课程10：
半监督学习

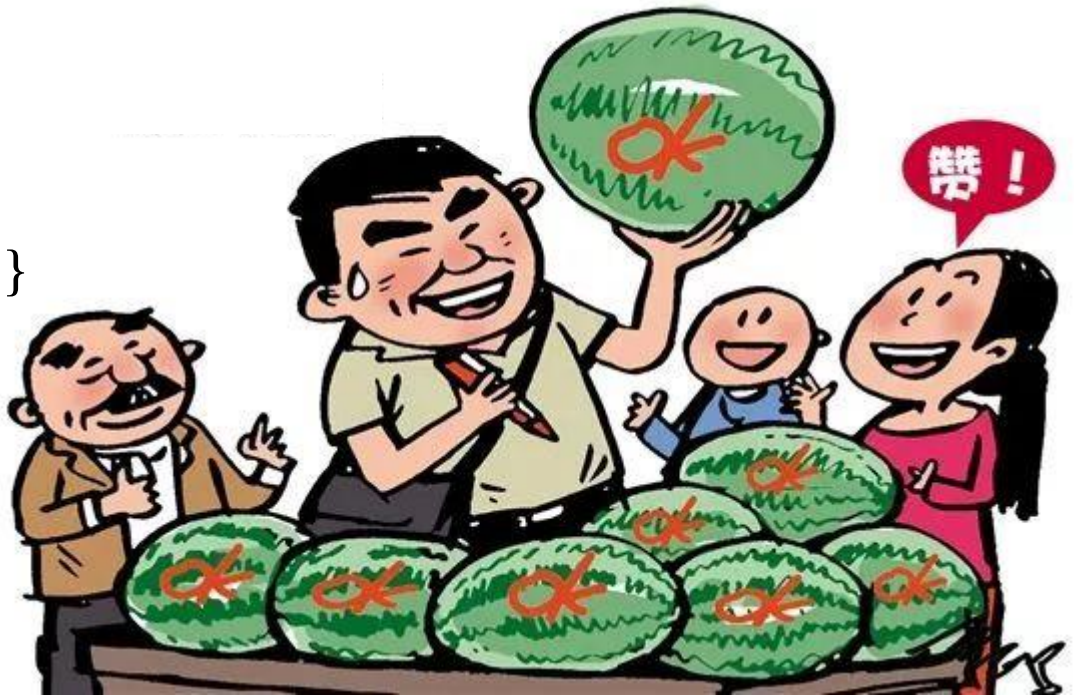
在丰收季节来到瓜田，满地都是西瓜，瓜农抱来五个西瓜说都是好瓜，又指着地里的五个西瓜说这些还不好，还需要再长几天，基于这些信息，我们能否构建一个模型，用来判断地里的哪些瓜是可以采摘的好瓜？ 训练样本过少

类别标记（即是否好瓜）

训练样本集： $D_l = \{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\}$

无标记样本集： $D_u = \{(x_{l+1}, y_{l+1}), (x_{l+2}, y_{l+2}), \ldots, (x_{l+u}, y_{l+u})\}$

