



---

# PLIN021 SÉMANTICKÁ ANALÝZA V PRAXI

ZUZANA NEVĚŘILOVÁ

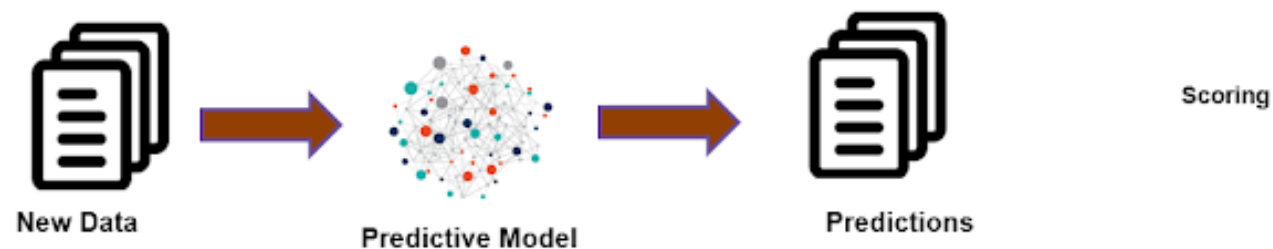
2020/21

# STROJOVÉ UČENÍ

- Učení = změna vnitřního stavu (k lepší schopnosti přizpůsobit se) na základě informací zvenčí
- **Strojové učení** = změna stavu počítačového systému tak, že zefektivní schopnost se přizpůsobit (zároveň se nemění kód počítačového programu)
- *Pokud je v blízkosti slova kočka i slovo pes, učím se, že jde o kočku ve významu 1.*

