

Deň otvorených dverí

Stratení v náhodnom lese. Príbeh o jednom machine-learningovom algoritme

Lukáš Lafférs

  UMBmath
www.umbmath.sk

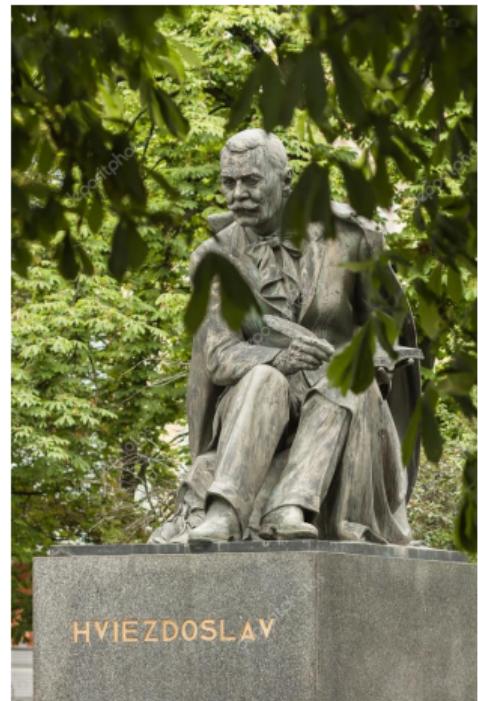
November 21, 2024

Náhodný les

Náhodný les

Pozdravujem vás, lesy, hory,
z tej duše pozdravujem vás!

P. O. Hviezdoslav

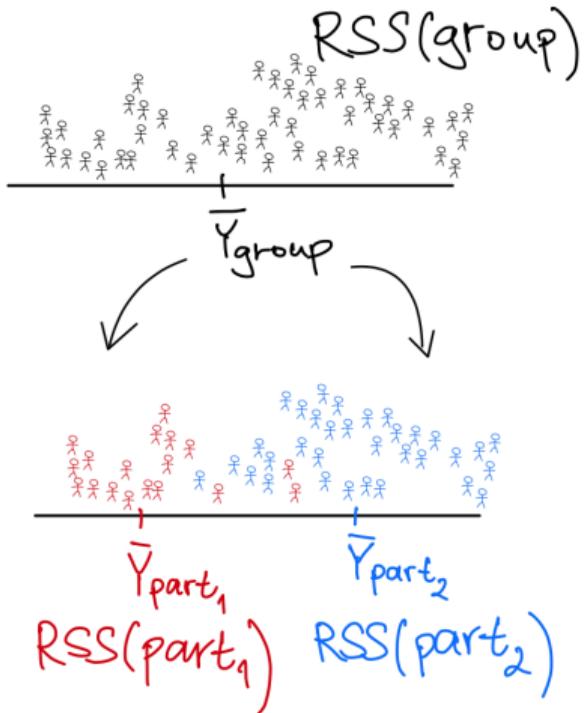
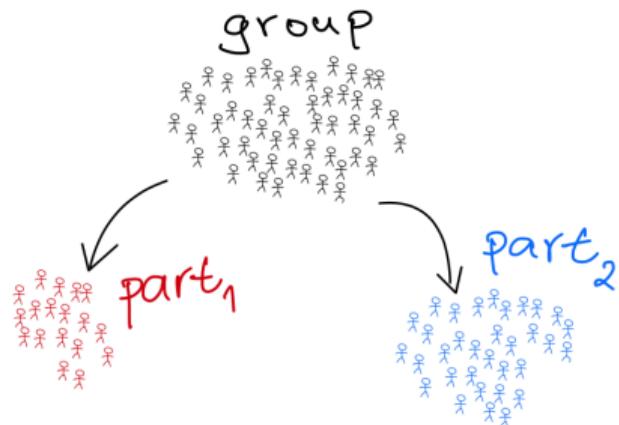


Prídeť k lekárovi a ten sa vás pýta kadejaké áno/nie otázky.

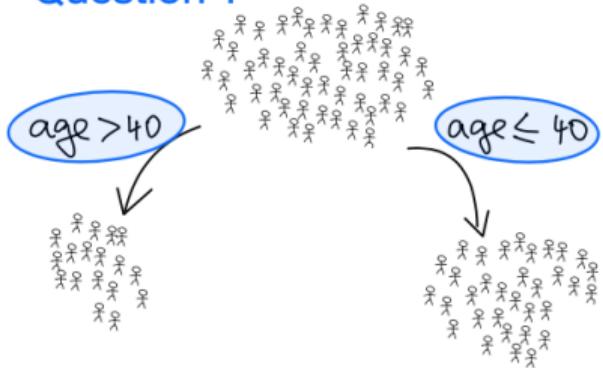
Na základe toho by chcel vedieť ako na tom ste.

Otázky musí klášť múdro.

