Local DP have a high accuracy but
Also, simple anonimization can be broken → Not a good plan

Global Differential Privacy

Add noise after the data has been aggregated by a function

⇒ Similar protection with less accuracy but
⇒ Compared to Local Diff Pri
⇒ Requires trust on the third party aggregator

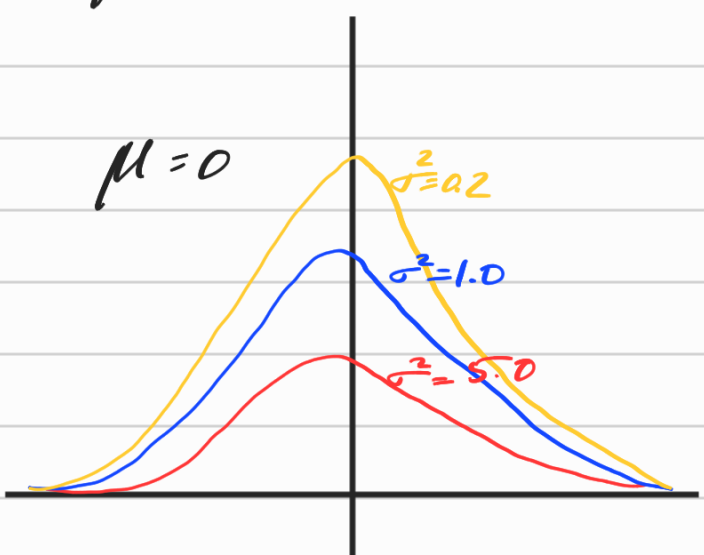
Eg A sum query with Global DP is as follows

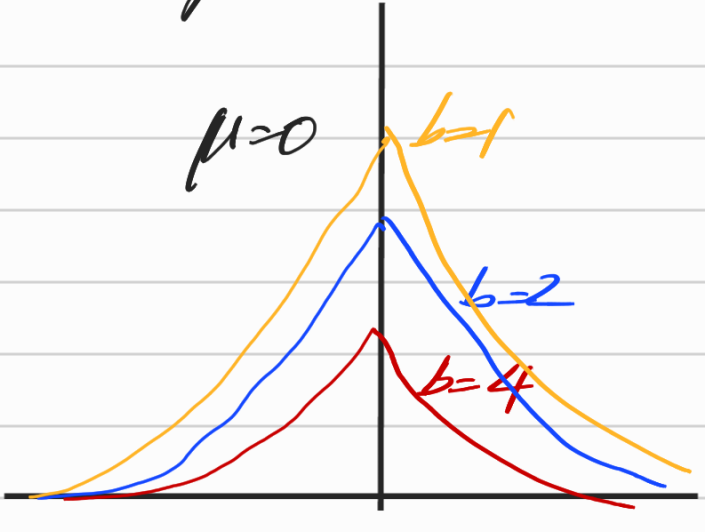
```
def query(db):
    return torch.sum(db.float())
                + noise
```

// Add noise to the op of query/func

Noise can be of two types :-

1) Gaussian Noise
2) Laplacian Noise





How to measure how much privacy is being leaked
→ For that, first define Diff Priv.

Formal Definition of Differential Privacy

A randomized algorithm M with domain $\mathbb{N}^{|x|}$ is (ϵ, δ)-differentially private if $\forall S \subseteq Range(M)$ and $\forall x, y \in \mathbb{N}^{|x|}$ such that $\|x - y\|_1 \leq 1$:

$$Pr[M(x) \in S] \leq \exp(\epsilon) Pr[M(y) \in S] + \delta$$

Say M is:

```
def M(db):
    return query(db) + noise
```

Not the above query, but sth like sum, avg, threshold etc

ie M is a randomized algo.
Could be either randomized after query (Global DP)
or at the db level (Local DP)

$S \subseteq Range(M)$
ie S is all the potential outputs $M(x)$ could predict

(Domain & Range of f)

Note: $M(x) = s$
$\in Domain(M)$
$\in S \subseteq Range(M)$

Note $x, y \in \mathbb{N}^{|x|}$ such that $\|x - y\|_1 \leq 1$

means that x & y are parallel dbs that differ by 1

Ex x is the full db
y is the full db with 1 entry missing

All of this such that the following holds:

$$Pr[M(x) \in S] \leq \exp(\epsilon) Pr[M(y) \in S] + \delta$$

$M(x)$: Query over full db
$M(y)$: Query over the parallel db with one entry missing

Both may return something in S

$Pr[M(x) \in S]$: Prob dist of all things in the db

$Pr[M(y) \in S]$: Prob dist of all things in db minus one entry

The above constraint equation asks
How much different are the two distributions
or How much does the removal of an entry change the output of the randomized function M

This difference is measured by two parameters
ie ϵ & δ

Eg: if $\epsilon = 0$, then $e^{\epsilon} = 1$ therefore

$$Pr[M(x) \in S] = Pr[M(y) \in S]$$

Note: ignoring δ

∴ If $\epsilon = 0$ & $\delta = 0$, perfect privacy

ie M shows no privacy leakage

Goal: Develop M such that we have DP with the lowest ϵ & δ possible

ϵ, δ called privacy budget

So, add a minimum amount of noise that satisfies the $\epsilon\delta$ privacy budget

How much Noise should we add:

Amount of noise depends upon:

1. Type of Noise (Gaussian/Laplacian)
2. Sensitivity of Query
3. Desired ϵ & δ

For Laplacian Noise:-

parameterized by b & μ

set $\mu = 0$

$b = \text{Sensitivity(query)} / \epsilon$

for Laplacian δ is assumed to be 0

np.random.laplace()