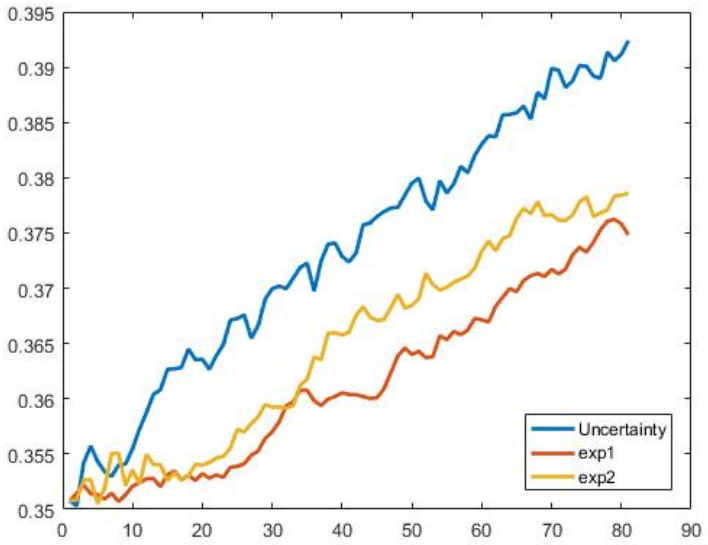
Cost-Effective Active Learning for Hierarchical Multi-Label Classification

U: uncertainty c: cost Y: groundtruth
H: hierarchical correlation

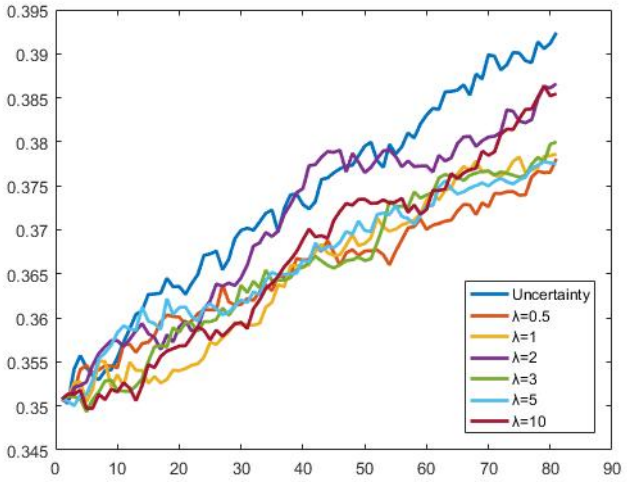
$$\min_{H,\alpha} \; -tr\left((UH \square P_1)\alpha\right) + \lambda c\left(P_2 \square \alpha\right)\mathbf{1} + \frac{1}{2}\| Y^T Y - H \|^2$$

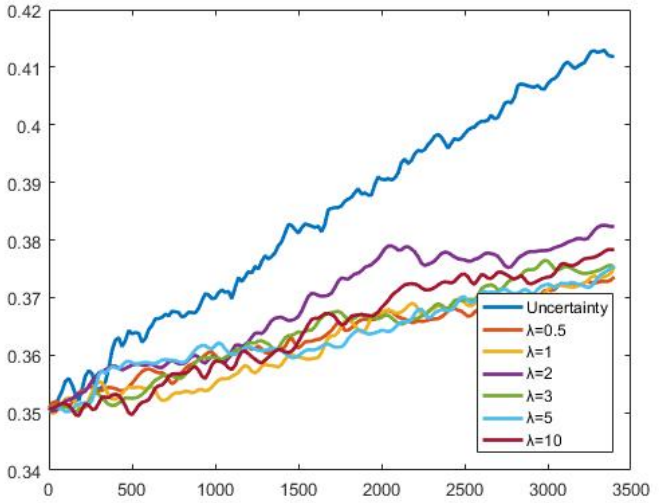
Exp1 $s.t. \;\; 0 \le \alpha_{ij} \le 1, \quad \sum \alpha_{ij} = k, \quad 0 \le h_{ij} \le h_{\max} \quad h_{\max} = \sqrt{P_{\inf}}$

