Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Mesterséges Intelligencia és Rendszertervezés Tanszék

VIMIAC16 2025/26/I.

Modell struktúra és paraméterezés

Paramétertanulás – Naiv Bayes-háló

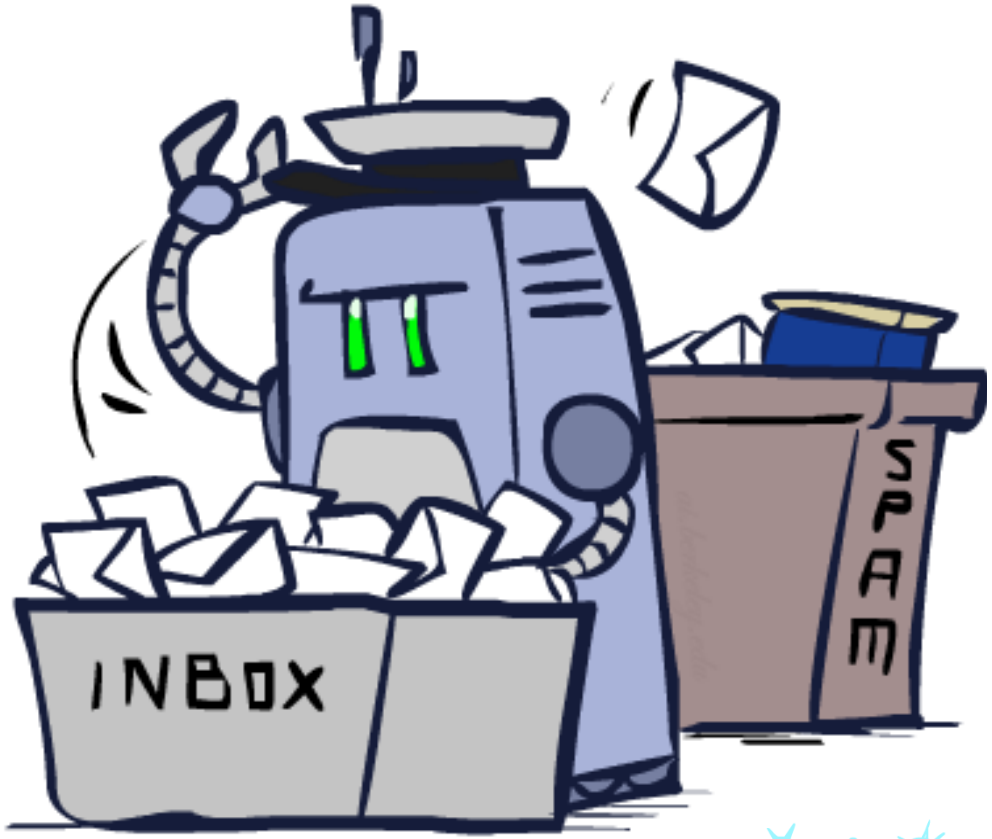
Előadó: Dr. Hullám Gábor

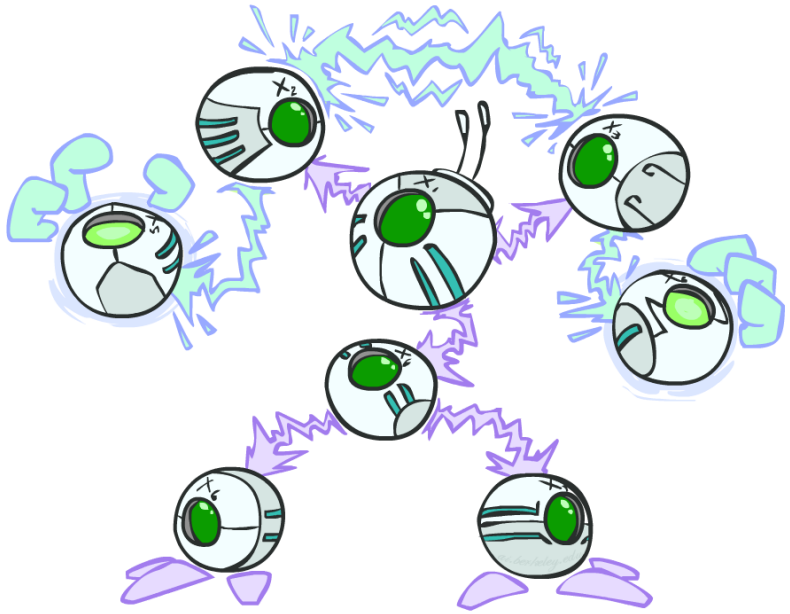
Artificial intelligence lectures

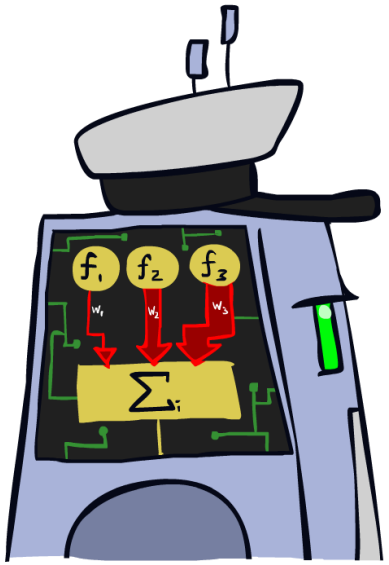
Az előadás diái az AIMA könyvre épülve (http://aima.cs.berkeley.edu) készültek a University of California, Berkeley mesterséges intelligencia kurzusának anyagainak felhasználásával (http://ai.berkeley.edu).

