ 图半监督学习：二分类的扩散方法和多类情况的迭代传播算法
- □ 基于分歧的方法：多视图协同训练
- □ 半监督聚类：先验约束和部分标记样本