$l \ll u$



类别标记（即是否好瓜）

训练样本集： $D_l=\{(x_1,y_1),(x_2,y_2),...,(x_l,y_l)\}$

无标记样本集： $D_u=\{(x_{l+1},y_{l+1}),(x_{l+2},y_{l+2}),...,(x_{l+u},y_{l+u})\}$

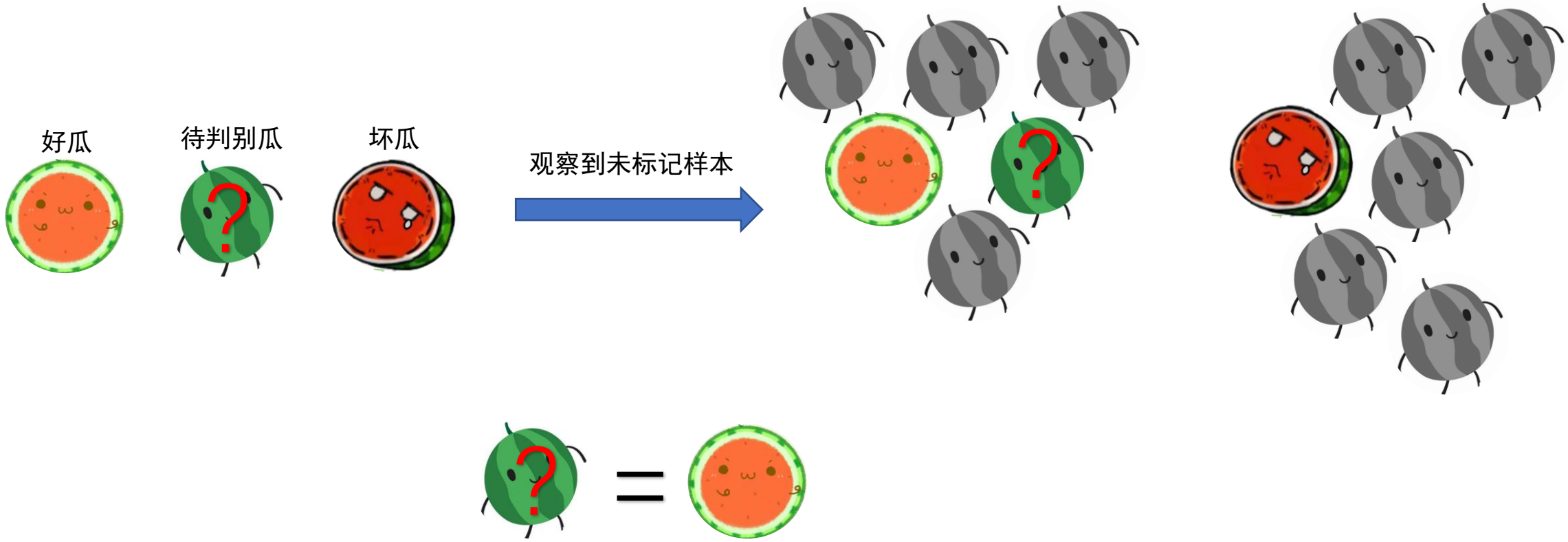
$l \ll u$

- 请农民将地里的瓜全部都检查一遍，告诉我们那些是好瓜，哪些是坏瓜。 时间成本过高
- 先用训练集训练一个样本，拿这个模型去地里挑一个瓜，询问瓜农好坏，然后将这个样本加入训练集中重新训练，…这样只用询问几次瓜农就能获得性能不错的模型。

有没有更好办法?