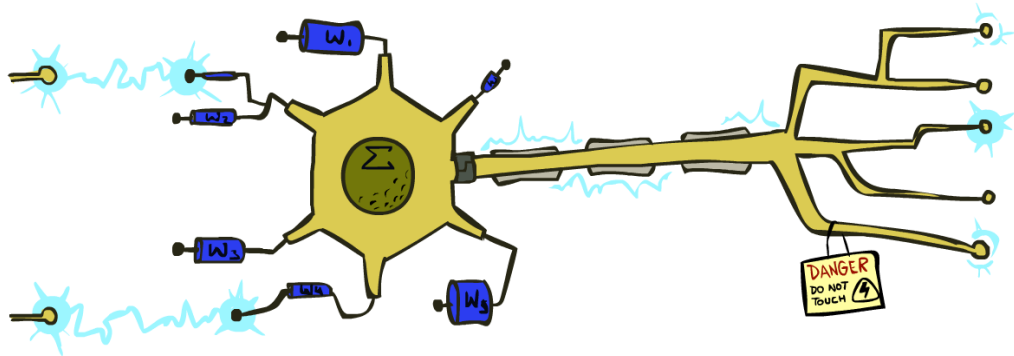
These slides are based on the AIMA book (http://aima.cs.berkeley.edu) and were adapted from the AI course material of University of California, Berkeley (http://ai.berkeley.edu).

Osztályozás









Példa: Spamszűrő

- Input: an email
- Output: spam/ham

- Setup:
  - Szükség van egy nagy gyűjteményre példa-e-mailekből, amelyek mindegyike "spam" vagy „ham" címkével van ellátva
  - Megjegyzés: valakinek kézzel kell címkéznie ezeket az adatokat!
  - Szeretnénk megtanulni megjósolni az új, jövőbeli e-mailek címkéit
- Jegyek: A ham/spam döntés meghozatalához használt attribútumok
  - Szöveg
  - Szövegminták: $dd, CAPS
  - Nem-szöveg:SenderInContacts

Dear Sir.

First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. …

TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.

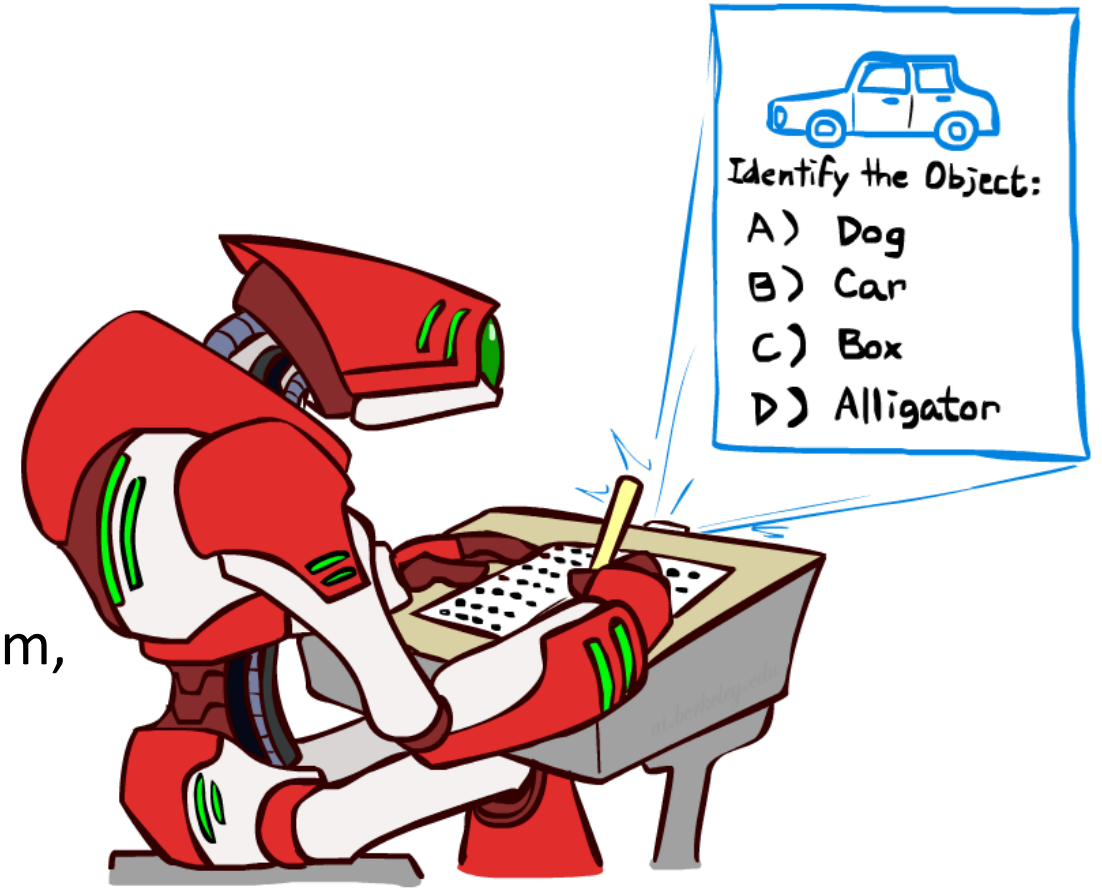
99 MILLION EMAIL ADDRESSES FOR ONLY $99

Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.

