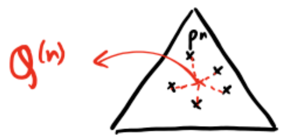
Lecture 16

Review: * Probability Assignment and Universal Portfolio

$$\rightarrow \; r(k,n) = \min_{Q^{(n)}} \max_{P} D(P^n \| Q^{(n)}) = \frac{k-1}{2}\log n + O_k(1)$$



$\mathcal{P}([k]^n)$

$$\rightarrow \; r_{k,n} = \min_{\hat\theta} \max_{P \in \mathcal{P}_k} \max_{\substack{\{x_t\}_{t=1}^n : \\ x_t \in \mathbb{R}_+^k}} \log \frac{\prod_{t=1}^n \sum_{j=1}^k P_j x_{t,j}}{\underbrace{\prod_{t=1}^n \sum_{j=1}^k \hat\theta(j|x^{t-1})\, x_{t,j}}_{T_n(\underline{x},\hat\theta)}}$$

$$T_n(\underline{x},\hat\theta) = \sum_{\underline{j} \in [k]^n} \underline{\theta}(\underline{j}|\underline{x})\, \underline{x}(\underline{j}),$$

where

$$\underline{x}(\underline{j}) = \prod_{t=1}^n x_{t,j_t} \quad \text{and} \quad \underline{\theta}(\underline{j}|\underline{x}) = \prod_{t=1}^n \hat\theta_t(j_t|x^{t-1}),$$

and so,

$$\sum_{\underline{j}} \hat\theta(\underline{j}|\underline{x}) = 1.$$

* Using probability assignment for universal portfolio

$$\rightarrow \; r_{k,n} = r(k,n)$$

We saw that any prob. assignment Q, the strategy

$$\hat\theta_Q(j|x^{t-1}) = \frac{\sum_{j^{t-1}} Q(j|j^{t-1})\, Q(j^{t-1})\, \underline{x}(j^{t-1})}{\sum_{j^{t-1}} Q(j^{t-1})\, \underline{x}(j^{t-1})}, \quad j = 1,\dots,k$$

satisfies

$$T_n(\underline{x},\hat\theta_Q) = \sum_{\underline{j} \in [k]^n} Q(\underline{j})\, \underline{x}(\underline{j}).$$

Agenda: * Bayesian interpretation of universal portfolios

* Prediction / multiarmed bandit / online learning problem

A Bayesian interpretation of universal portfolio

Consider a joint distribution on $\underline{j} \in [k]^n$, $\underline{x} \in \mathbb{R}^{kn}$ given by

$$(a) \quad Q(\underline{x} \mid \underline{j}) \propto \underline{x}(\underline{j}),$$

$$(b) \quad Q(j_{n+1} \mid \underline{x}, \underline{j}) = Q(j_{n+1} \mid \underline{j}).$$

Then,

$$Q(j_{n+1} \mid \underline{x}) = \frac{\sum_{\underline{j}} Q(\underline{j}, j_{n+1}, \underline{x})}{\sum_{\underline{j}} Q(\underline{j}, \underline{x})} = \frac{\sum_{\underline{j}} Q(j_{n+1} \mid \underline{j})\, Q(\underline{j})\, \underline{x}(\underline{j})}{\sum_{\underline{j}} Q(\underline{j})\, \underline{x}(\underline{j})}$$

$$= \hat{\theta}_n(j_{n+1} \mid \underline{x})$$

Furthermore, the specific Q that we used, Q_{KT}, itself is a Bayesian estimator obtained by using Dirichlet prior on $\mathcal{P}_k$. On the other hand, we are competing with experts who choose $Q(\underline{j}) = P^n(\underline{j})$ for some $P \in \mathcal{P}_k$.

Note that for $Q = Q_{KT}$, we have

$$\sum_{\underline{j}} Q_{KT}(\underline{j})\, \underline{x}(\underline{j}) = \sum_{\underline{j}} \int_{\mathcal{P}_k} \prod_{t=1} P(j_t)\, x_{t,j_t}\, \pi(dP)$$

$$= \int_{\mathcal{P}_k} \prod_{t=1}^{n} \left(\sum_{j=1}^{k} P(j)\, x_{t,j} \right) \pi(dP).$$

The integrand above can be computed in kn steps, but the k-fold integration makes the complexity exponential in k. Computing $\hat{\theta}_q$ just requires such computations. (3)

[B] Prediction with expert advice

The setting of universal portfolio reflects a typical situation in prediction:

(1) At each time, the algorithm predicts/makes a decision.

(2) Based on this decision, we get a reward and perhaps get to know some other things (in the universal portfolio, decision is $\hat{\theta}_t(\cdot|x^{t-1})$, reward is $r_t = \sum_i \hat{\theta}_t(i|x^{t-1}) x_{t,i}$, and we get to know x_t).

* Set of experts $\mathcal{E}$
* At each time t, we make a prediction $\hat{p}_t \in \mathcal{D}$ based on our past observations.

  Similarly, each expert $i \in \mathcal{E}$ makes a prediction $p^*_{t,i} \in \mathcal{D}$.
* We observe y_t after the decision.
* The loss for the prediction $p \in \mathcal{D}$ at time t is given by

$$l_t(p, y_t)$$

where $y_t \in \mathcal{Y}$ is the unknown state of the nature at at time t.

The "regret" our algorithm has for not following the advice of expert $i \in \mathcal{E}$ at time t is

$$\ell_t(\hat{p}_t, y_t) - \ell_t(p^*_{t,i}, y_t),$$

and the overall regret till time n is

$$\sum_{t=1}^{n} \ell_t(\hat{p}_t, y_t) - \ell_t(p^*_{t,i}, y_t).$$

Our goal is to compete with the best expert in hindsight. So, for any strategy $\hat{\theta}$,

$$r_n(\hat{\theta}, \underline{y}) = \max_{i \in \mathcal{E}} \sum_{t=1}^{n} \ell_t(\hat{p}_t, y_t) - \sum_{t=1}^{n} \ell_t(p^*_{t,i}, y_t)$$

$$= \sum_{t=1}^{n} \ell_t(\hat{p}_t, y_t) - \underbrace{\min_{i \in \mathcal{E}} \sum_{t=1}^{n} \ell_t(p^*_{t,i}, y_t)}_{\text{best expert in hindsight.}}$$

$$r_n(\mathcal{E}) = \min_{\hat{\theta}} \max_{\underline{y}} r_n(\hat{\theta}, \underline{y})$$

→ Universal portfolio is a special case with

- $\mathcal{D} \equiv \mathcal{P}_k$
- $\mathcal{E} \equiv \mathcal{P}_k$ and $p^*_{t,P} \equiv \hat{\theta}_{P^n}$, $P \in \mathcal{P}_k$,
- $y_t = x_t \in \mathbb{R}_+^k$,
- $\ell(\hat{p}_t, y_t) = \log \sum_{j=1}^{k} \hat{p}_t(j)\, y_{t,j}$

→ In many cases, we only need to handle finite (but large) set $\mathcal{E}$. In fact, we will see one such approximation for the portfolio problem.

C Multiarmed bandits

Another popular formulation which is very similar to the prediction formulation above puts two restrictions

(1) Decisions: we can only choose to follow one of the experts I_t at time t.

(2) Observations: we only get to know our loss at time t (and not the state of the nature y_t).

We need to minimize the regret for the best expert in the hindsight.