Slide 3

Riesce a spiegare come spesso la funzione **exp** non riesca a spiegare dei fenomeni che avvengono in natura, come le regole di distanziamento Covid.

Grazie però a delle visualizzazioni riesce a spiegare perché a volte è più semplice utilizzare una funzione logaritmica per vedere il comportamento di un **exp**.

Poiché la funzione **exp** non risulta essere adatta per spiegare l’andamento dei contagi; cosa che invece riesce a fare la **funzione sigmoide**.

Si considerano anche i contatti tra persone già contagiate.

Osservando tutti questi numeri, non si riesce a vedere l’andamento esponenziale ed il nostro cervello non riesce a capire subito che c’è una grande distanza tra 17mila e 21mila.

Se si fa una visualizzazione, si vede chiaramente che all’inizio lo spazio aumenta tra i numeri ma di poco, poi nei giorni successivi c’è sempre un aumento della distanza.

Una volta che ci si abitua a capire che sono indicati i giorni sull’asse x e numero di contagiati sull’asse y, il cervello capisce subito, grazie a tale visualizzazione, quanti contagiati ci sono al giorno.

Questa visualizzazione in un semplice grafico un po’ per i colori, un po perché il cervello umano è abituato a vedere le funzioni, una semplice visualizzazione fatta bene permette di andare ad eprimere un concetto ovverro il fatto che il contagio non può sempre crescere, ad un certo punto si arriva ad una stabilità.

Questo è il primo esempio di utilità di una visualizzazione.

Slide 4

Esempio di simulazione di una epidemia.

Slide 5

L’utente può capire.

Ci sono delle visualizzazioni complesse, ma se è fatta bene l’utente puo imparare a capirle.

Se ad un uomo del primitivo si dava una visualizzazione della slide 3 all’inzio non la capisce, ma continuando a studiarla poi la comprende.

***Playfair*** è stato il primo a realizzare una visualizzazione → di una funzione per spiegare ad un bambino una funzione.

La maggior parte delle informazioni entra dalla vista.

Ormai le visualizzazioni si trovano in tutti i campi.

Nascono dagli anni 2000, 2005, quando le capacità di generazione dei dati hanno iniziato a crescere.

Ci sono molti studi legati allo studio del ***data mining*** che cercano di studiare la distribuzione dei dati nel mondo.

Sono andati a vedere tutti i dati che possono essere prodotti all’interno di una casa (americana in questo caso).

Si tratta di calcoli puramente statistici, basati su stime.

[ Per **dati strutturati** si intendono quei dati che un computer può visualizzare, quei dati **digitalizzabili**, ossia email, foto, video,… ]

Slide 7

Ora facciamo 5 exabytes in 2 giorni perché la potenza computazionale è cresciuta.

**Dati non strutturati**: dati generati dai processi di manifattura.

***Digital twin***: è un gemello digitale di un’entità reale.

Un uomo è un’entità reale.

I digital twin sono nati dai macchinari.

Questi ultimi generano dei dati che possono essere inviati tramite dispositivi Wi-Fi.

Ad un certo tempo ci si è resi conto che tali dati possono essere valutati, processati e visualizzati, in modo magari da prevenire la rottura del macchinario → in questo modo si può prevedere quando si romperà il macchinario ed evitare che l’azienda spenda troppi soldi.

Più aumenta la capacità di immagazzinamento più aumenta la capacità di memorizzazione.

Slide 9

C’è un aumento esponenziale.

Le previsioni dicono che seguirà una sigmoide come quella vista alla slide 3 sul Covid, per cui ad un certo punto non si potrà più aumentare.

Slide 10

Se prima i dati non erano strutturati, con l’aumenta della conoscenza, la percentuale di dati capibili e strutturabili aumenta.

Come mai questo aumento dei dati e questo aumento di strutturazione?

- Perché sempre più aziende dedicano più soldi per l’analisi dei dati (dei clienti, dei dipendenti…)

- Perché ci sono sempre più persone che hanno Internet anche nei paesi in via di sviluppo.

- Perché ormai ogni persona ha un proprio cellulare, ciascuno che produce dei dati.

Bisogna sapere perciò **processare** questi dati e poterli **visualizzare**.

Slide 12

Bisogna imparare a strutturare questi dati.

Noi umani riusciamo a vedere massimo 3 dimensioni.

Se focalizziamo un foglio Excel con una sequenza di vettori non riusciamo a vederli.

Dobbiamo prendere un programma e fare il plotter dei dati che vediamo.

Un’alternativa più creativa:

Ci sono persone che hanno provato a visualizzare dei dati in modo più grafico.

→ **infografica**

L’infografica in slide proviene da un *Digital Information World Blog*.

Cerca di trasmettere la quantità di informazione che in ogni minuto nel 2018 era in giro per il mondo.

La scritta bianca sullo sfondo nero → il nero crea una sensazione di oppressione o serve per attirare l’attenzione.

Sulla DX un colore azzurro a diverse tonalità per andare a vedere le differenze tra le diverse cose che portano ai dati; si hanno bande schiacciate che danno un senso di oppressione anche loro → vul dire che ci sono tanti dati.

Se i dati fossero stati scritti in modo più snello, davano meno un senso di oppressiome.

Questa infografica mostra una miriade di dati.

Vuole dire che ogni giorno siamo bombardati da tantissime informazioni.

Vengono usate delle icone che in qualche modo riguardano l’applicazione da cui provengono quei dati (per Twitter hanno usato il simbolo di Twitter).

Slide 13

Le visualizzazioni servono per farci capire un concetto, vogliono incitarci a votare qualcuno, vogliono invogliarci a comprare un prodotto…

Domo si occupa di data analysis.

Slide 14

Netflix fa enorme analisi di dati, per proporre agli utenti contenuti sempre più attinenti ai loro gusti.

Grazie a questo, Netflix è riuscita ad ottenere più utenti.

Riesce a fornire subito ciò che l‘utente vuole (basti pensare a Rai Play che spesso non mostra buone raccomandazioni di contenuti).

Slide 15

La visualizzazione serve per dare un’idea visiva di ciò che l’utente sta vedendo, dei dati che sta guardando.

Si impara sempre di più a strutturare i dati per renderli più comprensibili, per vedere i malfunzionamenti di una macchina, per monitorare la salute delle persone…

[ Anche nel campo medico le visualizzazioni sono utilissime → per comprendere al meglio grandi quantità di dati. ]

[ Il Covid ha portato ad un aumento di visualizzazioni.]

Slide 16

BiG Data è una quantità di dati enorme.

Esempio

In un pronto soccorso ogni paziente viene rappresentato come un vettore con i vari dati, in più si avrà una rappresentazione del paziente tramite un’immagine.

Si ha il paziente visto con una rappresentazione multimodale, ossia vedo il paziente in diverse modalità (immagine, vettore..).

Se si riuscissero a visualizzare tali dati, riusciamo a scartare informazioni non utili al nostro scopo.

Esempio: si sta studiando chi sopravvive al Covid → se il paziente ha gli occhi azzurri o verdi non mi interessa, questa informazione posso scartarla.

Se si va invece sul sito di incontri, il colore degli occhi invece è utile come informazione.

Si può fare uno **scatter plot** del colore degli occhi rispetto alla razza.

Vedere molte volte vuol dire capire i dati: l’umano riesce a comprendere meglio come **processarli**, come **sfruttare la loro eterogeneità** per processarli.

Slide 17

La ***data visualization*** serve a capire questa mole infinita di dati.

[ **Grafo di conoscenza** - Se si vedesse il *Twitter graph* si vedono tantissimi dati. ]