Differential Privacy

Core Concepts

Definition: Differential privacy is a mathematical framework that guarantees privacy by adding calibrated noise to data or query results, ensuring individual contributions cannot be reliably identified while maintaining overall statistical accuracy.

Key Properties:

- **Privacy Budget (ε)**: Controls the privacy-utility tradeoff; smaller ε means stronger privacy but less accurate results
- **Composition**: Privacy guarantees degrade predictably when multiple queries are performed
- **Post-processing**: Privacy guarantees remain even after further processing of differentially private outputs

Implementation Approaches

Local vs. Central Differential Privacy:

- **Local DP**: Noise added on user devices before data collection (stronger privacy, lower utility)
- **Central DP**: Noise added by a trusted central entity after data collection (better utility, requires trust)

Common Mechanisms:

- **Laplace Mechanism**: Adds noise drawn from a Laplace distribution, suitable for numeric queries
- **Exponential Mechanism**: For non-numeric queries, selects outputs with probability proportional to utility
- **Gaussian Mechanism**: Uses Gaussian noise, often preferred for complex analytics

Applications:

- Census data
- Location data
- Healthcare analytics
- Machine learning
- Usage statistics

Selection Criteria:

- Data sensitivity level
- Required accuracy
- Trust in central authority
- Computational constraints
- Query complexity