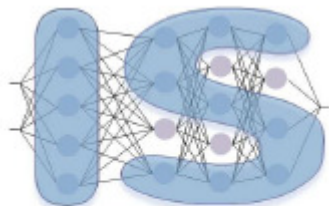




Budapesti Műszaki és Gazdaságtudományi Egyetem Méréstechnika és Információs rendszerek Tanszék



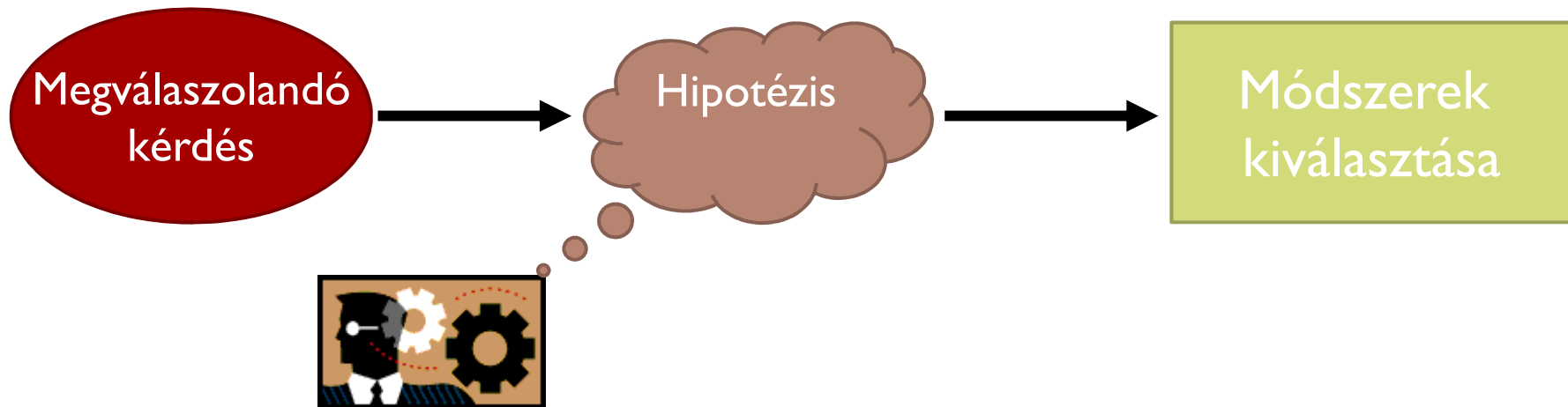
**Intelligent Systems
Research Group**

Adatelemzés intelligens módszerekkel

Hullám Gábor



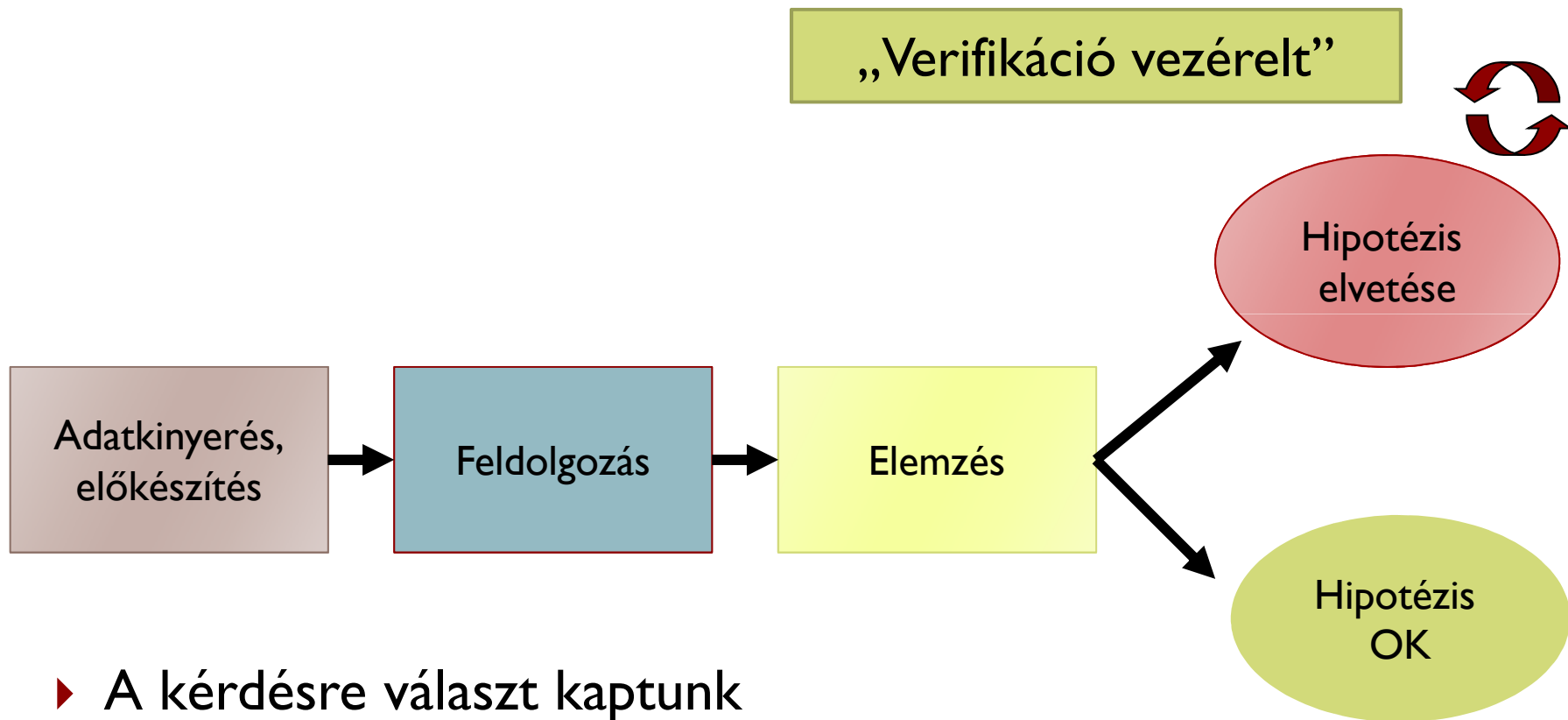
Adatelemzés hagyományos megközelítésben I.



