Multiplicative Weights & Differential Privacy

CompSci 590.03
Instructor: Ashwin Machanavajjhala
This Class

• Online Learning and Multiplicative Weights

• MWEM: Differentially private method to releasing data
Multiple Experts Problem

Will it rain today?

Yes
Yes
Yes
No

What is the best prediction based on these experts?
Multiple Experts Problem

• Suppose we know the best expert (who makes the least error), then we can just return what that expert says.
  – This is the best we can hope for.

• We don’t know who the best expert is.
  – But we can learn ... we know whether it rained or not at the end of the day.

Qn: Is there an algorithm that learns over time who the best expert is, and has an accuracy that close to the best expert?
Weighted Majority Algorithm

“Experts”

$W_1$ $W_2$ $W_3$ $W_4$

$Y_1$ $Y_2$ $Y_3$ $Y_4$

Algorithm

\[ \sum_i w_i y_i \]

\[ \sum_i w_i \]

\[ \frac{\sum_i w_i y_i}{\sum_i w_i} > \frac{1}{2} \]
Weighted Majority Algorithm

[Littlestone&Warmuth ‘94]

Yes Yes Yes No

Yes

Yes

Yes

No

1 - \(\varepsilon\)

1 - \(\varepsilon\)

1 - \(\varepsilon\)

1

\[\sum_i w_i y_i > \frac{1}{2}\]

Yes!
Multiplicative Weights Algorithm

- Maintain weights (or probability distribution) over experts.

Answering/Prediction:
- Answer using weighted majority, OR
- Randomly pick an expert based on current probability distribution. Use random experts answer.

Update:
- Observe truth.
- Decrease weight (or probability) assigned to the experts who are wrong.
Theorem:
After $t$ steps,
- let $m(t,j)$ be the number of errors made by expert $j$
- let $m(t)$ be the number of errors made by algorithm
- let $n$ be the number of experts,

$$\forall j, \quad m(t) \leq \frac{2 \ln n}{\varepsilon} + 2(1 + \varepsilon)m(t, j)$$
Error Analysis: Proof

• Let $\varphi(t) = \Sigma w_i$. Then, $\varphi(1) = n$.

• When the algorithm makes a mistake,
  $\varphi(t+1) \leq \varphi(t) \left( \frac{1}{2} + \frac{1}{2}(1-\varepsilon) \right) = \varphi(t)(1-\varepsilon/2)$

• When the algorithm is correct,
  $\varphi(t+1) \leq \varphi(t)$

• Therefore,
  $\varphi(t) \leq n(1-\varepsilon/2)^{m(t)}$
Error Analysis: Proof

• $\varphi(t) \leq n(1-\varepsilon/2)^{m(t)}$

• Also, $W_j(t) = (1-\varepsilon)^{m(t_j)}$

• $\varphi(t) \geq W_j(t) \implies n(1-\varepsilon/2)^{m(t)} \geq (1-\varepsilon)^{m(t_j)}$

• Hence, $m(t) \geq 2/\varepsilon \ln n + 2(1+\varepsilon)m(t_j)$
Multiplicative Weights

[Arora, Hazan, Kale ‘05]

• This algorithm technique has been used to solve a number of problems
  – Packing and covering Linear programs (Plotkin-Shmoys-Tardos)
  – Log n approximation for many NP-hard problems (set cover ...)
  – Boosting
  – Zero sum games
  – Network congestion
  – Semidefinite programs
This Class

• Online Learning and Multiplicative Weights

• MWEM: Differentially private method to releasing data
Workload-aware Synthetic Data Generation

Input:

- Q, a workload of (expected/typical) linear queries of the form $\Sigma_x q(x)$, and each $q(x)$ is in the range $[-1,1]$
- D, a database instance
- T, number of iterations
- $\epsilon$, differential privacy parameter

Output:

- A, a synthetically generated dataset such that for all $q$ in $Q$, $q(A)$ is close to $q(D)$
Multiplicative Weights Algorithm

• Let $n$ be the number of records in $D$, and $N$ be the number of values in the domain.

Initialization

• Let $A_0$ be a weight function that assigns $n/N$ weight to each value in the domain.
Multiplicative Weights

• Let \( n \) be the number of records in \( D \), and \( N \) be the number of values in the domain.
• Let \( A_0 \) be a weight function that assigns \( n/N \) weight to each value in the domain.

