

Hejsan! På föreläsningen ger vi mer exempel och går in närmare på algoritmerna. Vi ses!

Sammanfattning om Data Mining

Grundidé:

- ♣ Hitta information i stora datamängder.
 - Det ska vara tidigare okänd information som gömmer sig i datamassorna.
 - ♣ Information ska vara relevant för den uppgift du vill lösa.

Tre citat:

- ♣ ”Vi drunknar i data men är svältfödda på kunskap”
- ♣ ”Att med automatiska metoder få fram dold information ur stora datamängder”
- ♣ ”Det har alltid funnits människor som söker efter samband bland de 1000 senaste roulette-numren. Tyvärr lyckas de”

Exempel på användningsområden:

- ♣ Banker: ska en kund få kredit eller inte.
- ♣ Försäkringsbolag: Samkörning av medicinska uppgifter i register (Olagligt)
- ♣ Sportstatistik: Förutsäga motståndarens taktik.

Data minings' ursprung:

- ♣ Databaser, Statistik, Maskinlärande, Visualisering, Design av algoritmer.

Data mining är inte traditionell statistik, som kräver att man ställer upp hypoteser och designade experiment. Det är inte heller Customer Relations Management (CRM) eller visualisering eftersom inga modeller genereras automatiskt. Men fältet är svårt att avgränsa.

Eftersom sanningen ofta ligger i betraktarens öga för vad som är relevant information, finns det många som åberopar ”sin” variant av data mining.

- ♣ Stora datamängder verkar alltid innehålla spännande samband.
 - Dessa samband kan alltid ges en spännande förklaring – overfitting.

Data Mining har alltid funnits, vi har alltid tolkat gudarnas tecken. Det finns ett stort inflytande i vetenskapens utkanter – senast i somras: DaVincikoden. Den alternativa förklaringen har alltid lockat. Varför?

- ♣ Man hittar vad man söker.

Data mining-tekniker och algoritmer

- ♣ Beskrivande tekniker, summerar data och understryker de intressanta egenskaperna:
 - Generalisering – jämföra klasser för att hitta skillnader
 - Associering – relationer mellan egenskaper
- ♣ Förutsägande tekniker, bygger modeller för att förutsäga framtida beteenden:
 - Maskinlärande, (AI)
 - Bayesiska metoder
 - Neurala nätverk

- ♣ Det finns algoritmer med olika angreppssätt som utför data mining.