CMSC 35900 (Spring 2008) Learning Theory **Lecture: 5**

Game Playing, Boosting

Instructors: Sham Kakade and Ambuj Tewari

1 Game Playing

We can consider a game where, at each round, we choose a distribution $p \in \Delta[n]$ (a distribution over n outcomes) and nature chooses a distribution $q \in \Delta[m]$. The loss for us playing i and nature playing j is $M(i, j)$, in which case nature obtains a reward of $M(i, j)$ (which is the zero sum property). Our expected loss is:

$$p^\dagger M q$$

if we view M as a matrix.

The minimax theorem states that:

$$\min_p \max_q p^\dagger M q = \max_q \min_p p^\dagger M q$$

In other words, in the worst case, we can guarantee the same amount of reward regardless of if we play first or second. The strategies achieving this equality are the minimax strategies, and the associated value is the minimax value.

Let us now prove this theorem. One direction is simple:

$$\max_q \min_p p^\dagger M q = \min_p p^\dagger M q^* \leq \min_p \max_q p^\dagger M q$$

where q^* is the distribution which achieves the max.

We now prove the other direction. We also show how we can achieve the minimax value in a repeated game setting. Assume we know M. The repeated game setting is one where at each round, simultaneously, we choose a strategy p_t and nature chooses a strategy q_t. After round t, q_t is revealed to us. If we interpret Mq_t as our loss vector at time t, then this is a case of the "experts" setting, where the cost function at time t is:

$$c_t(w) = w^\dagger M q_t$$

Our regret is defined as:

$$R_T = \sum_t p_t^\dagger M q_t - \min_{p \in \Delta[n]} \sum_t p^\dagger M q_t$$

In particular, assume that nature chooses q_t with knowledge of our algorithm, so that:

$$q_t \in \mathrm{argmax}_{q \in \Delta[m]} p_t^\dagger M q$$

In other words, at each round nature is choosing her best strategy. Even if nature does this, we know there exists a strategy, such that:

$$\frac{R_T}{T} \leq 2\sqrt{\frac{\log n}{T}}\,.$$

Note that this quantity tends to 0 as $T \to \infty$.

Define $\overline{p} = \frac{1}{T}\sum_{t=1}^{T} p_t$ and $\overline{q} = \frac{1}{T}\sum_{t=1}^{T} q_t$. We have that:

$$\begin{aligned}
\min_p \max_q p^\dagger M q &\leq \max_q \overline{p}^\dagger M q \\
&\leq \max_q \frac{1}{T}\sum_{t=1}^{T} p_t^\dagger M q \\
&\leq \frac{1}{T}\sum_{t=1}^{T} \max_q p_t^\dagger M q \\
&= \frac{1}{T}\sum_{t=1}^{T} p_t^\dagger M q_t \\
&\leq \min_p \frac{1}{T}\sum_{t=1}^{T} p^\dagger M q_t + 2\sqrt{\frac{\log n}{T}} \\
&= \min_p p^\dagger M \overline{q} + 2\sqrt{\frac{\log n}{T}} \\
&\leq \max_q \min_p p^\dagger M q + 2\sqrt{\frac{\log n}{T}}
\end{aligned}$$

.

Now taking the limit as $T \to 0$, we have that:

$$\min_p \max_q p^\dagger M q \leq \max_q \min_p p^\dagger M q$$

which proves the other direction.

2 Weak and Strong Learning

Assume that we have a set of m examples $\{(x_1, y_1), (x_2, y_2), \ldots (x_m, y_m)\}$ where each $y \in \{-1, 1\}$. We would like to find a hypothesis h such that $h(x_t) = y_t$ for most training examples.

Ideally, we would like to find a hypothesis that is good on future examples, a point which we return to in a later lecture.

Let us now formalize the weak learning assumption. We say that we have a γ-weak learner if for every distribution w over the traning set, we can find a hypothesis $h : X \to [-1, 1]$ such that:

$$\sum_{i=1}^{m} w_i \frac{|h(x_i) - y_i|}{2} \leq \frac{1}{2} - \gamma$$

This L_1 error can be interpreted as the expected under under probabilistic predictions from h. Intuitively, we think of this as 'weak' learning when γ could be small — so the weak learner is only required to do slightly better than chance.

3 Boosting

For the case where γ is known, we now present the boosting algorithm. The algorithm enjoys the following performance guarantee:

Algorithm 1 Boosting

Input parameters: γ, T

Initialize $w_1 \leftarrow \frac{1}{T}\mathbf{1}$
for $t = 1$ to T **do**
Call γ-WeakLearner with distribution w_t, and receive hypothesis $h_t : X \to [-1, 1]$.
Set

$$l_{t,i} = 1 - \frac{|h_t(x_i) - y_i|}{2}$$

and update the weights

$$w_{t+1,i} = \frac{w_{t,i} e^{-\frac{\gamma}{2} l_{t,i}}}{Z} \quad , \quad Z = \sum_i w_{t,i} e^{-\frac{\gamma}{2} l_{t,i}}$$

end for
OUTPUT the 'majority vote' hypothesis:

$$h(x) = \text{sgn}\left(\frac{1}{T}\sum_{t=1}^{T} h_t(x)\right)$$

Theorem 3.1. *Let h be the output hypothesis of Boosting. Let M be the set of mistakes on the traning set, i.e. $M = \{i : \ h(x_i) \neq y_i\}$. We have:*

$$\frac{|M|}{m} \leq e^{-T\gamma^2/4}$$

Proof. We will appeal to the guarantee of our experts algorithm. For any w^*, we have that:

$$\sum_{t=1}^{T} w_t \cdot l_t \leq \sum_{t=1}^{T} w^* \cdot l_t + \frac{2KL(w^*||w_1)}{\gamma} + \frac{\gamma}{2} T$$

where we have used that $\eta = \gamma/2$ in boosting.

By the definition of weak learning, we have:

$$w_t \cdot l_t = 1 - \sum_i w_{t,i} \frac{|h_t(x_i) - y_i|}{2} \geq \frac{1}{2} + \gamma$$

for all t. So we have:

$$T(\frac{1}{2} + \gamma) \leq \sum_{t=1}^{T} w^* \cdot l_t + \frac{2KL(w^*||w_1)}{\gamma} + \frac{\gamma}{2} T$$

Rearranging:

$$\frac{T}{2} + \frac{T\gamma}{2} \leq \sum_{t=1}^{T} w^* \cdot l_t + \frac{2KL(w^*||w_1)}{\gamma}$$

which holds for all probilibity distributions w^*.

We will now choose w^* to be uniform over the set M. For $i \in M$, we know

$$\frac{|y_i - \frac{1}{T}\sum_{t=1}^{T} h_t(x_i)|}{2} \geq \frac{1}{2}$$

Hence, for $i \in M$

$$\frac{1}{T}\sum_{t=1}^{T} l_{t,i} = 1 - \frac{1}{T}\sum_{t=1}^{T} \frac{|y_i - h_t(x_i)|}{2} = 1 - \frac{|y_i - \frac{1}{T}\sum_{t=1}^{T} h_t(x_i)|}{2} \leq \frac{1}{2}$$

Hence,

$$\sum_{t=1}^{T} w^* \cdot l_t \leq \frac{T}{2}$$

Hence, we have that:

$$\frac{T}{2} + \frac{T\gamma}{2} \leq \frac{T}{2} + \frac{2\log(m/|M|)}{\gamma}$$

where we have used the definition of the KL distance with the uniform distribution. Rearranging completes the proof. $\square$