In iteration \( j \) in \( \{1, 2, \ldots, T\} \),
• Pick query \( q \) from \( Q \) with maximum error
  – Error = \( q(D) - q(A_{i-1}) \)
Multiplicative Weights

- Let $n$ be the number of records in $D$, and $N$ be the number of values in the domain.
- Let $A_0$ be a weight function that assigns $n/N$ weight to each value in the domain.

In iteration $j$ in $\{1, 2, \ldots, T\}$,

- Pick query $q$ from $Q$ with maximum error
- Compute $m = q(D)$
Multiplicative Weights

- Let $n$ be the number of records in $D$, and $N$ be the number of values in the domain.
- Let $A_0$ be a weight function that assigns $n/N$ weight to each value in the domain.

In iteration $j$ in $\{1, 2, ..., T\}$,
- Pick query $q$ from $Q$ with maximum error
- Compute $m = q(D)$
- Update Weights
  - $A_i(x) \propto A_{i-1}(x) \cdot \exp \left( q(x) \cdot \frac{(m - q(A_{i-1}))}{2n} \right)$

Output: $\text{average}_i(A_i)$
Update rule

\[ A_i(x) \propto A_{i-1}(x) \cdot \exp( q(x) \cdot (m - q(A_{i-1}))/2n ) \]

If \( q(D) - q(A) > 0 \),
then increase the weight of records with \( q(x) > 0 \), and decrease the weight of records with \( q(x) < 0 \)

If \( q(D) - q(A) < 0 \),
then decrease the weight of records with \( q(x) > 0 \), and increase the weight of records with \( q(x) < 0 \)
Error Analysis

Theorem:
For any database $D$, and any set of linear queries $Q$, MWEM outputs an $A$ such that:

$$\max_{q \in Q} |q(A) - q(D)| \leq 2n \sqrt{\frac{\log |\text{dom}|}{T}}$$
Error Analysis: Proof

\[
\max_{q \in Q} |q(A) - q(D)| = \max_{q \in Q} \left| q\left( \text{avg} A_i \right) - q(D) \right| \\
\leq \text{avg} \max_{i, q \in Q} |q(A_i) - q(D)| \leq \text{avg} \max_{i} \text{maxerr}_i
\]

Consider the potential function: \( \varphi(i) = \sum_{x \in \text{dom}} \frac{D(x)}{n} \log \left( \frac{D(x)}{A_i(x)} \right) \)

\( \varphi(i) \geq 0, \text{ and } \varphi(i) \leq \log |\text{dom}| \)
Error Analysis: Proof

Claim: \( \forall i, \phi(i - 1) - \phi(i) \geq \left( \frac{q(D) - q(A_i)}{2n} \right)^2 \)

\[
\text{avg maxerr}_i = \frac{1}{T} \sum_{i=1}^{T} |q(D) - q(A_i)|
\]

\[
\leq \frac{1}{T} \sum_{i=1}^{T} \left( 2n \sqrt{\phi(i - 1) - \phi(i)} \right)
\]

\[
\leq 2n \sqrt{\frac{1}{T} \sum_{i=1}^{T} (\phi(i - 1) - \phi(i))}
\]

\[
= 2n \sqrt{\frac{\phi(0) - \phi(T)}{T}} = 2n \sqrt{\frac{\log|\text{dom}|}{T}}
\]
Synthetic Data Generation with Privacy

Input:
- $Q$, a workload of (expected/typical) linear queries of the form $\Sigma_x q(x)$, and each $q(x)$ is in the range $[-1,1]$
- $D$, a database instance
- $T$, number of iterations
- $\varepsilon$, differential privacy parameter

Output:
- $A$, a synthetically generated dataset such that for all $q$ in $Q$, $q(A)$ is close to $q(D)$
MWEM

• Let $n$ be the number of records in $D$, and $N$ be the number of values in the domain.

Initialization
• Let $A_0$ be a weight function that assigns $n/N$ weight to each value in the domain.

[Hardt, Ligett & McSherry‘12]
MWEM

Let $n$ be the number of records in $D$, and $N$ be the number of values in the domain.

Let $A_0$ be a weight function that assigns $n/N$ weight to each value in the domain.

In iteration $j$ in $\{1, 2, \ldots, T\}$,

- Pick query $q$ from $Q$ with max error using Exponential Mechanism
  - Parameter: $\varepsilon/2T$
  - Score function: $|q(A_{i-1}) - q(D)|$

More likely to pick those queries for which the answer on the synthetic data is very different from the answer on the true data.