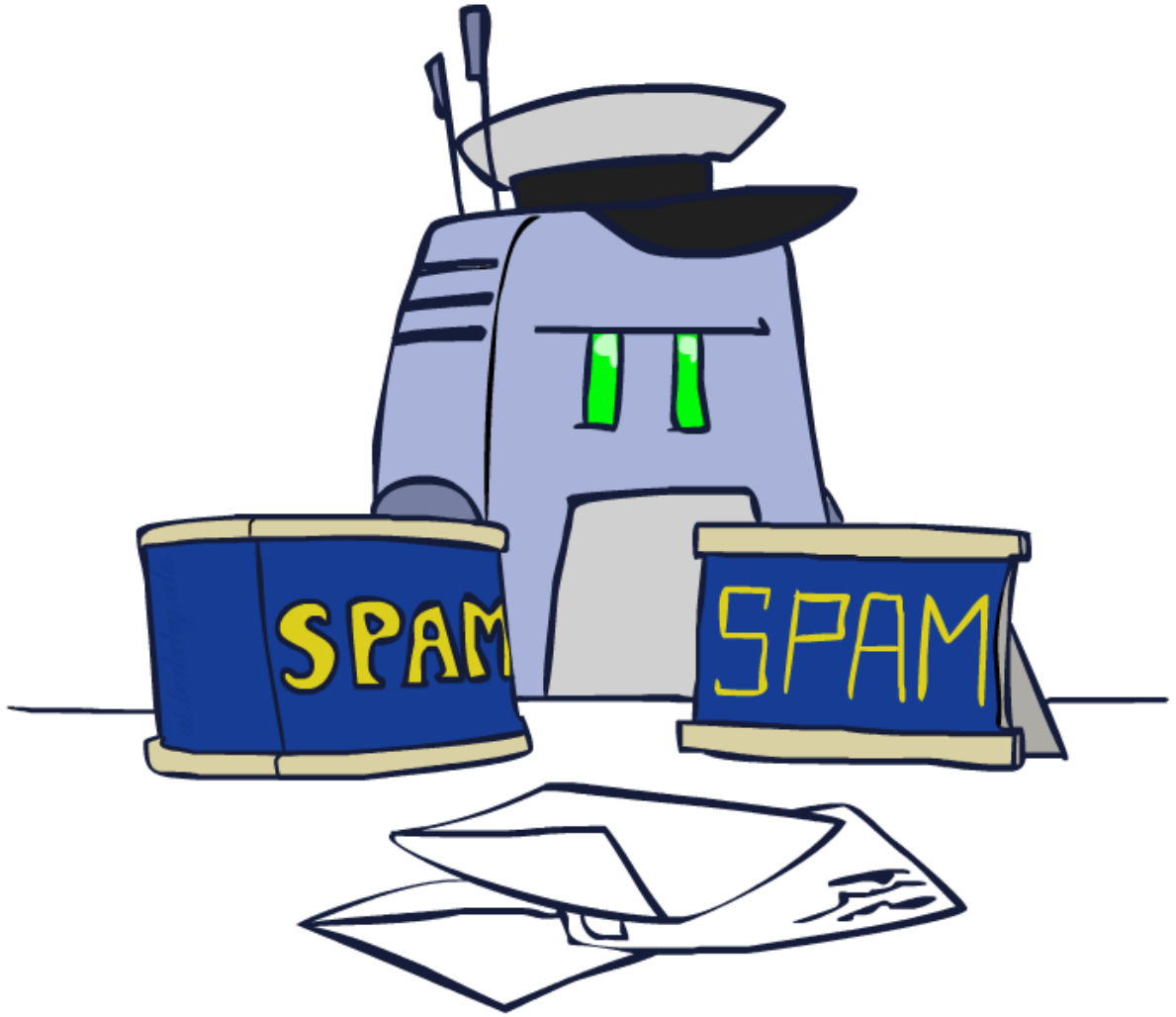
Egyéb osztályozási feladatok

- Osztályozás: adott x bemenetekre, y címkék (osztályok) előrejelzése
- Példák:
  - Spam észlelés (bemenet: dokumentum,
  - Osztályok: spam / ham)
  - OCR (input: Képek, Osztályok : Karakterek)
  - Orvosi diagnózis(input: Tünetek,
    Osztályok : Betegségek)
  - Automatikus esszéosztályozás (input: dokumentum,
    Osztályok : eredmény - érdemjegy)
  - Csalások felderítése(input: account tevékenység,
    Osztályok : csalás / nincs csalás)