- ▶ Milyen adaton?
- ▶ Milyen módszerrel?
- ▶ Mit keresünk?



Adatelemzés hagyományos megközelítésben II.



- ▶ A kérdésre választ kaptunk
- ▶ Az igazi kérdés: jó volt-e a kérdés?



Felmerülő kérdések

- ▶ A feltett hipotézis valóban a megválaszolandó problémára/kérdésre ad magyarázatot?
- ▶ Jól határoztuk meg, hogy mit veszünk figyelembe és mit nem?
- ▶ Nincs-e más, a vizsgált dolgokon kívüli, fontos információ az adatokban elrejtve?



Adatbányászat I.

- ▶ „The **nontrivial** extraction of **implicit**, **previously unknown**, and **potentially useful** information from data.”

*-Frawley, Piatetsky-Shapiro, and Matheus
Knowledge Discovery in Databases
AI Magazine Vol.13 Num. 3 (1992) AAAI*

- **Nontrivial:** Magas szintű tudás kinyerése, mint szabályok, kauzális kapcsolatok, predikciók, részhalmazok
(\leftrightarrow *SQL queryk, alap statisztikai adatok*)



Adatbányászat II.

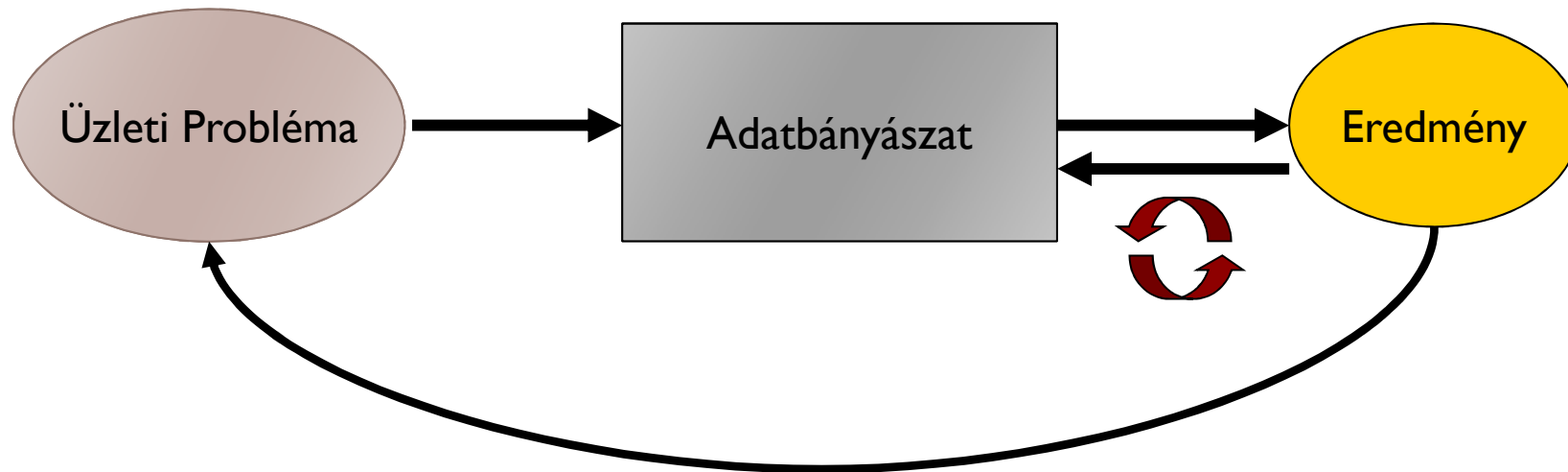
- ▶ **implicit:** Az adathalmazban kifejeződő rejtett szabályszerűségek
- ▶ **previously unknown:** korábban nem ismert összefüggések, mintázatok
- ▶ **potentially useful:** üzleti értéket hordozó kapcsolatok és szabályok kinyerése, melyek közvetlen döntéstámogatási inputként szolgálhatnak