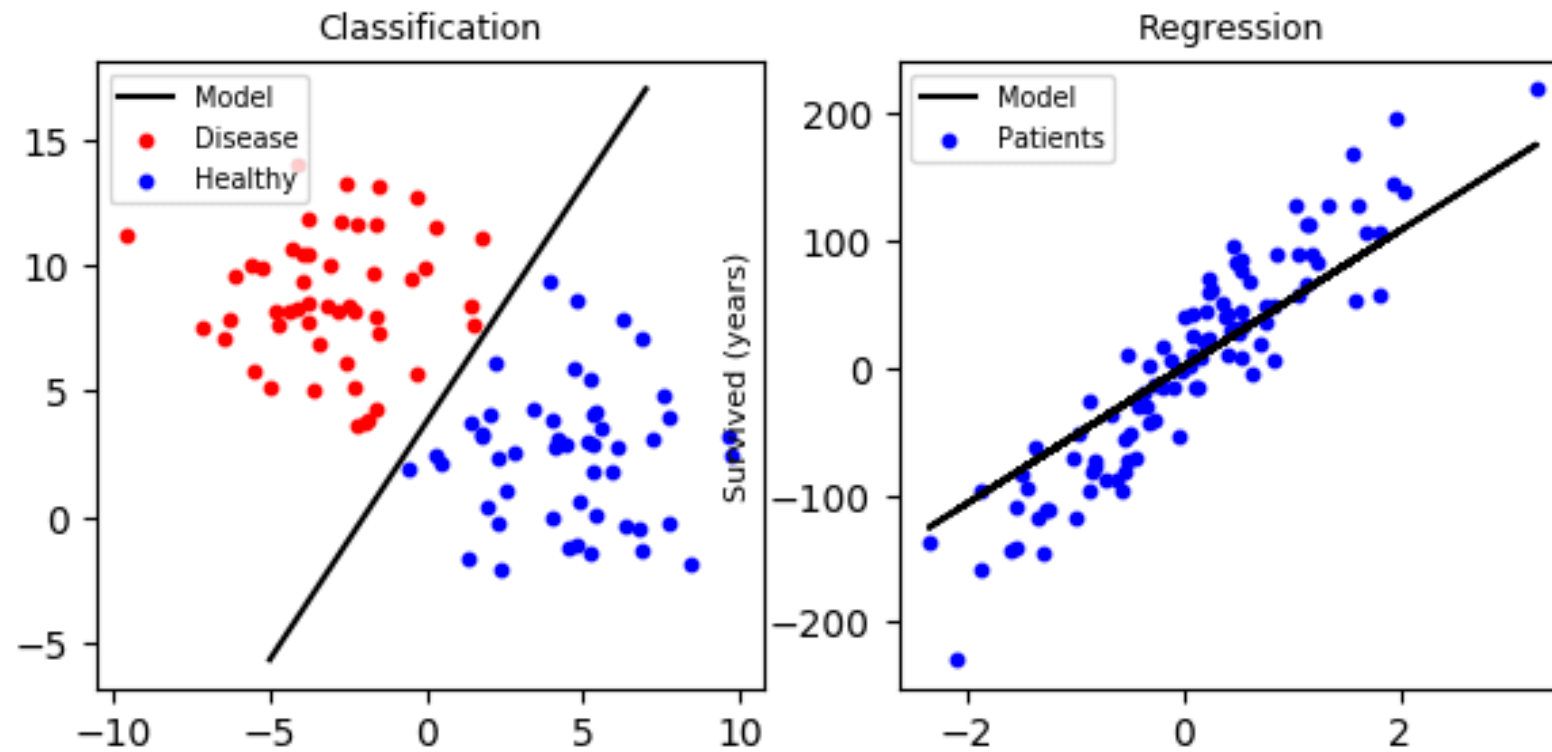
# STROJOVÉ UČENÍ

- **S učitelem** (supervised machine learning) – pro zadaný vstup máme i správný výstup (trénovací data)
- **Bez učitele** (unsupervised machine learning) – pro zadaný vstup neznáme správný výstup
- Kombinace – pro část dat máme trénovací data



<http://singaporebusinessintelligence.blogspot.com/2018/10/what-is-automated-machine-learning.html>

# STROJOVÉ UČENÍ: DISKRÉTNÍ A SPOJITÁ DATA



<https://mc.ai/between-classification-and-regression-how-to-know-which-is-which/>

# ROZHODOVACÍ SEZNAM

- `if (zvíře má chobot) then output(slon)`
- `if (zvíře má pruhy) then output(zebra)`
- `if (zvíře má ploutve & zvíře není ryba) then output(žralok)`



# ROZHODOVACÍ STROM

# NAIVNÍ BAYESOVSKÝ KLASIFIKÁTO R

## Historická pozorování

zvíře	velikost	barva	potrava
slon	velký	šedý	býložravec
kráva	velký	strakatý	býložravec
slon	střední	šedý	býložravec
kráva	velký	černý	býložravec
kráva	velký	bílý	býložravec
vlk	malý	šedý	masožravec
vlk	velký	černý	masožravec
kráva	střední	strakatý	býložravec

# NAIVNÍ BAYESOVSKÝ KLASIFIKÁTOR

## Historická pozorování

zvíře	velikost	barva	potrava
slon	velký	šedý	býložravec
kráva	velký	strakatý	býložravec
slon	střední	šedý	býložravec
kráva	velký	černý	býložravec
kráva	velký	bílý	býložravec
vlk	malý	šedý	masožravec
vlk	velký	černý	masožravec
kráva	střední	strakatý	býložravec

## Nové pozorování – Nepodmíněná

Aniž bychom **cokoli věděli**, můžeme spočítat, že pozorované zvíře je na 25 % vlk, na 25 % slon a 50 % kráva.

$$P(\text{slon}) = \frac{2}{8} \quad P(\text{vlk}) = \frac{2}{8} \quad P(\text{kráva}) = \frac{4}{8}$$

Poté, co víme, že jde o **velkého černého býložravce**, můžeme výpočet zpřesnit.



# NAIVNÍ BAYESOVSKÝ KLASIFIKÁTO R

## Historická pozorování

zvíře	velikost	barva	potrava
slon	velký	šedý	býložravec
kráva	velký	strakatý	býložravec
slon	střední	šedý	býložravec
kráva	velký	černý	býložravec
kráva	velký	bílý	býložravec
vlk	malý	šedý	masožravec
vlk	velký	černý	masožravec
kráva	střední	strakatý	býložravec

## Podmíněné pravděpodobnosti

$$P(\text{černý}|\text{slon}) = \frac{0}{2} \quad P(\text{černý}|\text{kráva}) = \frac{1}{4} \quad P(\text{černý}|\text{vlk}) = \frac{1}{2}$$

$$P(\text{velký}|\text{slon}) = \frac{1}{2} \quad P(\text{velký}|\text{kráva}) = \frac{3}{4} \quad P(\text{velký}|\text{vlk}) = \frac{1}{2}$$

$$P(\text{býložravec}|\text{slon}) = \frac{2}{2} \quad P(\text{býložravec}|\text{kráva}) = \frac{4}{4} \quad P(\text{býložravec}|\text{vlk}) = \frac{0}{2}$$

