

Subspace Clustering with Active Learning.

- Active learning: Iteratively annotating the potentially misclassified data.
- Subspace Clustering: Points are clustered according to their underlying subspaces.

1. Active Learning Framework

- ① 找出影响最大的误分类点
- ② 更新有标签点的聚类分配

1.1 K-Subspace Clustering

子空间 $S_k = \{x \in \mathbb{R}^p : x = V_k y\}$, V_k 是 S_k 的基

Loss function:

$$L(x_i, V_{k_i}) = \|x_i - V_{k_i} V_{k_i}^T x_i\|_2^2$$

Objective function:

$$f(X, V) = \min_{V_1, \dots, V_K} \sum_{i=1}^N \min_{k_i \in \{1, \dots, K\}} L(x_i, V_{k_i})$$

cluster assignment
(k_1, \dots, k_N) ???

1.2. Query Procedure

$$U_1(x_s, V_{k_s}) = \sum_{x \in \mathcal{X}_{k_s}} L(x, V_{k_s}) - \sum_{x \in \mathcal{X}_{k_s} \setminus \{x_s\}} L(x, \tilde{V}_{k_s})$$

从 k_s 聚类中移出点 x_s , 结构损失的减少量, 即 $U_1(x_s, V_{k_s})$
此时, x_s 将被分入聚类 k_s^*

$$k_s^* = \arg \min_{k \in \{1, \dots, K\} \setminus \{k_s\}} L(x_s, V_k)$$

同理, 定义 $U_2(x_s, V_{k_s^*})$ 为将 x_s 加入 k_s^* 聚类的结构损失的增量.

$$U_2(x_s, V_{k_s^*}) = \sum_{x \in \mathcal{X}_{k_s^*} \cup \{x_s\}} L(x, \tilde{V}_{k_s^*}) - \sum_{x \in \mathcal{X}_{k_s^*}} L(x, V_{k_s^*})$$

因此, 易误分类点 x_s^* :

$$x_s^* = \arg \max_{x_s \in \mathcal{X}_U} \{U_1(x_s, V_{k_s}) - U_2(x_s, V_{k_s^*})\}$$

其中, \mathcal{X}_U : 无 label 点集合.

\mathcal{X}_L : labelled 点集合

1.3 Update Procedure

属于一类的查询点应聚成一类

将 KSC 扩展成 iterative constrained clustering algorithm

① 估计子空间的基.

② 将每个点聚到最近的子空间

③ 满足分组约束条件 $\begin{cases} \mathcal{X}_U \\ \mathcal{X}_L \end{cases}$