

Problem

$$\min_{\alpha^T \mathbf{1}_{|\hat{P}-P|=b}} \left\| \frac{1}{|\hat{P}|} \sum_{x_i \in \hat{P}} \phi(x_i) - \frac{1}{p+b} \left(\sum_{x_i \in P} \phi(x_i) + \sum_{x_j \in \hat{P}-P} \alpha_j \phi(x_j) \right) \right\|^2$$

函数 f 预测结果为正类的分布

现有正例+待查集服从的分布

可能会含有被预测为
正类的负例样本

Reliable Predicted Positive Set

$D_{\hat{P}}$: $f(x)$ 预测为正类的样本

$g(x_i)$: 将 x_i 预测为正类的概率

$$threshold = \min g(x_i), x_i \in D_P \cap D_{\hat{P}}$$

可靠的预测正例集合 R :

$$R = \{x_i / x_i \in D_{\hat{P}-P} \text{ and } g(x_i) > threshold\}$$

- 可靠正例集合: $P + R$
- 候选集合: R

Selector

$$\min_{\alpha^T \mathbf{1}_{|R|=b}} \left\| \frac{1}{|P+R|} \sum_{x_i \in P+R} \phi(x_i) - \frac{1}{p+b} \left(\sum_{x_i \in P} \phi(x_i) + \sum_{x_j \in R} \alpha_j \phi(x_j) \right) \right\|^2$$

R : 当前分类器 f 预测结果为 + 的集合中的可靠正例

P : 已标记正例集合, $p = |P|$

b : 批量查询的个数

α : $\alpha_i \in \{0, 1\}$

If Query a Negative Example

现有的PU经验风险公式：

$$R_{emp}(f) = \frac{\pi}{p} \sum_{x_i \in D_P} \tilde{l}(f(x_i), 1) + \frac{1}{u} \sum_{x_i \in D_U} l(f(x_i), -1), \text{ where } p = |D_P|, u = |D_U|$$



$$R_{emp}(f) = \frac{\pi}{p} \sum_{x_i \in D_P} \tilde{l}(f(x_i), 1) + \frac{1}{u} \sum_{x_i \in D_U} l(f(x_i), -1) + \frac{\beta}{n} \sum_{x_i \in D_N} l(f(x_i), -1),$$

where $p = |D_P|, u = |D_U|, n = |D_N|$

为了让分类器将已标注的负类样本分对，设置一个较大的 β 值，加大对负类样本分错的惩罚力度。

$$f^* = \arg \min_{f \in H_k} \left\{ R_{emp}(f) + \lambda \|f\|_{H_k}^2 \right\}$$