Lecture 16

Review: * Probability Assignment and Universal Portfolio

\[ r(k, n) = \min_{\theta} \max_{P \in P_k} D(\textbf{P}^{\text{out}} \| Q^{(k)}) = \frac{k}{2} \log n + O(1) \]

\[ \text{with} \quad P \left( \{k\}^n \right) \]

\[ r_{k, n} = \min_{\hat{\theta}} \max_{P \in P_k} \max_{\{x_t^i \in \mathbb{R}^k_+\} \text{ s.t. } x_t^i \in \mathbb{R}^k_+} \log \frac{\prod_{t=1}^{n} \sum_{j=1}^{k} P_{t,j} x_{t,j}}{\prod_{t=1}^{n} \sum_{j=1}^{k} \hat{\theta}(j \mid x^{t-1}) x_{t,j}} \]

\[ \Gamma_n(x, \hat{\theta}) = \sum_{j \in \{k\}^n} \hat{\theta}(j \mid x) x(j), \]

where

\[ x(j) = \prod_{t=1}^{n} x_{t,j} \text{ and } \theta(j \mid x) = \prod_{t=1}^{n} \hat{\theta}(j_t \mid x^{t-1}), \]

and so

\[ \sum_{j} \hat{\theta}(j \mid x) = 1. \]

* Using probability assignment for universal portfolio

\[ r_{k, n} = r(k, n) \]

We saw that any prob assignment \( \theta \), the strategy

\[ \hat{\theta}_Q(j \mid x^{t-1}) = \frac{\hat{\theta}(j \mid x^{t-1})}{\sum_{j=1}^{k} \hat{\theta}(j \mid x^{t-1})} \]

satisfies

\[ \Gamma_n(x, \hat{\theta}_Q) = \sum_{j \in \{k\}^n} \hat{\theta}(j) x(j). \]
Agenda:

* Bayesian interpretation of universal portfolios

* Prediction / multiarmed bandit / online learning problem

A Bayesian interpretation of universal portfolio

Consider a joint distribution on \(j \in \mathbb{R}^n\), \(x \in \mathbb{R}^{kn}\) given by

(a) \(Q(x | j) \propto x(j)\),

(b) \(Q(j_{nm} | x, j) = Q(j_{nm} | j)\).

Then,

\[
Q(j_{nm} | x) = \frac{\sum_{j} Q(j_{nm}, j, x)}{\sum_{j} Q(j, x)} = \frac{\sum_{j} Q(j_{nm}, x | j) Q(j | x) x(j)}{\sum_{j} Q(j | x) x(j)}
\]

\[
= \hat{Q}_n(j_{nm}, x)
\]

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

Note that for \(Q = Q_{KT}\), we have

\[
\sum_{j} Q_{KT}(j) x(j) = \sum_{j} \int_{P_k} \prod_{t=1}^{n} P(j_t | x_{t,i_t}) \pi(dp) = \int_{P_k} \prod_{t=1}^{n} \left( \sum_{j=1}^{k} P(j | x_{t,i_t}) \right) \pi(dp).
\]
The integrand above can be computed in $kn$ steps, but the $k$-fold integration makes the complexity exponential in $k$. Computing $\hat{\Theta}_t$ just requires such computations.

**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 $\hat{\Theta}_t(x_t)$, reward is $r_t = \sum_i \hat{\Theta}_t(i|x_t) x_{t,i}$, and we get to know $x_t$).

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

Similarly, each expert $i \in \Xi$ makes a prediction $p_{t,i} \in \mathcal{B}$.
* We observe $y_t$ after the decision.
* The loss for the prediction $p \in \mathcal{B}$ at time $t$ is given by $l_t(p, y_t)$, where $y_t \in \mathcal{Y}$ is the unknown state of nature at time $t$. 
The "regret" our algorithm has for not following the advice of expert \( i \in \mathcal{E} \) at time \( t \) is
\[
l_t(\hat{\rho}_t, y_t) - l_t(\rho^*_{t,i}, y_t),
\]
and the overall regret till time \( n \) is
\[
\sum_{t=1}^{n} l_t(\hat{\rho}_t, y_t) - l_t(\rho^*_{t,i}, y_t).
\]
Our goal is to compete with the best expert in hindsight. So, for any strategy \( \hat{\mu} \),
\[
\mathcal{R}_n(\hat{\mu}, y) = \max_{i \in \mathcal{E}} \sum_{t=1}^{n} l_t(\hat{\rho}_t, y_t) - \sum_{t=1}^{n} l_t(\rho^*_{t,i}, y_t)
\]
\[
= \sum_{t=1}^{n} l_t(\hat{\rho}_t, y_t) - \min_{i \in \mathcal{E}} \sum_{t=1}^{n} l_t(\rho^*_{t,i}, y_t)
\]
\[
\mathcal{R}_n(\mathcal{E}) = \min_{\hat{\mu}} \max_{y} \mathcal{R}_n(\hat{\mu}, y) \quad \text{best expert in hindsight.}
\]
→ Universal portfolio is a special case with
- \( \mathcal{D} = \mathcal{P}_k \)
- \( \mathcal{E} = \mathcal{P}_k \) and \( \rho^*_{t,p} = \hat{\theta}_{\mathcal{P}_n}, p \in \mathcal{P}_k \),
- \( y_t = x_t \in \mathbb{R}^k \),
- \( l(\hat{\rho}_t, y_t) = \log \sum_{j=1}^{k} \hat{\rho}_t(j) y_t,j \)
→ In many cases, we only need to handle finite (but large) sets \( \mathcal{E} \). In fact, we will see one such approximation for the portfolio problem.
Multinomial 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.