直接使用未标记样本

- 考虑到未标记样本虽未直接包含标记信息，但若它们与有标记样本是从同样的数据源独立同分布采样而来的，那它们包含的信息对建立模型就有巨大作用。

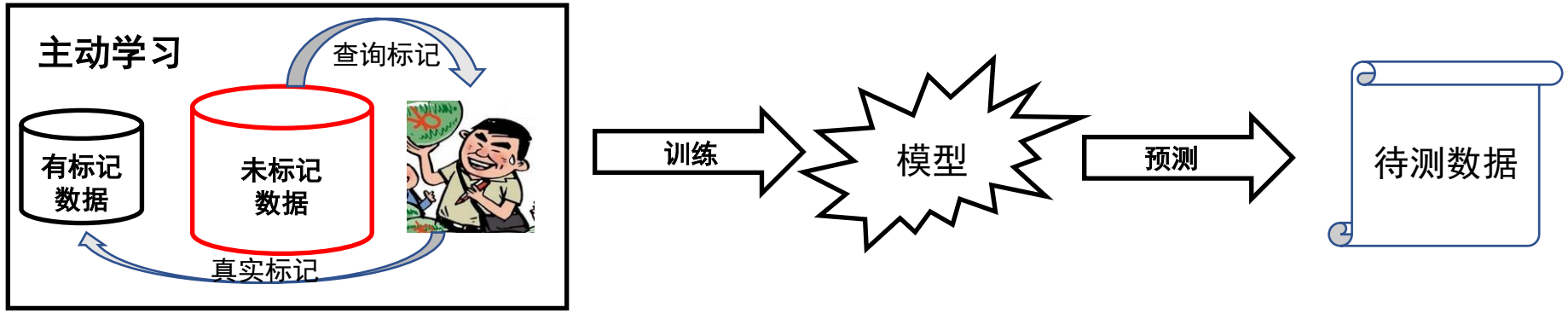


半监督学习

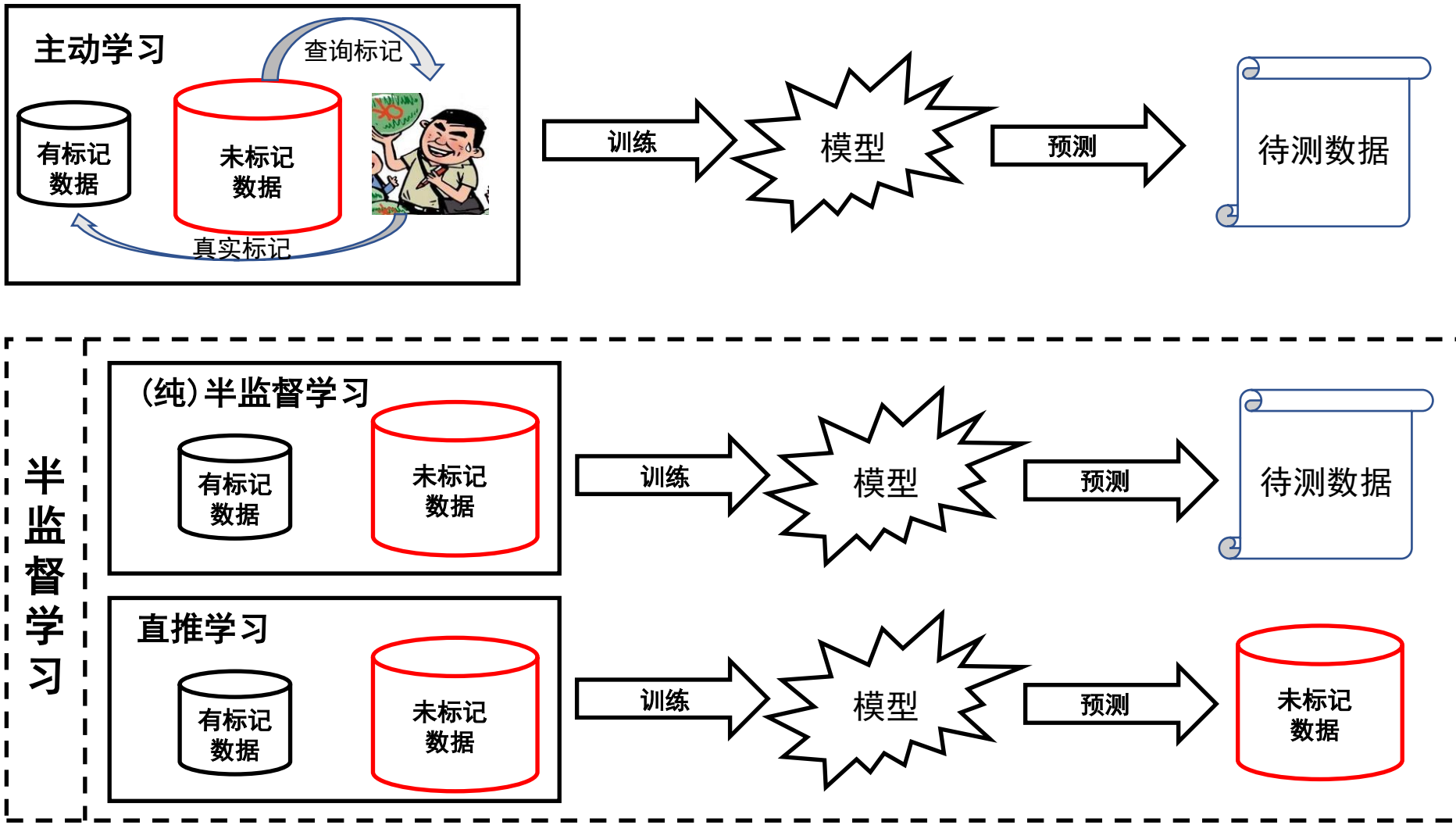
- 让学习器不依赖外界交互、自动地利用未标记样本来提升学习性能，就是半监督学习(semi-supervised learning)
- “聚类假设”：假设数据存在簇结构，同一个簇的样本属于一个类别。
- “流形假设”：假设数据分布在一个流形结构上，邻近样本拥有相似的输出值。

