

# MATEMATIKAI NYELVÉSZET OLVASÓSZEMINÁRIUM

Kornai András

2014 szept 12

- gépi fordítás (MT)
- beszédfelismerés (ASR)
- írásfelismerés (OCR)
- információkeresés (IR)
- nyelvmodellezés (LM)
- szófajmeghatározás (POS tagging)
- névelemkeresés (NER tagging)
- egyértelműsítés (disambiguation)
- szinonoma/parafrázis-keresés
- szótárépítés
- ... **Nem nyelvi:** arcfelismerés, tárgyfelismerés, ujjlenyomat, beszélő, ...

# AZ ELSŐ HASZNÁLHATÓ EREDMÉNYEK (HATVANAS ÉVEK)

## Tanítható osztályozók (HIGHLEYMAN, MUNSON)

*adat+cimke (arany)*

Tanulóalgoritmus

*Modell*

*csak adat*

Modell

*Hipotézis*

# A MUNKAMENET RÉSZLETESEBBEN

*nyers bemenő adat*

Jelfeldolgozás

*normalizált, jegyekre váltott, tisztított adat*

(szegmentálás)

*jegyvektor(sorozat)*

osztályozás (címkezés)

*hipotézis(sorozat/hálózat)*

kontextus-modell

*eredmény (hipotézis)*

- főkomponens-elemzés (PCA) Pearson (1901). Másnéven Karhunen-Loève transzformáció Ma elsősorban adatredukcióra használjuk
- Lineáris diszkrimináció (LDA) Fisher (1936) Másnéven ANOVA
- Ideghálók (ANN)
- Legközelebbi szomszédok (Nearest neighbor) A metrika (pl. tangens-távolság) maga is tanulható!
- Rejtett Markov (HMM)
- Maximum entrópia (maxent) Az SVM-ek maguk is tekinthetők diszkriminatív maxentnek Jaakkola (1999).

- Genetikus/evolúciós
- Feljavítás (boost) gyenge tanulók, súlyozás, költségfüggvény
- Bayes/grafikus
- Döntési fa
- szabályalapú, hibrid (szabályindukció)
- Kombinált (szimulált hűtés hőfoktól függően más és más tanulóalgt. tud lenni Rendell (1987))