[Hardt, Ligett & McSherry‘12]
MWEM

[Hardt, Ligett & McSherry’12]

- Let \( n \) be the number of records in \( D \), and \( N \) be the number of values in the domain.
- Let \( A_0 \) be a weight function that assigns \( n/N \) weight to each value in the domain.

In iteration \( j \) in \( \{1,2,...,T\} \),

- Pick query \( q \) from \( Q \) with max error using Exponential Mechanism
- Compute \( m = q(D) \) using Laplace Mechanism
  - Parameter: \( \varepsilon/2T \)
  - \( m = q(D) + \text{Lap}(2T/\varepsilon) \)
MWEM

Let $n$ be the number of records in $D$, and $N$ be the number of values in the domain.

Let $A_0$ be a weight function that assigns $n/N$ weight to each value in the domain.

In iteration $j$ in $\{1,2,\ldots, T\}$,

- Pick query $q$ from $Q$ with max error using Exponential Mechanism
- Compute $m = q(D)$ using Laplace Mechanism
- Update Weights
  
  \[ A_i(x) \propto A_{i-1}(x) \cdot \exp\left( q(x) \cdot \frac{m - q(A_{i-1})}{2n} \right) \]

Output: $\text{average}_i(A_i)$
Update rule

\[ A_i(x) \propto A_{i-1}(x) \cdot \exp(q(x) \cdot (m - q(A_{i-1}))/2n) \]

If noisy \( q(D) - q(A) > 0 \),
then increase the weight of records with \( q(x) > 0 \), and decrease the weight of records with \( q(x) < 0 \).

If noisy \( q(D) - q(A) < 0 \),
then decrease the weight of records with \( q(x) > 0 \), and increase the weight of records with \( q(x) < 0 \).
Error Analysis

Theorem:
For any database $D$, and any set of linear queries $Q$, with probability at least $1 - 2T/|Q|$, MWEM outputs an $A$ such that:

$$\max_{q \in Q} |q(A) - q(D)| \leq 2n \sqrt{\frac{\log |\text{dom}|}{T}} + \frac{10T \log |Q|}{\varepsilon}$$
Error Analysis: Proof

\[ \max_{q \in Q} |q(A) - q(D)| = \max_{q \in Q} \left| q\left( \text{avg} A_i \right) - q(D) \right| \]
\[ \leq \text{avg} \max_{i, q \in Q} |q(A_i) - q(D)| \leq \text{avg maxerr}_i \]

1. But exponential mechanism picks \( q_i \), which might not have the maximum error!

\[ P(|q_i(A_i) - q_i(D)| < \text{maxerr}_i - r) < |Q| \cdot e^{-\frac{\epsilon r}{4T}} \]

When \( r = \frac{8T \log|Q|}{\epsilon} \), we get w.p. \( 1 - \frac{1}{|Q|} \)

\[ \text{maxerr}_i \leq |q_i(A_i) - q_i(D)| + \frac{8T \log|Q|}{\epsilon} \]
Error Analysis: Proof

\[
\max_{q \in Q} |q(A) - q(D)| = \max_{q \in Q} \left| q \left( \frac{\text{avg} A_i}{i} \right) - q(D) \right| \\
\leq \text{avg} \max_{q \in Q} |q(A_i) - q(D)| \leq \text{avg} \max_{q \in Q} \text{err}_i
\]

1. In each iteration with probability at least \( 1 - 1/|Q| \), error in the query picked by exponential mechanism is smaller than max error by at most

\[
8T \frac{\log |Q|}{\varepsilon}
\]

2. **We add noise to \( m = q(D) \).** But with probability at least \( 1 - 1/|Q| \) in each iteration, the noise added by Laplace is at most

\[
2T \frac{\log |Q|}{\varepsilon}
\]
Error Analysis

Theorem:
For any database $D$, and any set of linear queries $Q$, with probability at least $1-\frac{2T}{|Q|}$, MWEM outputs an $A$ such that:

$$\max_{q \in Q} |q(A) - q(D)| \leq 2n \sqrt{\frac{\log |\text{dom}|}{T}} + \frac{10T \log |Q|}{\varepsilon}$$
Optimizations

• Output $A_T$ rather than the average
• In update step, use queries picked in all previous rounds for which $(m-q(A))$ is large.
• Can improve the solution by initializing $A_0$ with noisy counts.
Next Class

• The sparse vector technique and online query answering
References

Xu et al, “Differentially Private Histogram Publication”, ICDE 2012
Hardt & Rothblum, “A multiplicative weights mechanism for privacy-preserving data analysis”, FOCS ’10