“相似的样本拥有相似的输出”

半监督学习进一步可以分为纯半监督学习和直推学习



半监督学习


主动学习
查询标记
有标记数据
未标记数据
真实标记
训练
模型
预测
待测数据
半监督学习
(纯)半监督学习
有标记数据
未标记数据
训练
模型
预测
待测数据
直推学习
有标记数据
未标记数据
训练
模型
预测
未标记数据

半监督学习

- 自训练算法（self-training）
- 生成式方法（generative method）
- 半监督支持向量机（SVM）
- 图半监督方法（graph-based method）
- 多视图学习（multi-view learning）

自训练算法(self-training)

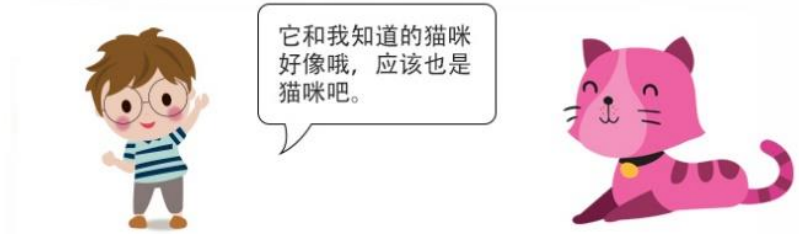
两个样本集合：$\text{Labeled} = \{(x_1, y_1), (x_2, y_2), ..., (x_l, y_l)\}$　$\text{Unlabeled} = \{x_{l+1}, x_{l+2}, ..., x_{l+u}\}, l \ll u, l + u = m$

1. 用Labeled集合生成分类策略F;
2. 用F分类Unlabeled，计算误差
3. 选取Unlabeled的子集u(误差小的)，加入标记。Labeled = Labeled + u
4. 重复上述步骤，直到Unlabeled为空集

(a)少量签数据集（两个标签数据）

具体实例：最近邻算法

1. 用Labeled集合生成分类策略F;
2. 选择 $x = \text{argmin}(d(x, x_o))$，其中$x \in \text{Unlabeled}, x_o \in \text{Labeled}$ 欧氏距离
3. 用F给x定一个类别F(x)，把(x, F(x))加入到Labeled中
4. 重复上述步骤，直到Unlabeled为空集



生成式方法（generative method）

给定标记样本集 $D_l=\{(x_1,y_1),(x_2,y_2),...,(x_l,y_l)\}$ 和未标记样本集 $D_u=\{x_{l+1},x_{l+2},...,x_{l+u}\},l\ll u,l+u=m$

假设所有样本独立同分布，且都是由一个高斯混合模型生成的，用极大似然法来估计高斯混合模型的参数 $\{(\alpha_i,\mu_i,\Sigma_i)|1\le i\le N\}$

$$p(x)=\sum_{i=1}^{N}\alpha_i\cdot p(x|\mu_i,\Sigma_i)$$

使用EM算法对高斯混合模型的参数进行更新：

- E步：根据当前模型参数计算未标记样本属于各高斯混合成分的概率

$$\gamma_{ji}=\frac{\alpha_i\cdot p(x_j|\mu_i,\Sigma_i)}{\sum_{i=1}^{N}\alpha_i\cdot p(x_j|\mu_i,\Sigma_i)};$$

- M步：基于上述概率更新模型参数，其中 l_i 表示第i类有标记样本数目

$$\mu_i=\frac{1}{\sum_{x_j\in D_u}\gamma_{ji}+l_i}\left(\sum_{x_j\in D_u}\gamma_{ji}x_j+\sum_{(x_j,u_j)\in D_{l\wedge y_i=i}}x_j\right)$$

