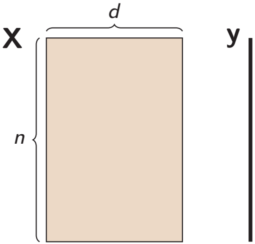
Leveraged volume sampling for linear regression

Michał Dereziński — UC Berkeley
Manfred K. Warmuth — UC Santa Cruz
Daniel Hsu — Columbia University

Linear regression



Loss: $$L(\mathbf{w}) = \sum_i (\mathbf{x}_i^\top \mathbf{w} - y_i)^2$$

Optimum: $$\mathbf{w}^* = \operatorname*{argmin}_{\mathbf{w}} L(\mathbf{w})$$

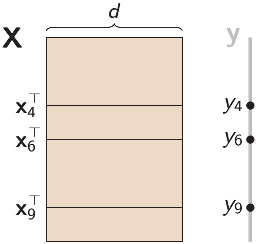
Leveraged volume sampling for linear regression

Michał Dereziński — UC Berkeley
Manfred K. Warmuth — UC Santa Cruz
Daniel Hsu — Columbia University

Linear regression with hidden responses

Sample
$S = \{4, 6, 9\}$

Receive
y_4, y_6, y_9



Loss: $$L(\mathbf{w}) = \sum_i (\mathbf{x}_i^\top \mathbf{w} - y_i)^2$$

Optimum: $$\mathbf{w}^* = \operatorname*{argmin}_{\mathbf{w}} L(\mathbf{w})$$

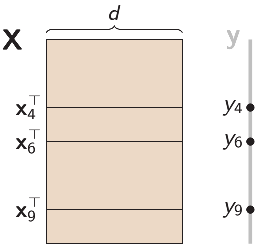
Leveraged volume sampling for linear regression

Michał Dereziński — UC Berkeley
Manfred K. Warmuth — UC Santa Cruz
Daniel Hsu — Columbia University

Linear regression with hidden responses

Sample
$S = \{4, 6, 9\}$

Receive
y_4, y_6, y_9



Goal: Best unbiased estimator $\widehat{\mathbf{w}}(S)$

$$\mathbb{E}\big[\widehat{\mathbf{w}}(S)\big] = \mathbf{w}^*$$

$$L\big(\widehat{\mathbf{w}}(S)\big) \le (1+\epsilon)\, L(\mathbf{w}^*)$$

Loss: $L(\mathbf{w}) = \sum_i (\mathbf{x}_i^\top \mathbf{w} - y_i)^2$

Optimum: $\mathbf{w}^* = \operatorname*{argmin}_{\mathbf{w}} L(\mathbf{w})$

Existing sampling methods:

1. leverage score sampling: i.i.d., biased
2. volume sampling: joint, unbiased

Leveraged volume sampling

Volume sampling
Jointly choose set S of $k \geq d$ indices s.t.

$$\Pr(S) \;\propto\; \det\Big(\sum_{i \in S} \mathbf{x}_i \mathbf{x}_i^\top\Big)$$

Leveraged volume sampling

Volume sampling
Jointly choose set S of $k \geq d$ indices s.t.

$$\Pr(S) \;\propto\; \det\Big(\sum_{i \in S} \mathbf{x}_i \mathbf{x}_i^\top\Big)$$

Theorem [DW17]

$$\mathbb{E}\big[\widehat{\mathbf{w}}(S)\big] = \mathbf{w}^*$$

where $\widehat{\mathbf{w}}(S) = \operatorname*{argmin}_{\mathbf{w}} \sum_{i \in S} (\mathbf{x}_i^\top \mathbf{w} - y_i)^2.$

Leveraged volume sampling

Volume sampling
Jointly choose set S of $k \geq d$ indices s.t.

$$\Pr(S) \;\propto\; \det\Big(\sum_{i\in S} \mathbf{x}_i \mathbf{x}_i^\top\Big)$$

Theorem [DW17]

$$\mathbb{E}\big[\widehat{\mathbf{w}}(S)\big] = \mathbf{w}^*$$

New Lower Bound
Volume sampling may need a sample of size $k = \Omega(n)$ to get a $\underbrace{(3/2)}_{\epsilon=1/2}$-approximation

Leveraged volume sampling

Volume sampling
Jointly choose set S of $k \geq d$ indices s.t.

