Lecture 24: Robust Mean Estimation in High Dimensional

Daniel Kane
Scribe: Yang Liu

May 26, 2017

Get samples from $N(\mu, I)$, ϵ-fraction have arbitrary errors, can we learn μ efficiently to small error, i.e $O(\epsilon)$? We want to return sample mean but it may not work if there is a sufficient bad error that corrupts the data to much.
Idea: remove outliers
Suppose $\mu_0 \approx \mu$, define outliers as those far from μ_0. How far from μ_0 do we expect, on average, it should be

$$|\mu - x|_2 \approx \sqrt{n}$$

we can throw out errors $\gg \sqrt{n}$ from μ but this doesn't solve the problem because ϵ-fraction corruption can corrupt mean by $\epsilon\sqrt{n}$
Another Idea: If $|\mu - \hat{\mu}|_2 > \delta$ then $\exists\ |v_2| = 1$ s.t. $v(\hat{\mu} - \mu) > \delta$. If we know v then we can detect outliers

$$\begin{aligned} var(v \cdot x) &> 1 + \epsilon\Big(\frac{\delta}{\epsilon}\Big)^2 \\ &= 1 + \frac{\delta^2}{\epsilon} \end{aligned}$$

Since $var(v \cdot x) = v^T cov(x) v$, consider eigenvalues of $cov(x)$,there are two cases:

- All eigenvalues $< 1 + \frac{\delta^2}{\epsilon}$, this implies there is no such v and $|\mu - \hat{\mu}|_2 \leq \delta$
- If there is some eigenvector v with eigenvalue $<1 + \frac{\delta^2}{\epsilon}$, then we have

  $$\begin{aligned} E[(vx - \mu_v)^2] &> var(v \cdot x) \\ &> 1 + \frac{\delta^2}{\epsilon} \end{aligned}$$

  where μ_v denotes the mean of vx, with good samples contribute 1 and bad samples contribute $\frac{\delta^2}{\epsilon}$ This implies that decent fraction of bad samples have $|vx - \mu_v| > \frac{\delta}{\epsilon}$ and very few good samples have such attribute. Then we can create a filter to throw out samples which satisfy $|vx - \mu_v| > \frac{\delta}{\epsilon}$

Definiton: A set S of points is good w.r.t some Gaussian G if

- $cov(S) = I \pm \epsilon(operator\_norm)$

- $\int_{lg(\frac{1}{\epsilon})}^{\infty} Pr_{x \in S}(v(x-\mu) > \sqrt{t}) dt = O(\epsilon lg(\frac{1}{\epsilon}))$ for any v with $|v|_2 = 1$
- $mean(S) \approx \mu$

Given $S^{'}$ with $\Delta(S, S^{'}) < \epsilon$, if S is good, an algorithm given any $S^{'}$ with $\Delta(S, S^{'}) < \epsilon$ can return a δ-approximation to μ

Proposal: Design an algorithm given $S^{'}$ returns either a δ-approximation to μ or $S^{''}$ with $\Delta(S^{''}, S^{'}) < \Delta(S, S^{'})$ where $\Delta(S, S^{'})$ is defined as:

$$\Delta(S, S^{'}) = \frac{|S \ \Delta S^{'}|}{|S|}$$

Let $S^{'} = S - S_L + S_E$, $\frac{|S_L|}{|S|} = \epsilon_L$, $\frac{|S_E|}{|S|} = \epsilon_E$, where S represents a good set, S_L represents elements removed from the good set and S_E represents new elements added by the adversary.

$$\begin{aligned} cov(S^{'}) &= E[X^T X] - E[X]^T E[X] \\ &= \frac{Cov(S) - \epsilon_L(cov(S_L)) + \epsilon_E(cov(S_E))}{1 - \epsilon_L + \epsilon_E} + cov(means) \end{aligned}$$

- $cov(S) = I \pm \epsilon$

- $\epsilon_L(cov(S_L)$ is small in operator norm
  Proof: We need to show $\epsilon_L(var(vS_L))$ is small,

$$\begin{aligned} \epsilon_L(cov(S_L)) &\leq \epsilon_L(var(vS_L)) \\ &\leq \epsilon_L E[(v(S_L - \mu)^2)] \\ &\leq \int_0^{\infty} \epsilon_L \ [fraction \ of \ elements \ in \ S_L \ with \ (v(x-\mu))^2 > t] \ dt \\ &\leq \int_0^{\infty} min(\epsilon_L, Pr_{x \in S}(|v(x-\mu)| > \sqrt{n})) dt \\ &\leq \epsilon_L log(\frac{1}{\epsilon}) \end{aligned}$$

- $\epsilon_E(cov(S_E))$ can be large but is positive semi-definite.
- cov(means) can be quite large if the μ we have is far from true μ.

Algorithm 1 Filter algorithm for a sub-Gaussian with unknown mean and identity covariance

compute $cov(S^{'})$
if there is no large eigenvalues
return $\hat{\mu}$
otherwise produce a filter which returns a set $S^{''}$ with $\Delta(S^{''}, S^{'}) < \Delta(S, S^{'})$

This is just a high level idea, a lot of technical details are omitted here. For more rigorous proof and more details, please refer to section 8 of paper http://cseweb.ucsd.edu/~dakane/robustLearnHighDim.pdf