

Cost-Effective Active Learning for Hierarchical Multi-Label Classification

U: uncertainty c: cost f: prediction y: groundtruth
H: hierarchical correlation

$$\min_{H, \alpha} -tr((UH \square P_1)\alpha) + \lambda c(P_2 \square \alpha) \mathbf{1} + \frac{1}{2} \|Y^T Y - H\|^2$$

$$s.t. \quad 0 \leq \alpha_{ij} \leq 1, \quad \sum \alpha_{ij} = k$$

Fix α

$$\begin{aligned} J &= -\text{tr}((UH \square P_1)\alpha) + \frac{1}{2} \|Y^T Y - H\|^2 + \text{constant} \\ &= -\text{tr}\left((\alpha^T \square P_1)^T UH\right) + \frac{1}{2} \sum_{i=1}^L (YY_i - h_i)^T (YY_i - h_i) + \text{constant} \\ &= -\sum_{i=1}^L m_i h_i + \frac{1}{2} \sum_{i=1}^L (h_i^T h_i - 2YY_i^T h_i + YY_i^T YY_i) + \text{constant} \\ &= \frac{1}{2} \sum_{i=1}^L h_i^T h_i - \sum_{i=1}^L (YY_i^T + m_i) h_i + \text{constant} \end{aligned}$$

$$M = (\alpha^T \square P_1)^T U$$

Fix $\mathbf{H}$

$$J = -\text{tr}((U\mathbf{H} \oslash \mathbf{P}_1)\boldsymbol{\alpha}) + \lambda c(\mathbf{P}_2 \oslash \boldsymbol{\alpha})\mathbf{1}$$

$$= -\text{tr}(\mathbf{S}\boldsymbol{\alpha}) + \sum_{i=1}^L \sum_{j=1}^n \lambda c_i p_{ij} \alpha_{ij}$$

$$= -\sum_{i=1}^L \alpha_i s_i + \sum_{i=1}^L \lambda c_i \alpha_i p_i^T$$

$$= \sum_{i=1}^L \alpha_i (\lambda c_i p_i^T - s_i)$$

$$= \hat{\boldsymbol{\alpha}} \hat{\mathbf{q}} \qquad q = \lambda c_i p_i^T - s_i$$

$$\text{s.t. } 0 \leq \hat{\alpha}_i \leq 1, \quad \sum_{i=1}^{Du*L} \hat{\alpha}_i = k$$

U: uncertainty c: cost f: prediction y: groundtruth

H: hierarchical correlation

$$\min_{H, \alpha} -tr((UH \square P_1)\alpha) + \lambda c(P_2 \square \alpha)\mathbf{1} + \frac{1}{2} \|Y^T Y - H\|^2 + \frac{1}{2} \|H\|^2$$

$$s.t. \quad 0 \leq \alpha_{ij} \leq 1, \quad \sum \alpha_{ij} = k, \quad 0 \leq h_{ij} \leq h_{\max} \quad h_{\max} = \sqrt{P_{\inf}}$$