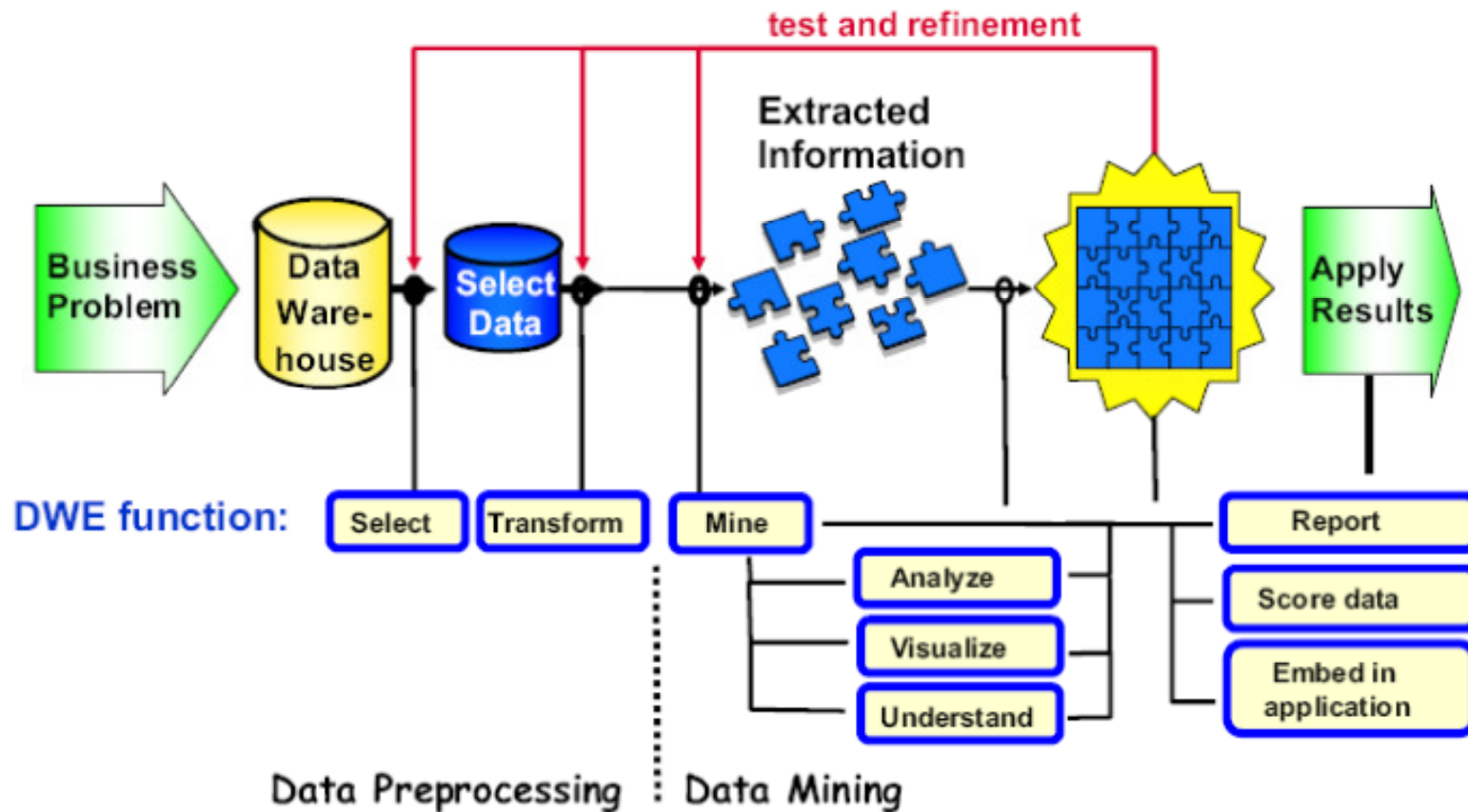
Adatelemzés adatbányászati megközelítésben