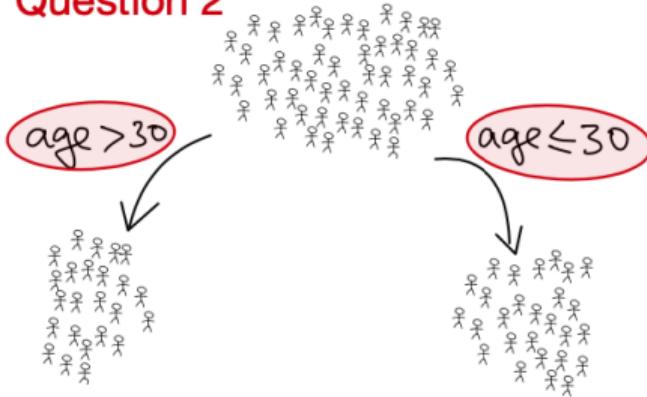
Čo je to múdra otázka?



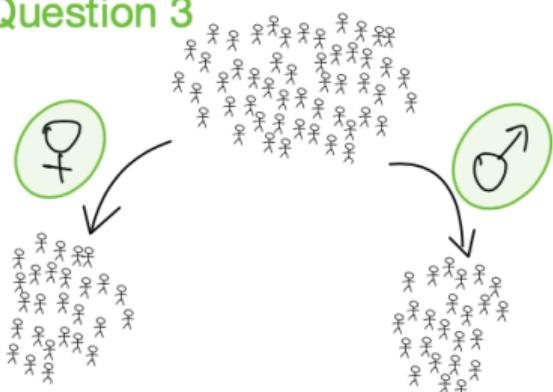
Question 1



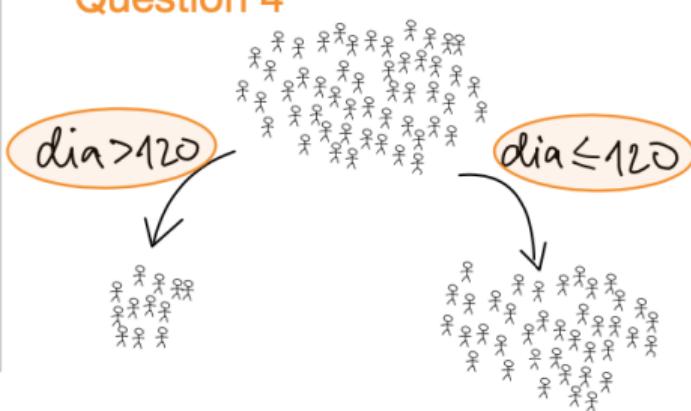
Question 2



Question 3



Question 4



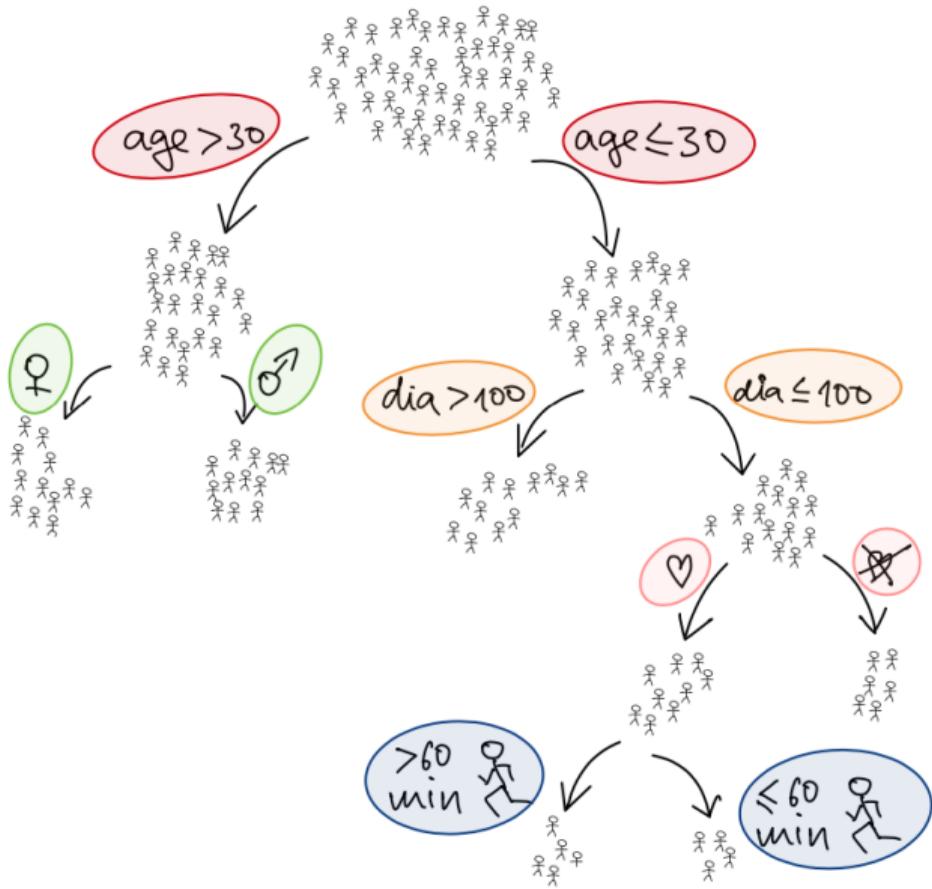
- Máme mnoho spôsobov ako rozdeliť strom.