… még sok más

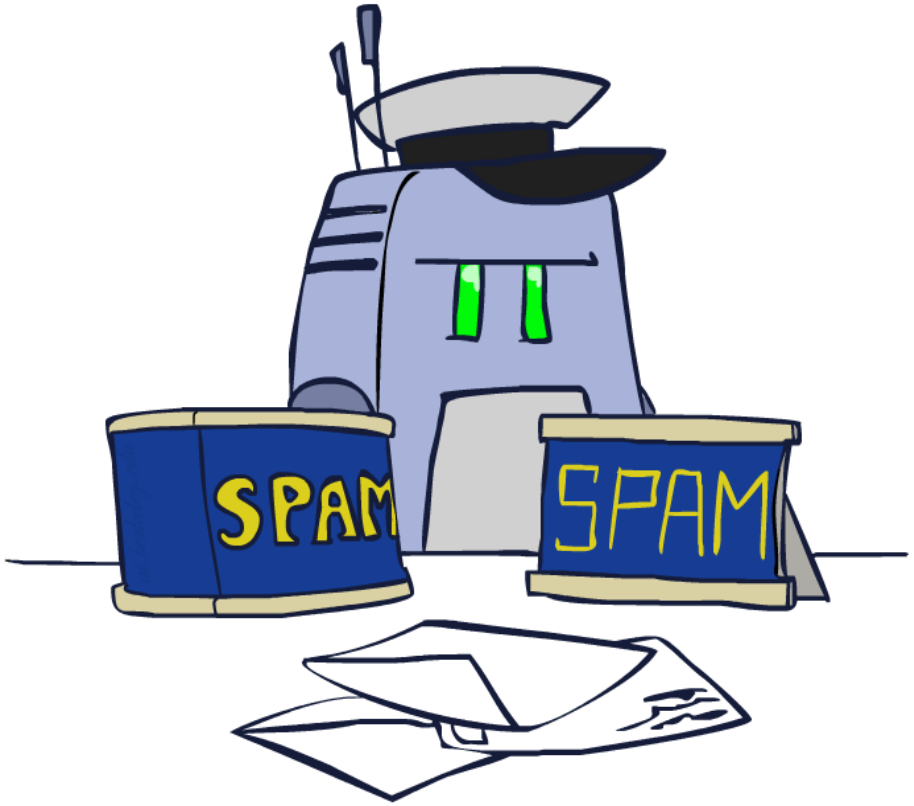


Modell alapú osztályozás



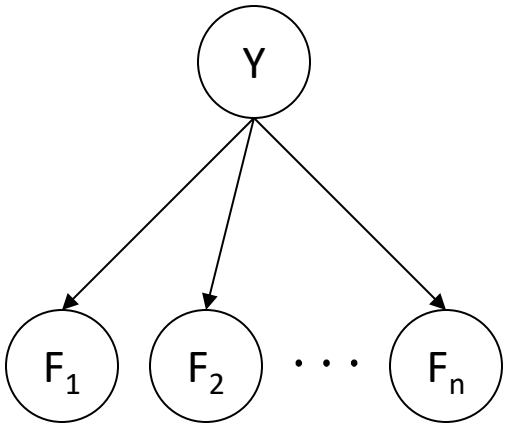
Modell alapú osztályozás

- Modellalapú megközelítés
  - Építsünk egy modellt (például Bayes-hálót), ahol mind a címke, mind a jellemzők véletlen változók
  - A megfigyelt jellemzőknek „értékadás” az adathalmaz alapján
  - A címke eloszlásának lekérdezése a jellemzőktől függően
- Kihívások
  - Milyen struktúrával kell rendelkeznie a Bayes-hálónak?
  - Hogyan kell megtanulnunk a paramétereit?



Naiv Bayes-háló

- Egy általános Naiv Bayes-háló modell:



|Y| paraméter értékei

$$P(\mathsf{Y}, \mathsf{F}_1 \ldots \mathsf{F}_n) = P(\mathsf{Y}) \prod_i P(\mathsf{F}_i|\mathsf{Y})$$

|Y| x |F|^n értékek

n x |F| x |Y| paraméterek

- Csak azt kell meghatároznunk, hogy az egyes jellemzők hogyan függenek az osztálytól
- A paraméterek száma *lineáris* n-ben
- A modell nagyon egyszerű, de gyakran így is működik

Következtetés naiv Bayes-hálóban

- Cél: a poszterior eloszlás kiszámítása az Y címkeváltozó felett
  - 1. lépés: címke és a evidenciák együttes valószínűségének meghatározása az egyes címkékhez

$$P(Y, f_1 \dots f_n) = \begin{bmatrix} P(y_1, f_1 \dots f_n) \\ P(y_2, f_1 \dots f_n) \\ \vdots \\ P(y_k, f_1 \dots f_n) \end{bmatrix} \Rightarrow \begin{bmatrix} P(y_1) \prod_i P(f_i|y_1) \\ P(y_2) \prod_i P(f_i|y_2) \\ \vdots \\ P(y_k) \prod_i P(f_i|y_k) \end{bmatrix}$$

$$\frac{\begin{bmatrix} P(y_1) \prod_i P(f_i|y_1) \\ P(y_2) \prod_i P(f_i|y_2) \\ \vdots \\ P(y_k) \prod_i P(f_i|y_k) \end{bmatrix}}{P(f_1 \dots f_n)} \Rightarrow P(Y|f_1 \dots f_n)$$

+

  - 2. lépés: összegzés az evidenciák valószínűségének számításához
  - 3. lépés: normalizálás az 1. lépés és a 2. lépés eredményének elosztásával