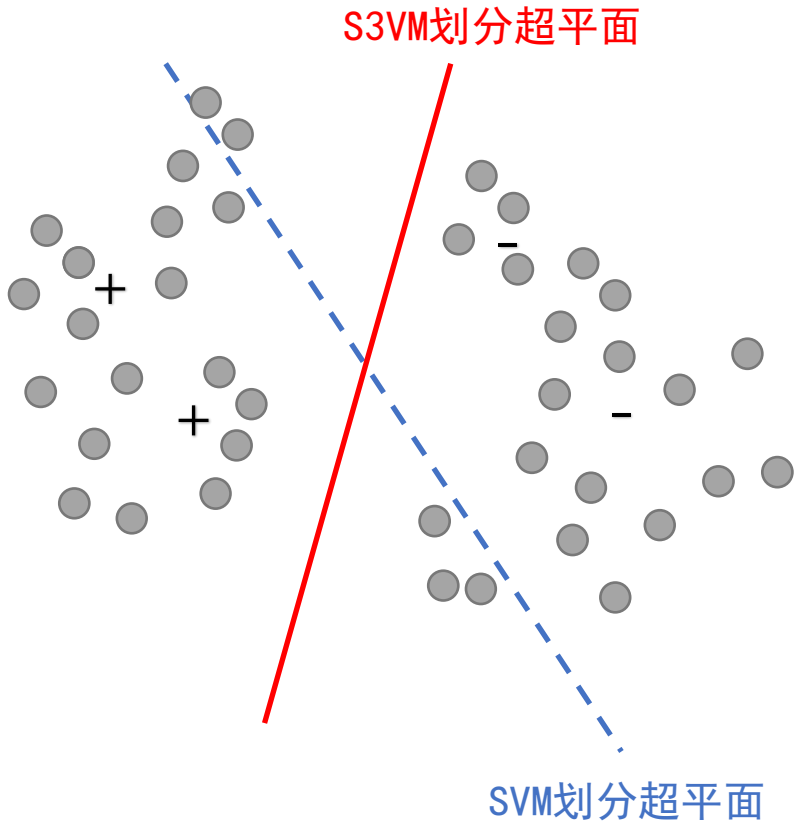
$$\alpha_i=\frac{1}{m}\left(\sum_{x_j\in D_u}\gamma_{ji}+l_i\right)$$

$$\Sigma_i=\frac{1}{\sum_{x_j\in D_u}\gamma_{ji}+l_i}\left(\sum_{x_j\in D_u}\gamma_{ji}(x_j-\mu_i)(x_j-\mu_i)^T+\sum_{(x_j,u_j)\in D_{l\wedge y_i=i}}(x_j-\mu_i)(x_j-\mu_i)^T\right)$$

EM算法具体介绍

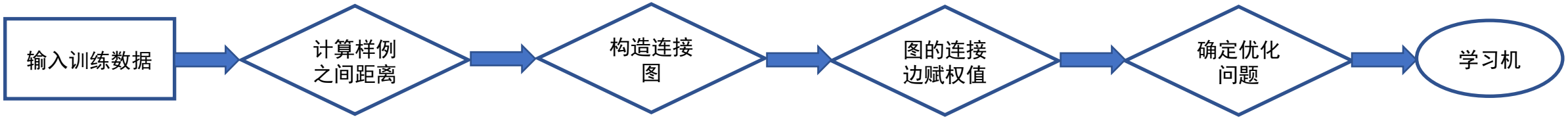
半监督支持向量机

在不考虑未标记样本时，支持向量机试图找到最大间隔划分超平面，而在考虑未标记样本后，S3VM(Semi-Supervised Support Vector Machine)试图找到能将两类有标记样本分开，且穿过数据低密度区域的划分超平面，如下所示：



图半监督方法(graph-based method)

