Differential Privacy

Parte IV

Indice

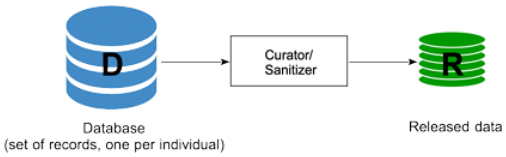
Capitolo 1

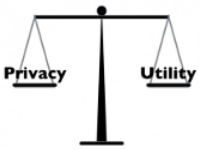
Introduzione

Introdotto nel 2006, ha avuto un'esplosione di utilizzo di questo concetto negli ultimi anni.
L'idea è la protezione dei dati, i dati sono privati e non basta che siano semplicemente anonimizzati.

Capitolo 2

Basic Scenario





1. **D:**Bisogna avere molti dati (se no D.P. non funziona bene), pensiamo di avere tanti record e ciascun record riguarda dati sensibili di individui.
2. **Curator/Sanitizer:** è una parte fidata che può accedere in chiaro ai dati
3. **R:** Dati rilasciati

2.1 L'intuizione classica per la privacy

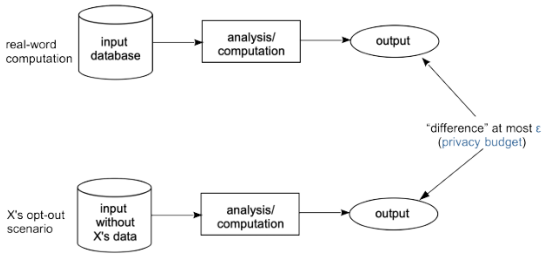
Cosa vuol dire mantenere la privacy degli individui?

- Ciò che viene rilasciato non dipende dai miei dati
- le informazioni apprese su un individuo dai risultati pubblicati R non sono più di quelle che possiamo apprendere su quell'individuo senza avere accesso a R.

Problema: Se gli individui non hanno impatto sui risultati allora i risultati non avrebbero utilità, serve equilibrio.

2.2 Differential privacy– Intuition

Quello che cerca di fare D.P. è che la presenza o l'assenza di un utente no cambi di molto il rischio di compromettere la sua privacy. Chiaramente il rilascio di informazioni impone che un minimo cambiamento ci sia, però sia minimale. Le inferenze che io posso fare su un individuo da un rilascio R deve essere lo stesos che io posso inferire utilizzando i dati di qualunque altro utente.



Più questo ϵ è piccolo più la differenza dei dati con, o senza, un determinato individuo è minimale. Tutto questo viene fatto inniettando del rumore all'interno della mia computazione. Deve essere rumore casuale (NON deterministico)
Due principali complicazioni:

- Capire quanto rumore inniettare
- Il risultato è sempre diverso poichè ogni volta viene inserita una quantità di rumore diversa (se viene posta la stessa domanda troppe volte si riesce a inferire il valore corretto, da questo nasce la necessità del **Privacy budget**)