Naiv Bayes-háló

- Mire van szükségünk a Naiv Bayes-háló használatához?
  - Következtetési módszer (most láttuk ezt a részt)
  - Helyi feltételes valószínűségi táblák becslése
    - P(Y), a priori eloszlás a címkékre
    - $P(F_i|Y)$ minden jellemzőre (evidencia változó)
    - Ezeket a valószínűségeket együttesen a modell *parametereinek* nevezzük és θ*-val jelöljük*
    - …Ezek általában a tanító adatok előfordulási gyakoriságaiból származnak

A Spam Filter

- Naiv Bayes spamszűrő
- Adat:
  - E-mailek gyűjtése, spam vagy ham címkével
  - Megjegyzés: valakinek kézzel kell felcímkéznie ezeket az adatokat!
  - Tanító, validációs és teszt adathalmazokra osztva
- Classifiers
  - Tanítás a tanító adathalmazon
  - Hangolás a validációs adathalmazon
  - Tesztelés új emaileken (tesztadathalmazon)

Dear Sir.

First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. …

TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.

99 MILLION EMAIL ADDRESSES FOR ONLY $99

Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.

Naiv Bayes-háló SPAM-hez

- Bag-of-words Naiv Bayes-háló:
  - Jellemzők: W_i a szó az i. pozícióban
  - Mint korábban: Jellemzőktől (változóktól) függő címke előrejelzése (spam vs. ham)
  - Mint korábban: Tegyük fel, hogy a jellemzők feltételesen függetlenek az adott címkétől
  - Új: mindegyik W_i azonos eloszlású

Szó az i-edik helyen, nem az i –edik szó a szótárban!

- Generatív modell: $$P(Y, W_1 \dots W_n) = P(Y) \prod_i P(W_i|Y)$$

- "Kötött" eloszlás és bag-of-words
  - “bag-of-words”-nek hívjuk, mert a modell érzéketlen a szórendre vagy az átrendezésre

Példa: Spamszűrés

- Modell: $P(Y, W_1 \dots W_n) = P(Y) \prod_i P(W_i|Y)$
- Mik a paraméterek?

$P(Y)$

```
ham : 0.66
spam: 0.33
```

$P(W|\text{spam})$

```
the :  0.0156
to  :  0.0153
and :  0.0115
of  :  0.0095
you :  0.0093
a   :  0.0086
with:  0.0080
from:  0.0075
...
```

$P(W|\text{ham})$

```
the :  0.0210
to  :  0.0133
of  :  0.0119
2002:  0.0110
with:  0.0108
from:  0.0107
and :  0.0105
a   :  0.0100
...
```