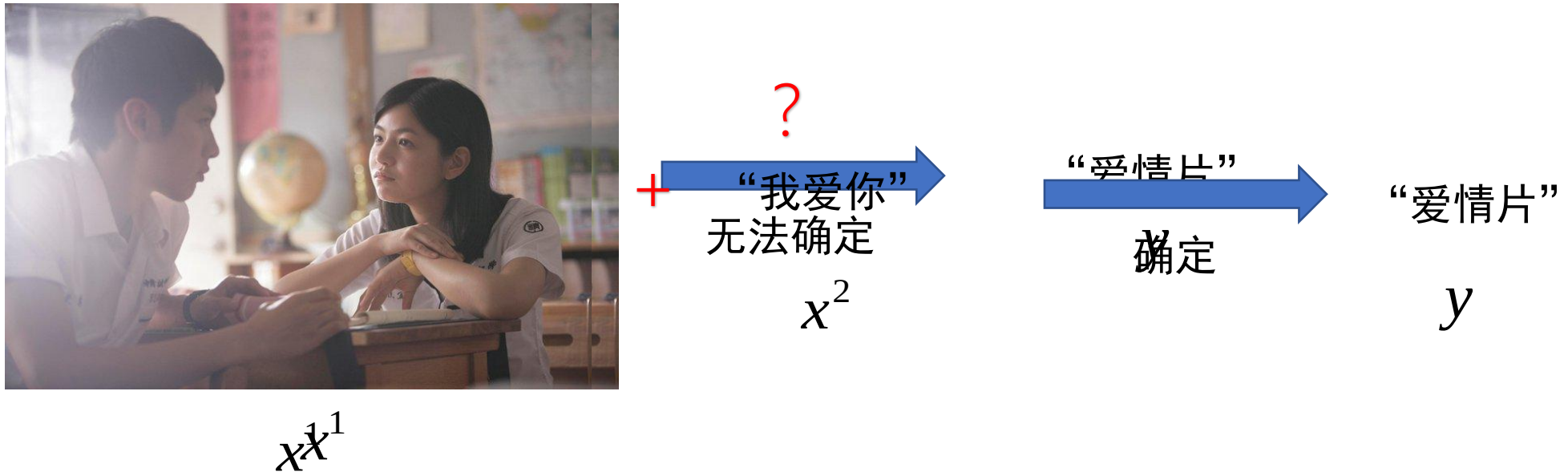
给定一个数据集，我们可将其映射为一个图，数据集中每个样本对应于图中一个结点，若两个样本之间相似度(相关性)很高，则对应结点之间存在一条边，边的“强度”正比于样本之间的相似度(相关性)。



多视图学习(multi-view learning)

给现实应用中，一个数据对象往往同时拥有多个“属性集”，每一个属性集就构成了一个“视图”(view)，多视图学习就会考虑视图在“相容性”的基础下，不同视图信息之间的“互补性”对学习器构建的作用。

一个电影片段可以表示为样本$(< x^1, x^2 >, y)$, 其中x1代表电影图像视图的属性向量，x2代表声音视图中的属性向量，y是标记，为电影的类型，例如“动作片”、“爱情片”等。



深度学习下的半监督学习

半监督深度学习算法：

1. 无标签数据预训练网络后，根据有标签数据进行微调(fine-tune)
2. 有标签数据训练网络，利用从网络中得到的深度特征来做半监督算法
3. 让网络work in semi-supervised fashion

深度学习下的半监督学习

半监督深度学习算法：

1. 无标签数据预训练网络后，根据有标签数据进行微调(fine-tune)