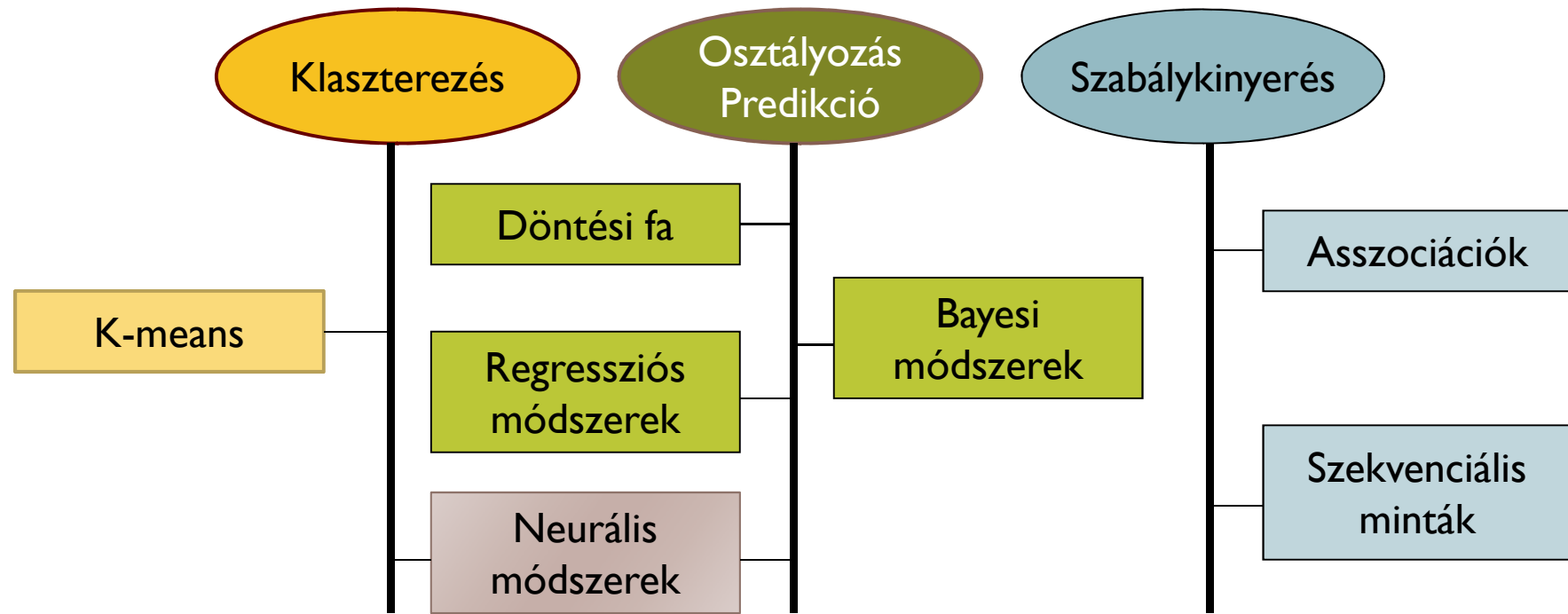
„Felfedezés vezérelt”



