

Cost-Effective Active Learning for Hierarchical Multi-Label Classification

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

$$\min_{H, \alpha} -tr(UH\alpha) + \lambda c \alpha \mathbf{1} + \sum_{i \in D_I} \frac{1}{2} \| f_i^T H - y_i^T \|^2$$

$$\min_{H, \alpha} -tr((UH \square P)\alpha) + \lambda c (P^T \square \alpha) \mathbf{1} + \sum_{i \in D_I} \frac{1}{2} \| f_i^T H - y_i^T \|^2$$

Fix α

$$\begin{aligned} J &= -\text{tr}((UH \square P)\alpha) + \sum_{k \in Dl} \frac{1}{2} \|f_k^T H - y_k^T\|^2 + \text{constant} \\ &= -\text{tr}(\alpha^T \square P)^T UH) + \frac{1}{2} \sum_{k \in Dl} (f_k^T H - y_k^T)(f_k^T H - y_k^T)^T + \text{constant} \\ &= -\sum_{i=1}^L m_i h_i + \frac{1}{2} \sum_{i=1}^L \sum_{k \in Dl} h_i^T f_k f_k^T h_i - \sum_{i=1}^L \sum_{k \in Dl} y_{ki} f_k^T h_i + \text{constant} \\ &= \frac{1}{2} \sum_{i=1}^L \sum_{k \in Dl} h_i^T f_k f_k^T h_i - \sum_{i=1}^L (\sum_{k \in Dl} y_{ki} f_k^T + m_i) h_i + \text{constant} \\ &= \frac{1}{2} \sum_{i=1}^L h_i^T G h_i - \sum_{i=1}^L b_i h_i + \text{constant} \end{aligned}$$

$$G = \sum_{k \in Dl} f_k f_k^T \quad b_i = \sum_{k \in Dl} y_{ki} f_k^T + m_i$$

Fix H

$$\begin{aligned}
 J &= -\text{tr}((UH \square P)\alpha) + \lambda c \left(P^T \square \alpha \right) \mathbf{1} \\
 &= -\text{tr}(S\alpha) + \sum_{i=1}^L \sum_{j=1}^n \lambda c_i p'_{ij} \alpha_{ij} \\
 &= -\sum_{i=1}^L \alpha_i s_i + \sum_{i=1}^L \lambda c_i \alpha_i p_i^T \\
 &= \sum_{i=1}^L \alpha_i \left(\lambda c_i p_i^T - s_i \right) \quad q = \lambda c_i p_i^T - s_i \\
 &= \hat{\alpha} \hat{q}
 \end{aligned}$$

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