Lecture 5: Random Projection

Instructor: Santosh Vempala *Lecture date: 9/20*

1 Random projection

The sample complexity for (ϵ,δ)-PAC learning a halfspace in $\mathbb{R}^n$ is $O(\frac{1}{\epsilon}(n\log\frac{1}{\epsilon}+\log\frac{1}{\delta}))$. For halfspaces that lie in high dimension with $\gamma-$margin, could we have a better complexity? For example, if we can project the data into $\mathbb{R}^k, k \ll n$, while preserving the linear separability of the data, then the sample complexity will be $O(\frac{1}{\epsilon}(k\log\frac{1}{\epsilon}+\log\frac{1}{\delta}))$.

As the initial step, consider the case when we randomly project the data into one dimension. Let $w \in \mathbb{R}^n$ be the normal vector of the subspace that separates the data. To preserve the separability, the candidate vectors lie in a cone around w. If the projection vector is randomly drawn uniformly from the unit ball in $\mathbb{R}^n$, the probability that the projection preserves separability is $O(\gamma^n)$, which is exponentially small. Thus it is not sufficient to project to one dimension. In the following section, we will figure out the dimension of the subspace that we can project the data to while preserving separability with high probability.

Random Projection Matrix Consider a random matrix $R \in \mathbb{R}^{n\times k}$, where $R_{ij} \sim N(0,1/k)$. For a vector $x \in \mathbb{R}^n$, it is projected to the vector $\tilde{x} = R^\top x \in \mathbb{R}^k$. We will next show that the random projection by R can approximately preserve the norm and inner product.

Lemma 1. *For $\tilde{x} = R^\top x, R_{ij} \sim N(0,1/k)$, we have*

$$\mathbb{E}\|\tilde{x}\|_2^2 = \|x\|_2^2$$

Proof.

$$\mathbb{E}\|\tilde{x}\|_2^2 = \mathbb{E}x^\top RR^\top x = x^\top(\mathbb{E}RR^\top)x$$

For $1 \le i,j \le n$, we have

$$\mathbb{E}(RR^\top)_{ij} = \sum_{l=1}^{k}\mathbb{E}R_{il}R_{jl} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Thus $\mathbb{E}RR^\top$ is the identity matrix. So we have

$$\mathbb{E}\|\tilde{x}\|_2^2 = x^\top x = \|x\|_2^2$$

□

Theorem 2. *For $\tilde{x} = R^\top x, R_{ij} \sim N(0,1/k)$, we have*

$$\mathbb{P}\left(|\|\tilde{x}\|_2^2 - \|x\|_2^2| > \epsilon\|x\|_2^2\right) \le 2e^{-(\epsilon^2-\epsilon^3)k/4}$$

Corollary 3. *Consider doing random projection for m samples $x^{(1)},\cdots,x^{(m)}$ with $\tilde{x}^{(i)} = R^\top x^{(i)}$, where $R \in \mathbb{R}^{n\times k}, R_{ij} \sim N(0,1/k)$. Let $0 < \epsilon < 1/2, 0 < \delta < 1$. Then for $k \ge \frac{8}{\epsilon^2}\log\frac{2m}{\delta}$, with probability $\ge 1-\delta$,*

$$\forall i \in [m], (1-\epsilon)\|x^{(i)}\|_2^2 \le \|\tilde{x}^{(i)}\|_2^2 \le (1+\epsilon)\|x(i)\|_2^2.$$

Proof. For m samples $x^{(1)},\cdots,x^{(m)}$, by Theorem 2,

$$\mathbb{P}(\exists i \in [m], |\|\tilde{x}^{(i)}\|_2^2 - \|x^{(i)}\|_2^2 > \epsilon\|x^{(i)}\|_2^2|) \le 2me^{-(\epsilon^2-\epsilon^3)k/4}$$

