IE613: Online Machine Learning **Jan-Apr 2019**

Solution to Midterm 1

February 23M. K. HanawalYour-name

Solution 1

(a) *We know that* $M_{Hav}(\mathcal{H}) \geq Ldim(\mathcal{H})$ *and* $Ldim(\mathcal{H}) = 1$. *Whenever halving algorithms errs for the first time, all the hypothesis except the correct one will be eliminated, hence* $M_{Hav}(\mathcal{H})$*=1.*

(b) *Let* $\mathcal{H}$ *be the set of all binary maps on* $\mathcal{X} = \{x_1, x_2, \ldots, x_d\}$. *Then* $|\mathcal{H}| = 2^d$. *For any* $x \in \mathcal{X}$, *exactly half of the hypotheses in* $\mathcal{H}$ *assign label* 1 *and the other half assign label* 0. *Thus, whenever the halving algorithm makes an error, exactly half of the hypothesis will be eliminated. Then, after* $\log_2(|\mathcal{H}|)$ *mistakes a single hypothesis remains which is the correct hypothesis (by realizability assumption.). Hence for any input sequences the halving algorithm using the hypothesis class* $\mathcal{H}$ *makes at most* $\log_2(|\mathcal{H}|)$ *mistakes.*

(c) *For the above hypothesis class, the consistent algorithm also makes the same number of errors.*

Solution 2

(a) *Uniform on* $[K]$.

(b) *Lagrangian function is*

$$L(P, \lambda, \beta_i) = \sum_{s=1}^{t} \sum_{i=1}^{K} l_s^i P_i + \frac{1}{\eta} \sum_{i=1}^{K} P_i \log\left(\frac{1}{P_i}\right) + \lambda(\sum_{i=1}^{K} P_i - 1) + \sum_{i=1}^{K} \beta_i P_i$$

where λ *is unrestricted and* $\beta_i \geq 0 \forall i = 1...., K$.
Now, KKT conditions are given as follows:

$$\frac{\partial L(P, \lambda, \beta_i)}{\partial P_i} = \sum_{i=1}^{K} l_s^i + \frac{1}{\eta} \log\left(\frac{1}{P_i}\right) + \frac{1}{\eta}(-1) + \lambda + \beta_i = 0 \quad \forall i = 1...., K; \tag{1.1}$$

$$\lambda(\sum_{i=1}^{K} P_i - 1) = 0; \tag{1.2}$$

$$\beta_i P_i = 0 \quad \forall i = 1...., K. \tag{1.3}$$

Let $L_t^i = \sum_{i=1}^{K} l_s^i$ *from* (1.1), *we get*

$$L_t^i + \frac{1}{\eta}\left(\log\left(\frac{1}{P_i}\right) + 1\right) + \lambda + \beta_i = 0$$

$$\textit{Simplifying we get} \quad P_i = \exp(\eta(L_t^i - \frac{1}{\eta} + \lambda + \beta_i))$$

Since, $P_i = \exp(\eta(L_t^i - \frac{1}{\eta} + \lambda + \beta_i)) \geq 0, \implies \beta_i = 0$
Also. $\sum_{i=1}^{K} P_i = 1$

$$\implies \sum_{i=1}^{K} \exp(\eta(L_t^i - \frac{1}{\eta} + \lambda)) = 1$$

$$\implies \exp(\eta\lambda - 1) = \frac{1}{\sum_{i=1}^{K} \exp(\eta L_t^i)}$$

Substituting this in expression of P_i*, we get*

$$P_i = \frac{\exp(\eta L_t^i)}{\sum_{j=1}^{K} \exp(\eta L_t^j)} \quad \forall i = 1, 2, ..., K$$

(c) *For weighted majority,*

$$P_t^i = \frac{\exp(-\eta' L_t^i)}{\sum_{j=1}^{K} \exp(-\eta' L_t^j)} \quad \textit{for some } \eta' > 0$$

P_t *in FRL has the same structure for as in weighted majority.*

(d) *Weighted majority performs optimally when*

$$\eta' = \sqrt{\frac{2 \log K}{T}}$$

So, we would set

$$\eta = -\sqrt{\frac{2 \log K}{T}}$$

Solution 3 *The error rate of a sensor is given by* $\gamma_i = \mathcal{P}_r\{Y \neq \hat{Y}^i\}$ *where*Y *is the true label and* $\hat{Y}^i$ *is that estimated by sensor* i*.*

We can treat the sensors as the arms of a bandit & estimate their error by comparing prediction of the sensor with the true labels. We will have smaller number of mistakes if we can identify the sensor with smallest error rate. Since the instance come in an i.i.d. fasion, the underlying distribution is fixed and the error rate is fixed for the underlying distribution. We can treat each sensor as an arm in a stochastic mult-armed bandit problem with the mean of i-th arm set to γ_i *(loss)*

(a) *We define loss incurred by playing sensor* i *in round* t *as* $1_{\{Y_t \neq \hat{Y}_t^i\}}$*. The regret of an algorithm that plays arm/sensor* I_t *in round* t *as*

$$\mathcal{R}_T = \sum 1_{\{Y_t \neq \hat{Y}_t^{I_t}\}} - \min_i \sum 1_{\{Y_t \neq \hat{Y}_t^i\}}$$

Expected (pseudo) regret is then,

$$\begin{aligned}\mathbb{E}[\mathcal{R}_T] &= \sum_{t=1}^{T} \gamma_{I_t} - \min_i \sum_{t=1}^{T} \gamma_i \quad \textit{where} \\ &= \sum_{t=1}^{T} \gamma_{I_t} - T\gamma^* \quad \textit{where } \gamma^* = \min_i \gamma_i\end{aligned}$$

(b) *Since this is stochastic multi-armed bandit, we can use any of UCB, KL-UCB or Thomson sampling algorithms.*

(c) *We get regret bound of* $O\left(\sqrt{TK\log K}\right)$ *or* $O\left(\sum_{i=1}^{K}\frac{\log T}{\Delta_i}\right)$ *where* $\Delta_i = \gamma_i - \gamma^*$ *If observation are delayed by* S *rounds, we will incur an extra regret of* $\Delta_{\max}S$*, where* $\Delta_{\max} = \max_{i\in[K]}\Delta_i$.

Solution 4

(a) *If we don't know the environment is stochastic, we will consider the worst case & use the algorithms for adversarial setting.*

(b) *If the environment remains stochastic for a fixed duration, it is better to use stochastic bandit algorithm is those rounds. Since we don't know at which point the environment changes the arm distribution, one option is to detect the point when changes point distribution (How?) and when the is detected restart the algorithm. Other option possibility is we only take few past samples (within a window) to estimate the mean reward instead of all the samples we have. In this case, new optimal arm will be eventually detected if it distribution changes (what should be the window size?).*