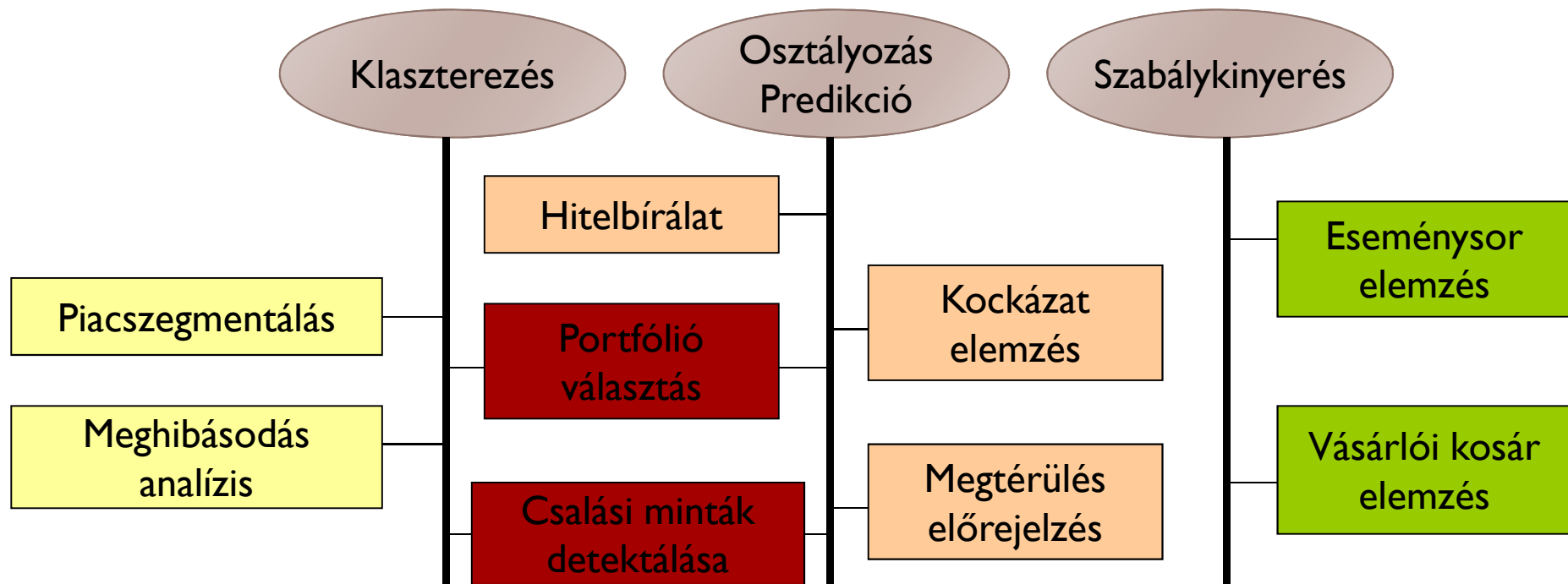
Az adatbányászat folyamata



Adatbányászati eszközök



Alkalmazások

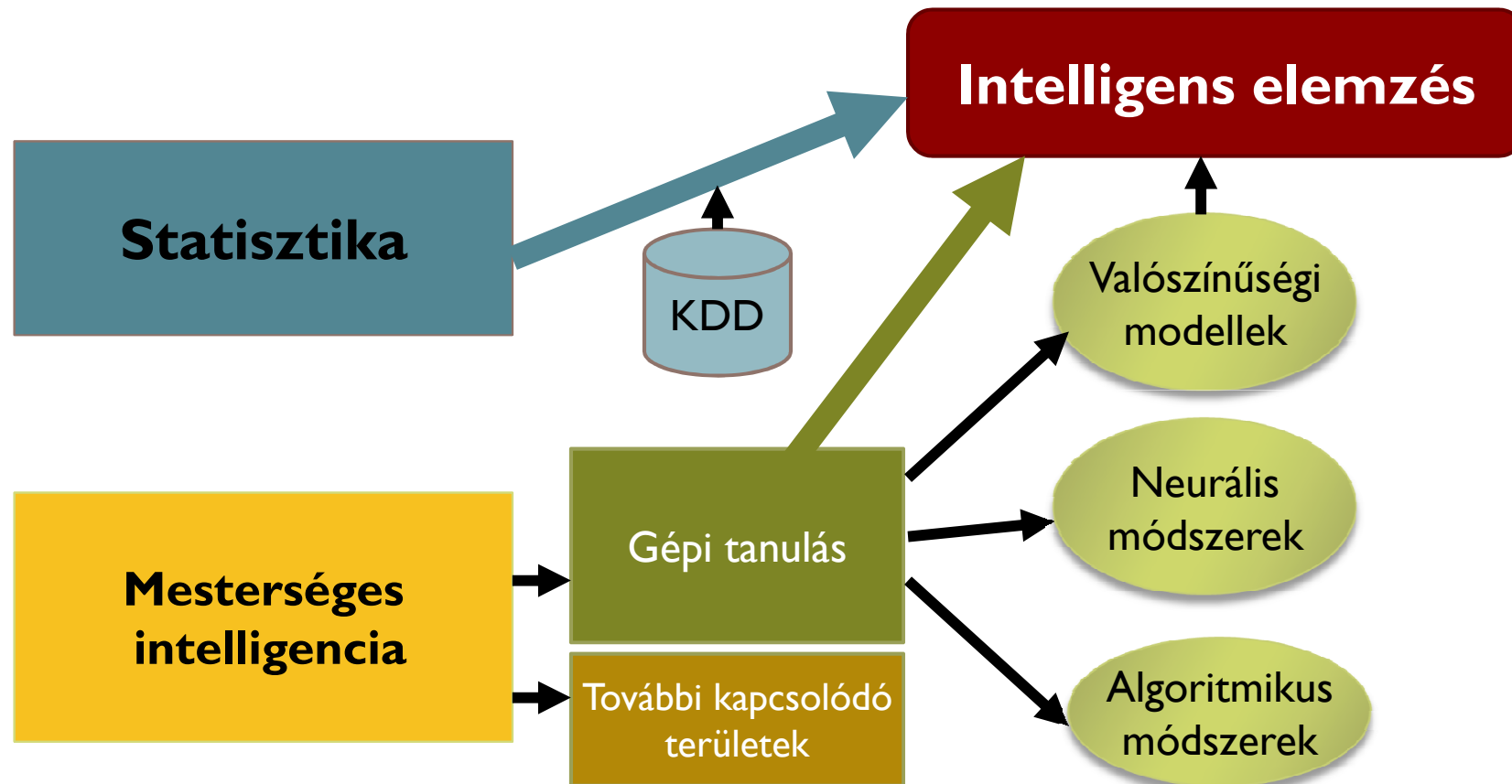


Intelligens adatelemzés

Hozzávetőleges meghatározás:

- ▶ Statisztikai és gépi tanulási módszerek alkalmazása komplex adatelemzési és modellezési feladatok megoldására
- ▶ A hangsúly az alkalmazott módszereken van
- ▶ A megközelítés a korábbi kétpólusú (adatbányász vs. klasszikus) felálláshoz képest hibrid