In other words, with probability $\geq 1 - 2me^{-(\epsilon^2-\epsilon^3)k/4}$, all samples can preserve their norm within ϵ. That is,

$$\mathbb{P}(\forall i \in [m], (1-\epsilon)\|x^{(i)}\|_2^2 \leq \|\tilde{x}^{(i)}\|_2^2 \leq (1+\epsilon)\|x(i)\|_2^2) \geq 1 - 2me^{-(\epsilon^2-\epsilon^3)k/4}$$

To let this probability to be $\geq 1-\delta$, we need

$$2me^{-(\epsilon^2-\epsilon^3)k/4} \leq \delta$$

This implies

$$k \geq \frac{4\log\frac{2m}{\delta}}{\epsilon^2(1-\epsilon)}$$

For $\epsilon < 1/2$,

$$\frac{4\log\frac{2m}{\delta}}{\epsilon^2(1-\epsilon)} < \frac{8\log\frac{2m}{\delta}}{\epsilon^2}$$

So it suffices to let

$$k \geq \frac{8\log\frac{2m}{\delta}}{\epsilon^2}$$

□

To preserve linear separability, we need also to preserve the inner product within ϵ. We give the result in the following lemma.

Lemma 4. *For $u, v \in \mathbb{R}^n$ with $\|u\|_2, \|v\|_2 \leq 1$, we use the same random projection and get $\tilde{u}, \tilde{v}$. Then*

$$\mathbb{P}\left(|\langle u, v\rangle - \langle \tilde{u}, \tilde{v}\rangle| \leq \epsilon\right) \geq 1 - 4e^{-(\epsilon^2-\epsilon^3)k/4}$$

Proof. By Theorem 2, with probability $\geq 1 - 4e^{-(\epsilon^2-\epsilon^3)k/4}$,

$$\begin{aligned}
4\langle \tilde{u}, \tilde{v}\rangle =& \|\tilde{u}+\tilde{v}\|_2^2 - \|\tilde{u}-\tilde{v}\|_2^2 \\
\leq& \|u+v\|^2(1+\epsilon) - \|u-v\|_2^2(1-\epsilon) \\
=& 4\langle u, v\rangle + \epsilon(\|u+v\|_2^2 + \|u-v\|_2^2) \\
=& 4\langle u, v\rangle + 2\epsilon(\|u\|_2^2 + \|v\|_2^2) \\
\leq& 4\langle u, v\rangle + 4\epsilon
\end{aligned}$$

This implies

$$\langle \tilde{u}, \tilde{v}\rangle \leq \langle u, v\rangle + \epsilon$$

We can prove the second part similarly. □

Corollary 5. *To learn a halfspace in $\mathbb{R}^n$ with margin γ, we can project the data into a $k-$dimensional subspace with $k = O(\frac{1}{\gamma^2}\log\frac{1}{\gamma^2\epsilon\delta})$. Then we need only $m = O(\frac{1}{\epsilon\gamma^2}\log\frac{1}{\gamma^2\epsilon\delta}\log\frac{1}{\epsilon} + \frac{1}{\epsilon}\log\frac{1}{\delta}) = \tilde{O}(\frac{1}{\epsilon\gamma^2})$ samples to $(\epsilon, \delta)-$PAC learn the halfspaces.*

Proof. We first randomly project the data into $\mathbb{R}^k$ with $k = O(\frac{1}{\gamma^2}\log\frac{m}{\delta})$. We get a margin of at least $\gamma/2$ with high probability. To learn the halfspace in $\mathbb{R}^k$, we need $m = O(\frac{k}{\epsilon}\log\frac{1}{\epsilon} + \frac{1}{\epsilon}\log\frac{1}{\delta})$ samples. Therefore we have

$$k = O(\frac{1}{\gamma^2}\log\frac{1}{\gamma^2\epsilon\delta}), \quad m = O(\frac{1}{\epsilon\gamma^2}\log\frac{1}{\gamma^2\epsilon\delta}\log\frac{1}{\epsilon} + \frac{1}{\epsilon}\log\frac{1}{\delta})$$

□

Extension: sparse random projection / using random signs $R \in \{\pm 1\}^{n\times k}$.