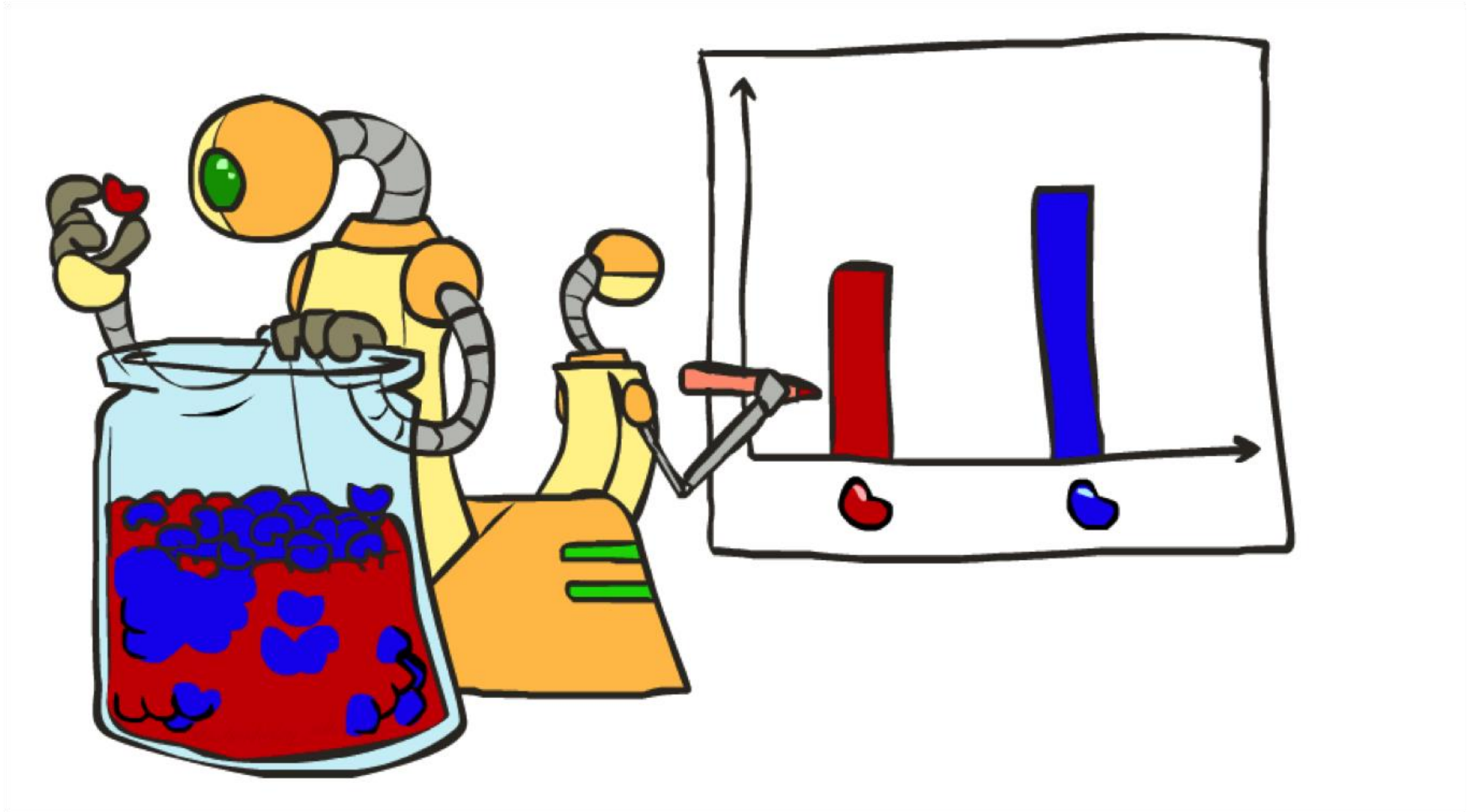
Általánosítás és túltanulás

- A relatív gyakorisági paraméterek túlilleszkednek a tanító adatokhoz!
  - Nem valószínű, hogy a "perc" minden előfordulása 100% spam
  - Nem valószínű, hogy a "komolyan" minden előfordulása 100% ham
  - Mi a helyzet azokkal a szavakkal, amelyek egyáltalán nem fordulnak elő a tanító adathalmazon?
  - Általában nem tehetjük meg, hogy a nem látott események nulla valószínűséget kapjanak

- Szélsőséges eset: Teljes e-mail az egyetlen jellemző
  - Tökéletessé tenné a tanító adathalmazt (ha determinisztikus a címkézés)
  - Egyáltalán nem általánosítana
  - A bag-of-words a feltételezés ad némi általánosítást, de nem elég

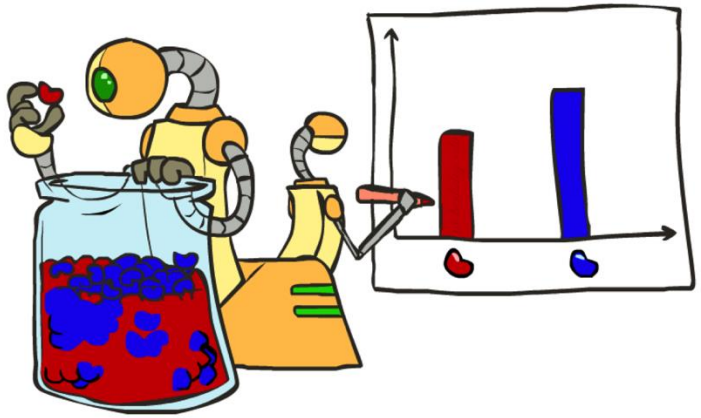
- Jobb általánosításhoz: simítanunk vagy regularizálnunk kell a becsléseket

Paraméter becslés



Parameter becslés

- Véletlen változó eloszlásának becslése
- *Szakértővel:* Kérdezzünk meg egy embert!
- *Empirikusan: tanító adatok használata (tanulás!)*
  - Pl.: Minden x eredménynél nézzük meg az adott érték empirikus arányát:



$$P_{\text{ML}}(x) = \frac{\text{count}(x)}{\text{total samples}}$$

$$P_{\text{ML}}(\text{r}) = 2/3$$

  - Ez az a becslés, amely maximalizálja az adat *likelihoodját*

$$L(x,\theta) = \prod_i P_\theta(x_i) = \theta \cdot \theta \cdot (1-\theta)$$

$$P_\theta(x = \text{red}) = \theta$$

$$P_\theta(x = \text{blue}) = 1 - \theta$$

Példa: Paraméterbecslésre

- Azt mondja egy ismerősöd: Van egy rajszögem, ha feldobom, mekkora a valószínűsége annak, hogy a szöggel felfelé esik?
- Azt mondod: Kérlek, dobd fel néhányszor:

  - Azt mondod: A valószínűség:
    - P(H) = 3/5
  - **Azt mondja: Miért???**
  - Azt mondod: Mert...

Példa: Paraméterbecslésre

- P(„Fej”) = θ, P(„Írás”) = 1-θ

- feldobások *i.i.d. -k*: $D=\{x_i \mid i=1 \ldots n\}, \; P(D \mid \theta) = \Pi_i P(x_i \mid \theta)$
  - Független események
  - Ismeretlen eloszlás szerint azonos eloszlásúak
  - *D:* egy sorozat α_H Fejből és α_T írásból áll

$$P(\mathcal{D} \mid \theta) = \theta^{\alpha_H}(1-\theta)^{\alpha_T}$$

Maximum Likelihood becslés (MLE)

- **Adat:** Megfigyelt D halmaz : α_H Fej és α_T Írás
- **Hipotézis tér:** Binomiális eloszlások
- **Tanulás:** θ megtalálása egy optimalizálási probléma
  - Mi a célfüggvény?