Intelligens adatalemzés

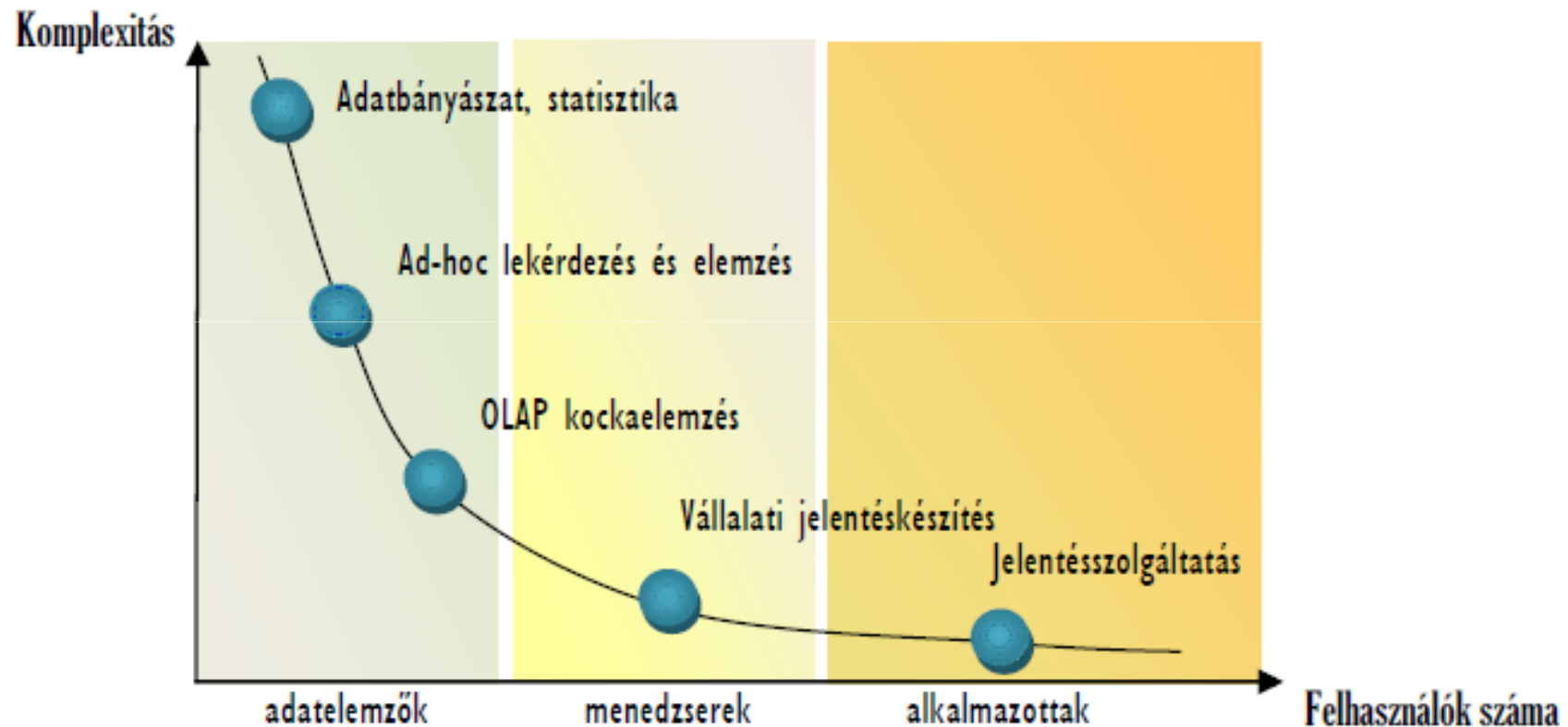


Kihívások

- ▶ Nagy adathalmazok kezelése (BIG DATA)
- ▶ Számítási igény kezelése
- ▶ **Eredmények kiértékelése, értelmezése**

- ▶ Hiába van eredmény, ha nincs megfelelő értelmezés.
Ehhez szükség van:
 - ▶ Kiértékelést segítő „ügyes módszerekre”
 - ▶ Vizualizációs eszközökre
 - ▶ Esetenként szakértőkre

Elemzések felhasználása

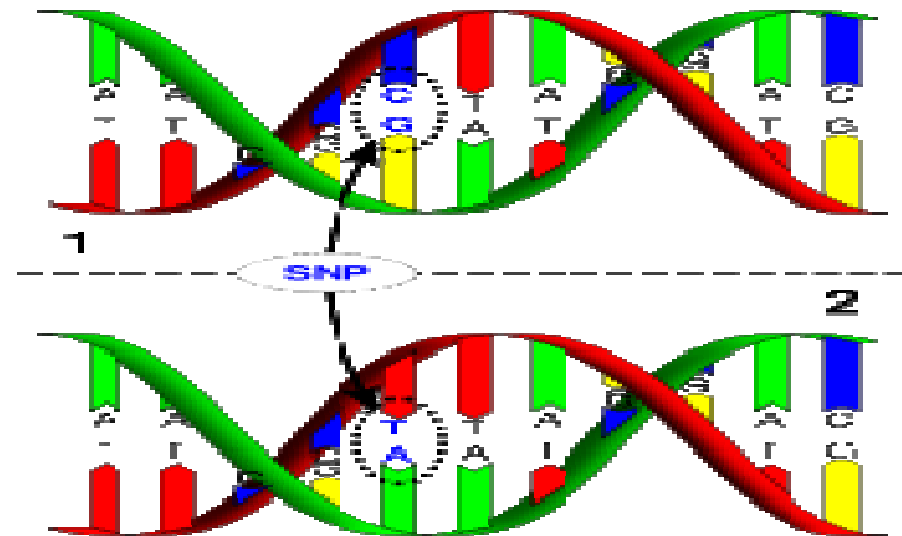




**Bayes-háló alapú adatelemzés
bioinformatikai adatokon**

Bioinformatika

- Orvos-biológiai, genetikai adatok elemzésén alapul
- Céljai:
 - Új elemzési módszerek kialakítása
 - Eredmények intelligens feldolgozása és kiértékelése
 - Tudásbázisok létrehozása
 - Biomarkerek kutatása