# NAIVNÍ BAYESOVSKÝ KLASIFIKÁTOR

## Nepodmíněné pravděpodobnosti

$$P(\text{slon}) = \frac{2}{8}$$

$$P(\text{vlk}) = \frac{2}{8}$$

$$P(\text{kráva}) = \frac{4}{8}$$

$$P(\text{velký}) = \frac{5}{8}$$

$$P(\text{černý}) = \frac{2}{8}$$

$$P(\text{býložravec}) = \frac{6}{8}$$

## Podmíněné pravděpodobnosti

$$P(\text{černý}|\text{slon}) = \frac{0}{2}$$

$$P(\text{černý}|\text{kráva}) = \frac{1}{4}$$

$$P(\text{černý}|\text{vlk}) = \frac{1}{2}$$

$$P(\text{velký}|\text{slon}) = \frac{1}{2}$$

$$P(\text{velký}|\text{kráva}) = \frac{3}{4}$$

$$P(\text{velký}|\text{vlk}) = \frac{1}{2}$$

$$P(\text{býložravec}|\text{slon}) = \frac{2}{2}$$

$$P(\text{býložravec}|\text{kráva}) = \frac{4}{4}$$

$$P(\text{býložravec}|\text{vlk}) = \frac{0}{2}$$

## Výpočet podle Bayesova vzorce

$P(\text{kráva}|\text{velký, černý, býložravec})$

$$= \frac{P(\text{kráva}) \cdot P(\text{černý}|\text{kráva}) \cdot P(\text{velký}|\text{kráva}) \cdot P(\text{býložravec}|\text{kráva})}{P(\text{velký, černý, býložravec})} =$$

$$= \frac{0.5 \cdot 0.25 \cdot 0.75 \cdot 1}{\frac{5}{8} \cdot \frac{2}{8} \cdot \frac{6}{8}} = 0.8 = 80\%$$

# NAIVNÍ BAYESOVSKÝ KLASIFIKÁTO R

## Nepodmíněné pravděpodobnosti

$$P(\text{slon}) = \frac{2}{8}$$

$$P(\text{velký}) = \frac{5}{8}$$

$$P(\text{vlk}) = \frac{2}{8}$$

$$P(\text{černý}) = \frac{2}{8}$$

$$P(\text{kráva}) = \frac{4}{8}$$

$$P(\text{býložravec}) = \frac{6}{8}$$

## Podmíněné pravděpodobnosti

$$P(\text{černý}|\text{slon}) = \frac{0}{2}$$

$$P(\text{velký}|\text{slon}) = \frac{1}{2}$$

$$P(\text{býložravec}|\text{slon}) = \frac{2}{2}$$

$$P(\text{černý}|\text{kráva}) = \frac{1}{4}$$

$$P(\text{velký}|\text{kráva}) = \frac{3}{4}$$

$$P(\text{býložravec}|\text{kráva}) = \frac{4}{4}$$

$$P(\text{černý}|\text{vlk}) = \frac{1}{2}$$

$$P(\text{velký}|\text{vlk}) = \frac{1}{2}$$

$$P(\text{býložravec}|\text{vlk}) = \frac{0}{2}$$

## Výpočet podle Bayesova vzorce

$P(\text{slon}|\text{velký, černý, býložravec})$

$$= \frac{P(\text{slon}) \cdot P(\text{černý}|\text{slon}) \cdot P(\text{velký}|\text{slon}) \cdot P(\text{býložravec}|\text{slon})}{P(\text{velký, černý, býložravec})} =$$

$$= \frac{0.25 \cdot 0 \cdot 0.5 \cdot 1}{\frac{5}{8} \cdot \frac{2}{8} \cdot \frac{6}{8}} = 0 = 0\%$$



# LITERATURA

- scikit-learn developers: 1.10. Decision Trees. 2007-2020. <https://scikit-learn.org/stable/modules/tree.html>
- Rohith Gandhi: **Naive Bayes Classifier**. Towards Data Science. 2018. <https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c>