

# Semi-supervised Learning

Yu Xueqing

2022 年 7 月 27 日

## 1 Semi-supervised learning

two assumption:

- low-density separation: 非黑即白, 类间有明显的界限  
self-training: 用 training set 训练模型  $f^*$ , 用  $f^*$  去标记 unlabeled data, 再去训练模型
  - entropy-based regularization: 衡量分类的 distribution 是否好, 计算 distribution 的 entropy, entropy 越小 (属于某类确信度高) 越好  
design loss function:  $L = \sum C(y, \hat{y}) + \sum H(y)$
- smoothness assumption: 近朱者赤近墨者黑  
 $x^1$   $x^2$  在 high density region 近,  $y^1 = y^2$ 
  - cluster and then label(在图像上直接聚类效果不好)
  - graph-based approach: 建图, 有 edge 相连的点归为一类  
 $smoothness S = \frac{1}{2} \sum w_{i,j} (y^i - y^j)^2 = y^T L y$ ,  $L = \sum C(y, \hat{y}) + \lambda S$