A bioinformatika kihívásai

Genetikai szabályozás

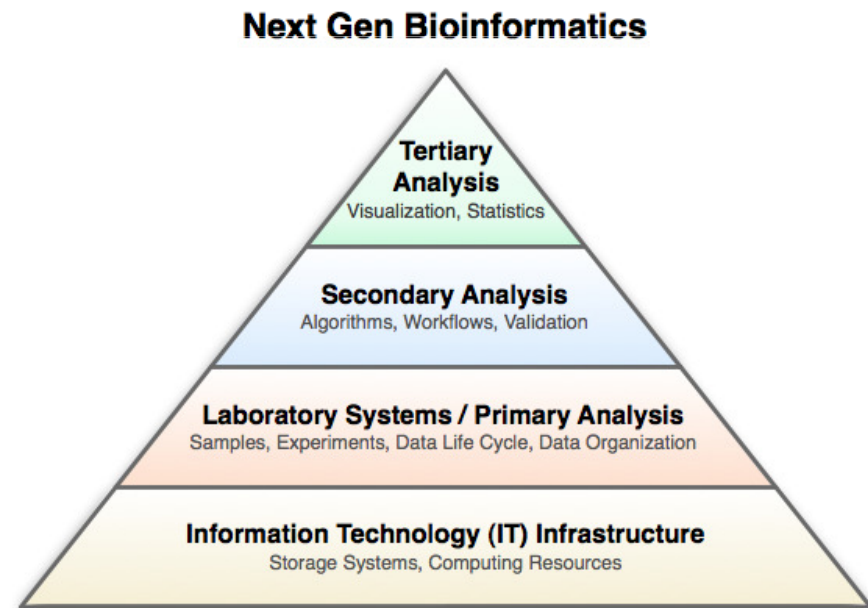
- ▶ Újabb szabályozó elemek keresése, célponthoz való kapcsolása

Genome Wide Association Studies (GWAS)

- ▶ Sok változó, változók számához képest kevés eset
- ▶ Feature Subset Selection