$$P(\mathcal{D} \mid \theta) = \theta^{\alpha_H}(1-\theta)^{\alpha_T}$$

- **MLE:** θ kiválasztása úgy, hogy maximalizálja D valószínűségét

$$\begin{aligned} \hat{\theta} &= \arg\max_{\theta} P(\mathcal{D} \mid \theta) \\ &= \arg\max_{\theta} \ln P(\mathcal{D} \mid \theta) \end{aligned}$$

Maximum Likelihood Becslés

$$\begin{aligned}\widehat{\theta} &= \arg\max_{\theta} \ln P(\mathcal{D} \mid \theta) \\ &= \arg\max_{\theta} \ln \theta^{\alpha_H}(1-\theta)^{\alpha_T}\end{aligned}$$

- Állítsuk a deriváltat nullára, és oldjuk meg!

$$\begin{aligned}\frac{d}{d\theta} \ln P(\mathcal{D} \mid \theta) &= \frac{d}{d\theta}\left[\ln \theta^{\alpha_H}(1-\theta)^{\alpha_T}\right] \\ &= \frac{d}{d\theta}\left[\alpha_H \ln\theta + \alpha_T \ln(1-\theta)\right] \\ &= \alpha_H \frac{d}{d\theta}\ln\theta + \alpha_T \frac{d}{d\theta}\ln(1-\theta) \\ &= \frac{\alpha_H}{\theta} - \frac{\alpha_T}{1-\theta} = 0\end{aligned}$$

$$\widehat{\theta}_{MLE} = \frac{\alpha_H}{\alpha_H + \alpha_T}$$

Maximum Likelihood

- A relatív gyakoriságok maximum likelihood becslések

$$\theta_{ML} = \arg\max_{\theta} P(\mathbf{X}|\theta)$$
$$= \arg\max_{\theta} \prod_i P_\theta(X_i)$$

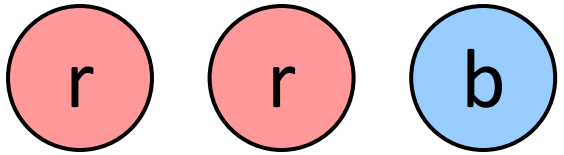
$$\Rightarrow \quad P_{\mathsf{ML}}(x) = \frac{\mathsf{count}(x)}{\mathsf{total\ samples}}$$

- Egy másik lehetőség az adatok alapján a legvalószínűbb paraméterérték figyelembevétele

$$\theta_{MAP} = \arg\max_{\theta} P(\theta|\mathbf{X})$$
$$= \arg\max_{\theta} P(\mathbf{X}|\theta)P(\theta)/P(\mathbf{X}) \quad \Rightarrow \quad ????$$
$$= \arg\max_{\theta} P(\mathbf{X}|\theta)P(\theta)$$

Laplace simítás

- Laplace becslés:
  - Tegyünk úgy, mintha minden eredményt még egyszer látnánk, mint amennyiszer valójában



$$P_{LAP}(x) = \frac{c(x)+1}{\sum_x [c(x)+1]}$$

$$= \frac{c(x)+1}{N+|X|}$$

$$P_{ML}(X) =$$

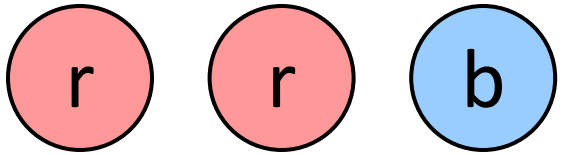
$$P_{LAP}(X) =$$

Laplace simítás

- Laplace becslés (kiterjesztett):
  - Tegyünk úgy, mintha minden eredményt láttunk volna k extra alkalommal