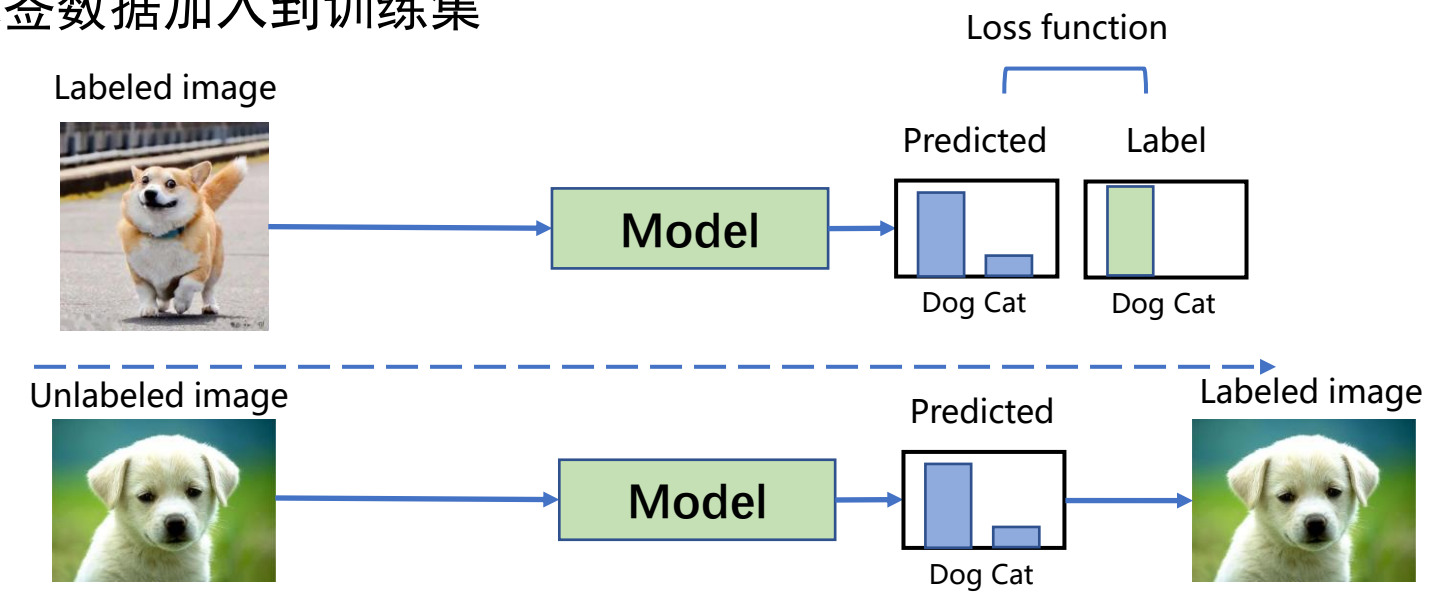
   - 无监督预训练：用所有训练数据训练自动编码器(AutoEncoder)，然后把自编码网络的参数作为初始参数，用有标签数据微调网络。
   - 伪有监督预训练：通过半监督算法或聚类算法等方式，给无标签数据附上伪标签信息，先用伪标签信息来预训练网络，然后再用有标签的数据微调网络

深度学习下的半监督学习

半监督深度学习算法：

2. 有标签数据训练网络，利用从网络中得到的深度特征来做半监督算法

- 先用有标签数据训练网络(容易过拟合)
- 通过隐藏层提取特征，以这些特征来用某种分类算法对无标签数据进行分类
- 挑选认为分类正确的无标签数据加入到训练集
- 重复上述过程



深度学习下的半监督学习

半监督深度学习算法：

3. 让网络work in semi-supervised fashion

Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks(ICML2013)

该文章在伪标签学习中使用深度学习网络作为分类器，将网络对无标签数据的预测，作为其伪标签，再来对网络进行训练。

其主要贡献在于损失函数的构造

$$Loss = \sum_{m=1}^{n} \sum_{i=1}^{C} L(y_i^m, f_i^m) + \alpha(t) \sum_{m=1}^{n'} \sum_{i=1}^{C} L(y_i'^m, f_i'^m)$$

损失函数的第一项是有标签数据的损失，第二项是无标签数据的损失。在无标签数据的损失中，y′ 为无标签数据预测得到的伪标签，是直接取网络对无标签数据的预测的最大值为标签。其中 α（t）决定着无标签数据的代价在网络更新的作用，选择合适的 α（t） 很重要，太大性能退化，太小提升有限。一般初始设置为0，随着训练慢慢增加。

总结

- 传统半监督学习一般分为：
  - 自训练算法
  - 生成式方法
  - 半监督支持向量机
  - 图半监督方法
  - 多视图学习
- 深度学习下的半监督学习一般分为：
  - 无标签数据预训练网络后，根据有标签数据进行微调(fine-tune)
  - 有标签数据训练网络，利用从网络中得到的深度特征来做半监督算法
  - 让网络work in semi-supervised fashion

参考资料：

1. R.J.Brachman, Thomas Dietterich. Synthesis Lectures on Artificial Intelligence and Machine Learning[J].2009
2. D.H. Lee. Pseudo-Label: The Simple and Effcient Semi-Supervised Learning Method for Deep Neural Networks[C]. Proceedings of the International Conference on Machine Learning (ICML), 2013.
3. 半监督深度学习 - Moonx5 - 博客园 (cnblogs.com)

相关论文会放到课程网页中，如有需要请自行下载。