Exp2 $s.t. \;\; 0 \le \alpha_{ij} \le 1, \quad \sum \alpha_{ij} = k, \quad 0 \le h_{ij} \le 1$

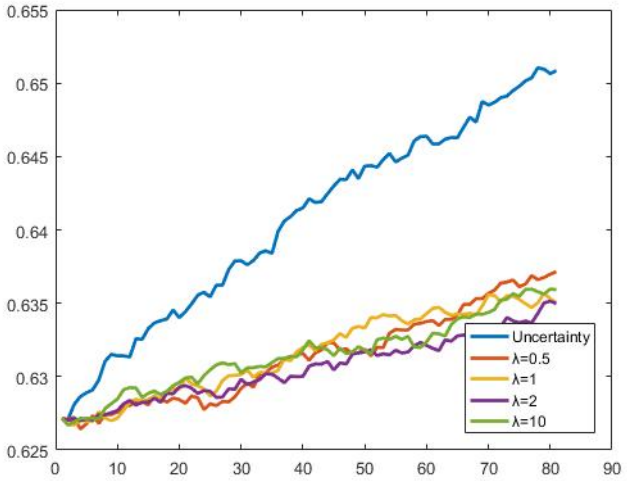


$$\min_{H,\alpha} \; -tr\left((UH \square P_1)\alpha\right) + \lambda c\left(P_2 \square \alpha\right)\mathbf{1} + \frac{1}{2}\| Y^T Y - H \|^2 \qquad s.t. \;\; 0 \le \alpha_{ij} \le 1, \;\; \sum \alpha_{ij} = k, \;\; 0 \le h_{ij} \le 1$$





Yeast-go



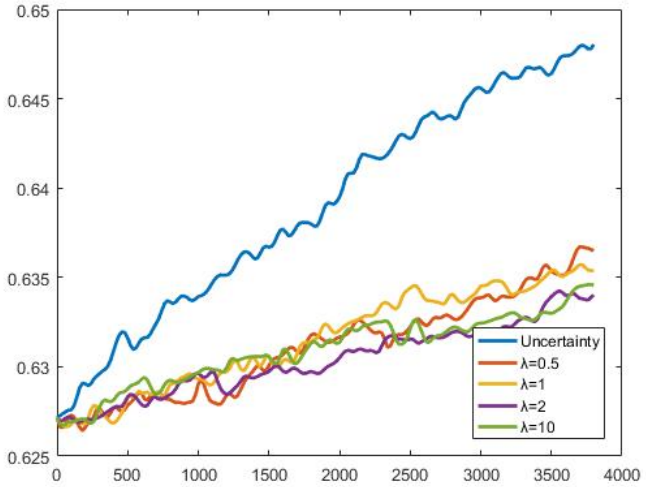


Image07D

Fix α

$$J = -tr((UH \square P_1)\alpha) + \frac{1}{2}\| Y^T Y - H \|^2 + \text{constant}$$

$$= -tr\left((\alpha^T \square P_1)^T UH\right) + \frac{1}{2}\sum_{i=1}^{L}(YY_i - h_i)^T (YY_i - h_i) + \text{constant}$$

$$= -\sum_{i=1}^{L} m_i h_i + \frac{1}{2}\sum_{i=1}^{L}(h_i^T h_i - 2YY_i^T h_i + YY_i^T YY_i) + \text{constant}$$

$$= \frac{1}{2}\sum_{i=1}^{L} h_i^T h - \sum_{i=1}^{L}(YY_i^T + m_i)h_i + \text{constant}$$

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

Fix H

$$J = -tr((UH \square P_1)\alpha) + \lambda c\left(P_2 \square \alpha\right)\mathbf{1}$$

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

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

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

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

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