$$P_{LAP,k}(x) = \frac{c(x) + k}{N + k|X|}$$

  - k a prior erőssége



$$P_{LAP,0}(X) =$$

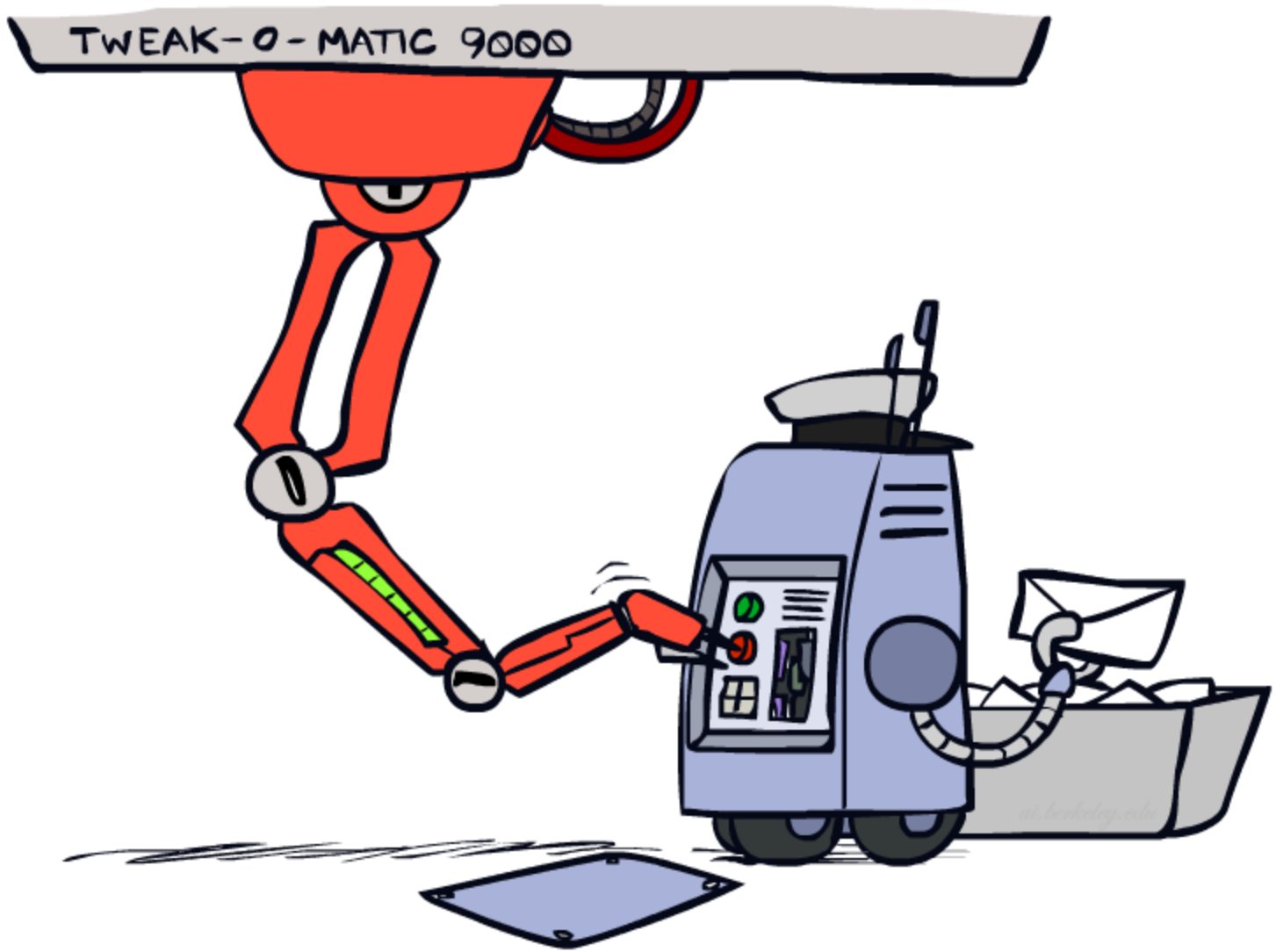
$$P_{LAP,1}(X) =$$

$$P_{LAP,100}(X) =$$

- Laplace feltételes valószínűségekhez :
  - Simítsuk az egyes feltételes valószínűségeket egymástól függetlenül:

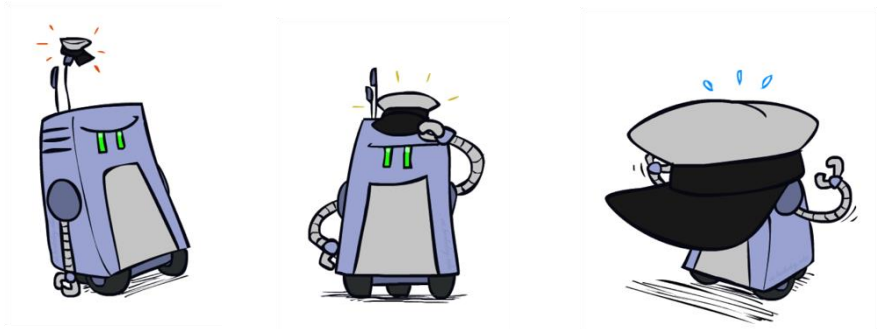
$$P_{LAP,k}(x|y) = \frac{c(x,y) + k}{c(y) + k|X|}$$

Tuning



Finomhangolás validációs adathalmazon

- Most kétféle ismeretlennel állunk szemben
  - Paraméterek: valószínűségekP(X|Y), P(Y)
  - Hiperparaméterek: Pl. a simítás összege / típusa: k, α
- Mit hol kell megtanulnunk?
  - Paraméterek megismerése tanító adatokból
  - Hiperparaméterek hangolása validációs adatokon
  - Válasszuk ki a legjobb értéket, és végezzünk végső tesztet a tesztadatokon



Kiértékelés – Baseline (alapmodell)

- Első lépés: hozzunk létre egy alapmodellt
  - Az alapmodellek nagyon egyszerű eljárások
  - Segítsenek meghatározni, hogy mennyire nehéz a feladat
  - Segítsen megtudni, mi a "jó" pontosság

- Gyenge baseline: Leggyakoribb címke alapú osztályozó
  - Az összes tesztpéldányt a tanító adathalmazban leggyakoribb címkével látja el
  - Például spamszűrés esetén mindent ham-nek címkézhet
  - A pontosság nagyon magas lehet, ha a probléma nem kiegyensúlyozott
  - Pl. ha mindent ham-nek hívunk, az 66%-ot kap, tehát az az osztályozó, amelyik 70%-ot kap, nem túl jó...
- A valódi kutatáshoz általában egy korábbi munkát használjuk (erős) kiindulási alapként