ECE 6980
Algorithmic and Information-Theoretic Methods in Data Science

Instructor: Jayadev Acharya — Lecture #6
Scribe: Qian Zhang — September 13, 2017

1 Review

- Reduction from learning to testing,
- Lower bound of $O(\frac{k}{\varepsilon^2})$ for learning distributions over $[k]$,
- Coding theoretic: $p_1, \cdots, p_{2^{ck}}, d_{\text{TV}}(p_i, p_j) > 3\varepsilon$ for any i, j,
- KL divergence adds up over product distribution,
- Proving lower bound on learning Gaussian

2 Le Cam's two point method

Given $\mathbb{P}$ and $\mathbb{Q}$ as two distribution collections, the goal is to prove lower bounds on testing $\mathbb{P}$ and $\mathbb{Q}$. Uniformity testing is a special case when $\mathbb{P} = \{U\}$, $\mathbb{Q} = \{q \mid d_{\text{TV}}(q, U) \geq \varepsilon\}$ where U is uniform distribution over support $[k]$.

An easy attempt is to take one distribution $p \in \mathbb{P}$ and $q \in \mathbb{Q}$. Intuitively, lower bound on testing p and q should be a lower bound for testing $\mathbb{P}$ and $\mathbb{Q}$. With p, q and $d_{\text{TV}}(p, q) = \epsilon$, we have

$$\Omega(\frac{1}{\varepsilon}) < n_{\text{test}} < O(\frac{1}{\varepsilon^2})$$

The easy attempt above may run into trouble in complexity when the dimension of the problem increases.

A more sophisticated attempt:

Consider *any* distribution (priors) μ over $\mathbb{P}$ and ν over $\mathbb{Q}$, with probability $\frac{1}{2}$, pick $\mathbb{P}(u)$ or $\mathbb{Q}(u)$.

For $\mathbb{P}$, we will pick $p \in \mathbb{P}$ according to μ. Then, we generate n samples from p. Similarly, for $\mathbb{Q}$, we will pick $q \in \mathbb{Q}$ according to ν. Then, we generate n samples from q.

Given $X_1, \cdots, X_n$ i.i.d, we have

$$\mathbb{P}(X_1, \cdots, X_n) = \sum_{p \in \mathbb{P}} p(X_1, \cdots, X_n)\mu(p) = \mu\mathbb{P}^{\otimes n}$$

Similarly,

$$\mathbb{Q}(X_1, \cdots, X_n) = \sum_{q \in \mathbb{Q}} q(X_1, \cdots, X_n)\nu(q) = \nu\mathbb{Q}^{\otimes n}$$

Now, we formally define *Le Cam's two points method*:

Definition 1. For any u,v, the number of samples needed to test between $\mu\mathbb{P}^{\otimes n}$ and $\nu\mathbb{Q}^{\otimes n}$ is a lower bound on the number of samples to test $\mathbb{P}, \mathbb{Q}$.

3 Example in uniform testing

Given $\mathbb{P} = \{U\}$, $\mathbb{Q} = \{q : d_{\text{TV}}(q, U) \geq \varepsilon\}$. We can set $\mu(U) = 1$. However for $q(i)$, we have following two constructions:

We can set a distribution with $q(i)$ for $i = 1, \cdots, k$ such that

$$q(i) = \{\frac{1+2\varepsilon}{k} \quad \text{or} \quad \frac{1-2\varepsilon}{k}\}$$

We can also set $\bar{Z} = Z_1, Z_2, \cdots, Z_{\frac{k}{2}} \in \{\pm 1\}^{\frac{k}{2}}$ such that

$$q_{\bar{Z}}(2i-1) = \frac{1+2\varepsilon Z_i}{k}, \quad q_{\bar{Z}}(2i) = \frac{1-2\varepsilon Z_i}{k}$$

Based on the second construction of q, we have $\mu \to \{U\}$, $\nu \to \{2^{\frac{k}{2}}$ distributions$\}$. Unless $n > \sqrt{k}/\varepsilon^2$ with $(\varepsilon < 0.2)$, $d_{\text{TV}}(U^{\otimes n}, \nu q_{\bar{Z}}^{\otimes n})$ is small.

Now, we will introduce χ^2-distance since d_{TV} does not have nice property with product distribution. Given p, q over $\mathcal{X}$, χ^2-distance is

$$\chi^2(p, q) := \sum_{x \in \mathcal{X}} \frac{(p(x) - q(x))^2}{q(x)} = (\sum_x \frac{p(x)^2}{q(x)}) - 1$$