$$\Pr(S) \propto \det\Big(\sum_{i\in S} \mathbf{x}_i \mathbf{x}_i^\top\Big)$$

Solution: Use i.i.d. and joint sampling

$$\Pr(S) \propto \overbrace{\Big(\prod_{i\in S} \ell_i\Big)}^{\text{leverage scores}} \overbrace{\det\Big(\sum_{i\in S} \frac{1}{\ell_i}\mathbf{x}_i \mathbf{x}_i^\top\Big)}^{\text{volume sampling}}$$

Theorem [DW17]

$$\mathbb{E}\big[\widehat{\mathbf{w}}(S)\big] = \mathbf{w}^*$$

New Lower Bound
Volume sampling may need a sample of size $k = \Omega(n)$ to get a $\underbrace{(3/2)}_{\epsilon=1/2}$-approximation

Leveraged volume sampling

Volume sampling

Jointly choose set S of $k \geq d$ indices s.t.

$$\Pr(S) \;\propto\; \det\Big(\sum_{i \in S} \mathbf{x}_i \mathbf{x}_i^\top\Big)$$

Solution: Use i.i.d. and joint sampling

$$\Pr(S) \;\propto\; \underbrace{\Big(\prod_{i \in S} \ell_i\Big)}_{\text{leverage scores}} \; \underbrace{\det\Big(\sum_{i \in S} \frac{1}{\ell_i} \mathbf{x}_i \mathbf{x}_i^\top\Big)}_{\text{volume sampling}}$$

Theorem [DW17]

$$\mathbb{E}\big[\widehat{\mathbf{w}}(S)\big] = \mathbf{w}^*$$

$$\widehat{\mathbf{w}}(S) \;=\; \operatorname*{argmin}_{\mathbf{w}} \sum_{i \in S} \frac{1}{\ell_i} (\mathbf{x}_i^\top \mathbf{w} - y_i)^2$$

New Lower Bound

Volume sampling may need a sample of size $k = \Omega(n)$ to get a $\underbrace{(3/2)}_{\epsilon = 1/2}$-approximation

New Theorem For $k = O(d \log d + d/\epsilon)$

$$\mathbb{E}\big[\widehat{\mathbf{w}}(S)\big] = \mathbf{w}^* \quad \text{and}$$

$$\text{w.h.p.} \quad L\big(\widehat{\mathbf{w}}(S)\big) \leq (1 + \epsilon) L(\mathbf{w}^*)$$

New volume sampling algorithm

Determinantal rejection sampling trick

---

repeat

  Sample $i_1, \dots, i_s$ i.i.d. $\sim (\ell_1, \dots, \ell_n)$

  Sample *Accept* $\sim$ Bernoulli $\left[\frac{\det\left(\sum_{t=1}^{s} \frac{1}{\ell_{i_t}} \mathbf{x}_{i_t} \mathbf{x}_{i_t}^\top\right)}{\det(\mathbf{X}^\top \mathbf{X})}\right]$

until *Accept* $=$ true

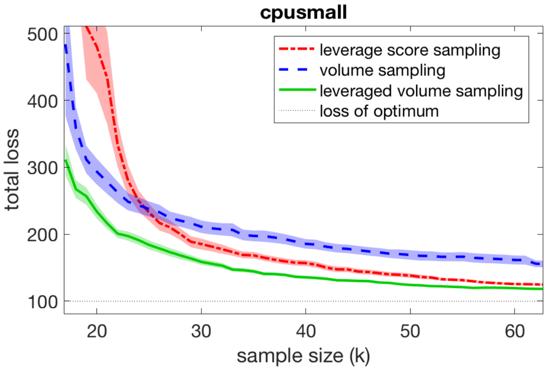
$$\underbrace{\text{preprocessing } O(nd^2)}_{\text{improvable to } \widetilde{O}(nd+\text{poly}(d))} + \underbrace{\text{sampling } O(d^4)}_{\text{no dependence on } n}$$

New volume sampling algorithm

Determinantal rejection sampling trick
repeat
Sample $i_1, \ldots, i_s$ i.i.d. $\sim (\ell_1, \ldots, \ell_n)$
Sample *Accept* $\sim$ Bernoulli$\left[\frac{\det\left(\sum_{t=1}^{s} \frac{1}{\ell_{i_t}} \mathbf{x}_{i_t}\mathbf{x}_{i_t}^\top\right)}{\det(\mathbf{X}^\top\mathbf{X})}\right]$
until *Accept* = true

$$\underbrace{\text{preprocessing } O(nd^2)}_{\text{improvable to } \widetilde{O}(nd+\text{poly}(d))} + \underbrace{\text{sampling } O(d^4)}_{\text{no dependence on } n}$$

Experiments – 7 datasets from Libsvm



Check out poster **#151**