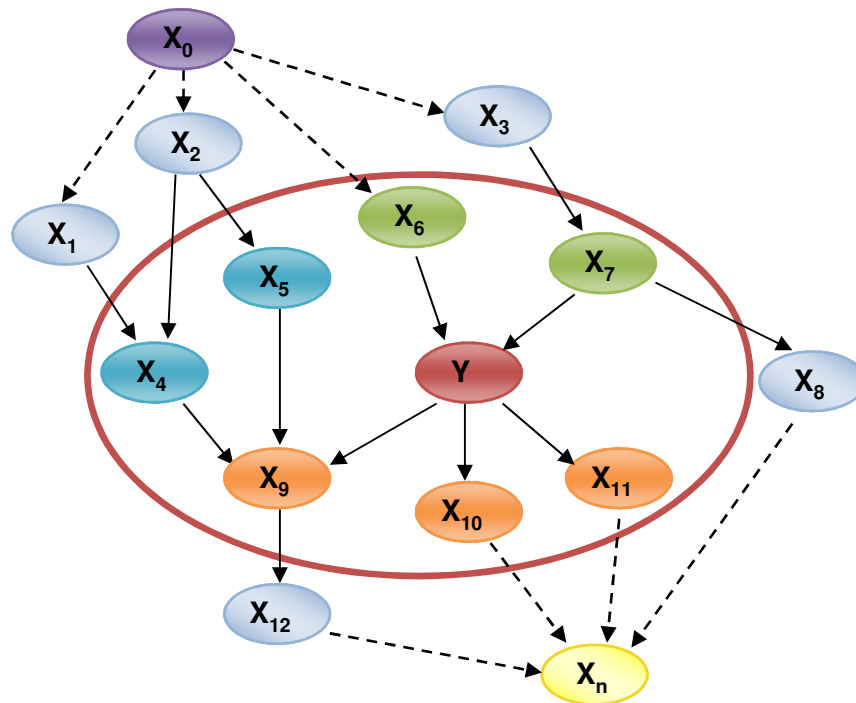
Újabb labor technológiák támogatása

- ▶ RNS szekvenálás
- ▶ *de novo* illesztés
- ▶ Variáns detektálás



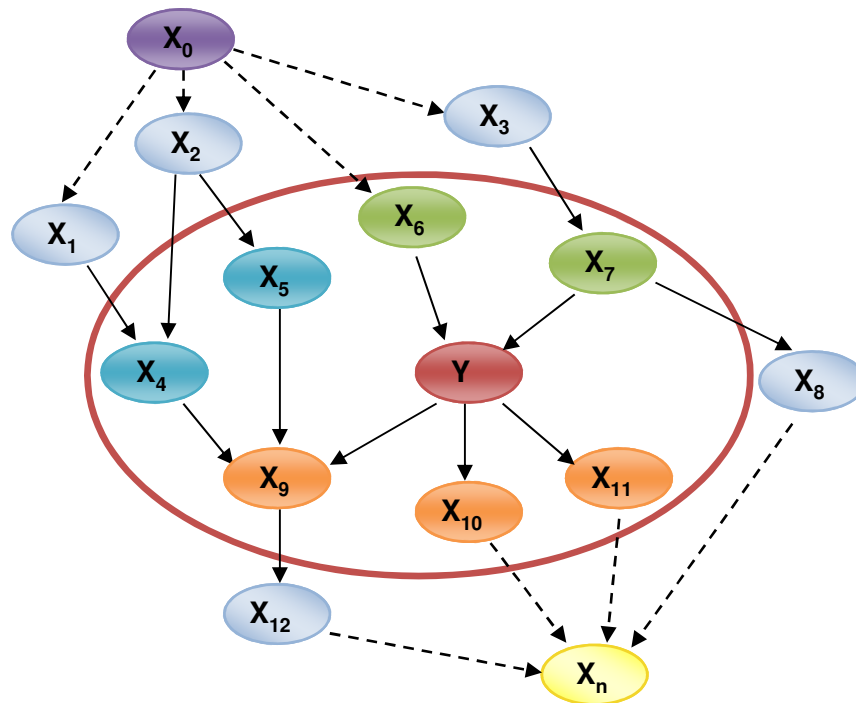
-
- ▶ <http://www.kurzweilai.net/international-conference-on-bioinformatics-computational-biology-genomics-and-chemoinformatics-bcbgc-10>

Bayes-háló alapú relevanciaelemzés I.



- ▶ A Bayes-háló , mint modellosztály lehetővé teszi a változók közötti függőségi kapcsolatok rendszer szemléletű modellezését
- ▶ Többváltozós függőségi minták modellezésére alkalmazható
- ▶ Strukturális és parametrikus tulajdonságok vizsgálhatók

Bayes-háló alapú relevanciaelemzés II.

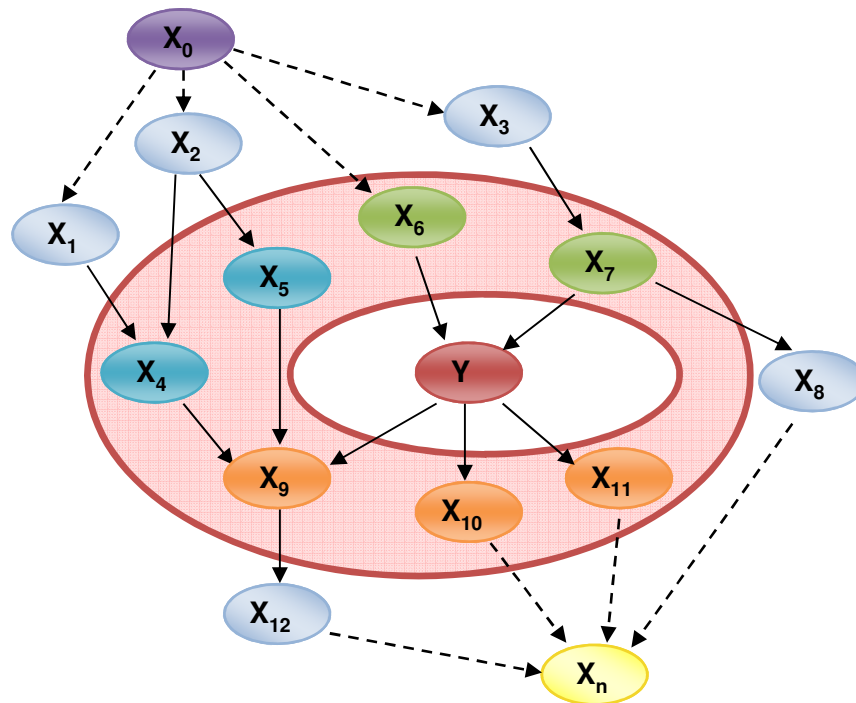


A módszer célja:

lehetséges Bayes-háló

- ▶ struktúrák,
- ▶ részstruktúrák,
- ▶ strukturális tulajdonságok tanulása az elérhető adathalmaz alapján
- ▶ Ezek alapján következtethetünk egyes változók relevanciájára

Bayes-háló alapú relevanciaelemzés III.



- ▶ Relevancia mindig egy (vagy több) célváltozóhoz képest definiált
- ▶ Bayesi **strukturális** és **parametrikus** relevancia mértékek segítségével azonosíthatók a célváltozó szempontjából releváns változók
 - ▶ **Strukturális**: erős relevancia
 - ▶ **Parametrikus**: hatáserősség

Génasszociációs adatok elemzése



- ▶ **Célváltozó:** betegségleíró (indikátor) változó
- ▶ **Változók:** genetikai faktorok, klinikai és környezeti faktorok
- ▶ **Adat:** genetikai minta és kiegészítő adatok beteg és egészséges páciensektől
- ▶ **Cél:** releváns genetikai faktorok azonosítása

Köszönöm a figyelmet!

- ▶ Gabor Hullam (gabor.hullam-at-mit.bme.hu)



Budapest University of Technology and Economics
Department of Measurement and Information Systems