Hedge and Exp3 for sequential decision-making

- Written by Tommaso R. Cesari and Nicolo Cesa-Bianchi
- https://cesa-bianchi.di.unimi.it/Algo2/Note/hedge-exp3.pdf

问题描述

- 算法对一系列依次到达的请求做出相应
- 随机/对抗环境
- 存在多个专家给出预测
- 目标：最小化累计损失，使其接近最好专家的表现

Prediction from expert advice

- 专家{1,...,K},可以被决策者和(对抗)环境选择
- 在每一轮t,决策者选择专家I_t,产生损失$\ell_t(I_t)$,并揭示所有专家的损失$\ell_t(i) \in [0,1], \ell_t \in [0,1]^K$
- 决策的评估:
  - $\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\ell_t(I_t)\right] - \min_{i=1,...,K}\left(\frac{1}{T}\sum_{t=1}^{T}\ell_t(i)\right)$
  - 目标: 找到合适的决策算法，使$T \to \infty$时，上式趋近于0
- 决策算法1：选择在过去表现最好的专家
  - $I_t = \arg \min_{i=1,\dots,K} \sum_{s=1}^{t-1} \ell_s(i)
  - 缺陷:考虑每次为0的环境,专家1决策$0,1,0,1......$，专家2决策$\frac{1}{2},0,1,0......$，则每次都会基于之前的损失选择最差的专家，产生线性遗憾
- 决策算法2：Hedge
  - 参数:学习率$\gamma \in (0,1)$
  - 初始化:权重$w_1(i) = 1$
  - for t=1,2,...,T:
    - $W_t = \sum_{i=1}^{K} w_t(i)$,概率分布$p_t(i) = \frac{w_t(i)}{W_t}$
    - 根据概率分布p_t选择专家I_t
    - 产生损失$\ell_t(I_t)$,更新权重$w_{t+1}(i) = w_t(i) \cdot e^{-\gamma \ell_t(i)}$
-

$$\frac{W_{t+1}}{W_t} = \sum_{i=1}^{K}\frac{w_{t+1}(i)}{W_t} = \sum_{i=1}^{K}\frac{w_t(i)e^{-\gamma\ell_t(i)}}{W_t} = \sum_{i=1}^{K}p_t(i)e^{-\gamma\ell_t(i)} \qquad \leq \sum_{i=1}^{K}p_t(i)\left(1-\gamma\ell_t(i)+\frac{\gamma^2}{2}\ell_t(i)^2\right) \quad = 1-\gamma\sum_{i=1}^{K}p_t(i)\ell_t(i)+\frac{\gamma^2}{2}\sum_{i=1}^{K}p_t(i)\ell_t(i)^2$$

$\ln\frac{W_{t+1}}{W_t} \leq ln(1-\gamma\sum_{i=1}^{K}p_t(i)\ell_t(i)+\frac{\gamma^2}{2}\sum_{i=1}^{K}p_t(i)\ell_t(i)^2)$$ \ln \left(\frac{W_T}{W_1} \right) = \sum_{t=1}^{T} \ln \left(\frac{W_{t+1}}{W_t} \right) \leq -\gamma \sum_{t=1}^{T} \sum_{i=1}^{K} p_t(i)\ell_t(i) + \frac{\gamma^2}{2} \sum_{t=1}^{T} \sum_{i=1}^{K} p_t(i)\ell_t(i)^2\quad[\ln(1+z)\leq z]\\\ln \left(\frac{W_T+1}{W_1} \right) \geq \ln \left(\frac{w_{T+1}(k)}{W_1} \right) = -\gamma \sum_{t=1}^{T} \ell_t(k) - \ln(K) $