- Máme mnoho spôsobov ako rozdeliť strom.
- Aký veľký strom?

- Máme mnoho spôsobov ako rozdeliť strom.
- Aký veľký strom?
 - Zastaň ak sú pozorovania v **listoch** dosť podobné

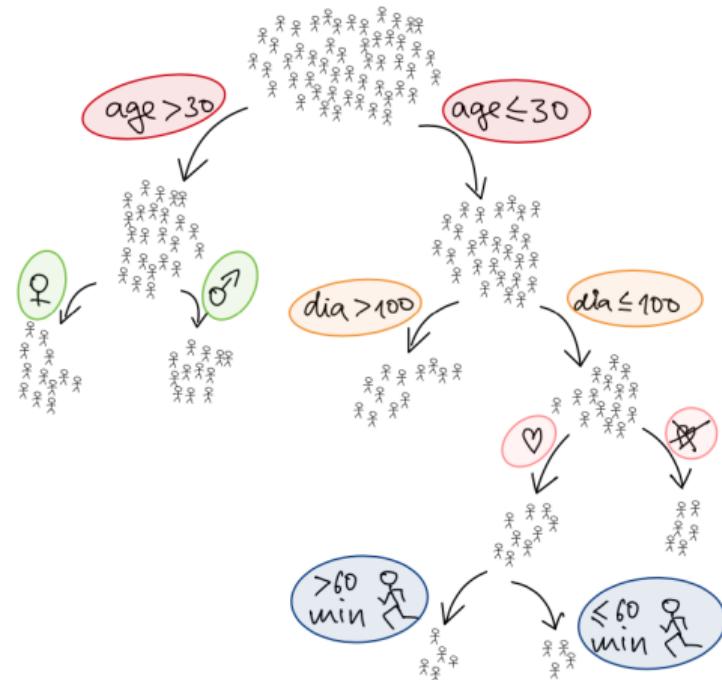
- Máme mnoho spôsobov ako rozdeliť strom.
- Aký veľký strom?
 - Zastaň ak sú pozorovania v **listoch** dosť podobné
 - Nariast' veľký strom a osekaj **vetvičky**

- Máme mnoho spôsobov ako rozdeliť strom.
- Aký veľký strom?
 - Zastaň ak sú pozorovania v **listoch** dosť podobné
 - Nariast' veľký strom a osekaj **vetvičky**
 - Vyber strom, ktorý balansuje **predikčnú schopnosť** a **zložitosť**



Ono je to vlastne regresia:

$$\begin{aligned}Y &= \beta_1 I(\text{age} > 30) \cdot I(\text{gender} = \text{♀}) \\&+ \beta_2 I(\text{age} > 30) \cdot I(\text{gender} = \text{♂}) \\&+ \beta_3 I(\text{age} \leq 30) \cdot I(\text{dia} > 100) \\&+ \beta_4 I(\text{age} \leq 30) \cdot I(\text{dia} \leq 100) \\&+ \beta_5 I(\text{age} \leq 30) \cdot I(\text{dia} \leq 100) \cdot I(\heartsuit) \cdot I(\text{sport} > 60) \\&+ \beta_6 I(\text{age} \leq 30) \cdot I(\text{dia} \leq 100) \cdot I(\heartsuit) \cdot I(\text{sport} \leq 60) \\&+ \beta_7 I(\text{age} \leq 30) \cdot I(\text{dia} \leq 100) \cdot I(\times) \\&+ \varepsilon\end{aligned}$$



Predpoved' v koncovom **liste** je jednoduchý priemer.

VÝHODY

- Stromy sú pekné a jednoduché.
- Nerobíme silné predpoklady.
- Vieme zachytiať zložité súvzťažnosti

VÝHODY

- Stromy sú pekné a jednoduché.
- Nerobíme silné predpoklady.
- Vieme zachytiať zložité súvzťažnosti

NEVÝHODY

- Citlivé ak je mnoho pozorovaní okolo bodov rozdelenia.
- Sú variabilné. Malá zmena vie zmeniť celý strom.

Náhodný les

Jeden strom je príliš nestabilný.

Náhodný les

Jeden strom je príliš nestabilný.

Necháme narásť mnoho stromov.

Náhodný les

Jeden strom je príliš nestabilný.

Necháme narásť mnoho stromov.

A urobíme priemer predpovedí.

Náhodný les

Jeden strom je príliš nestabilný.

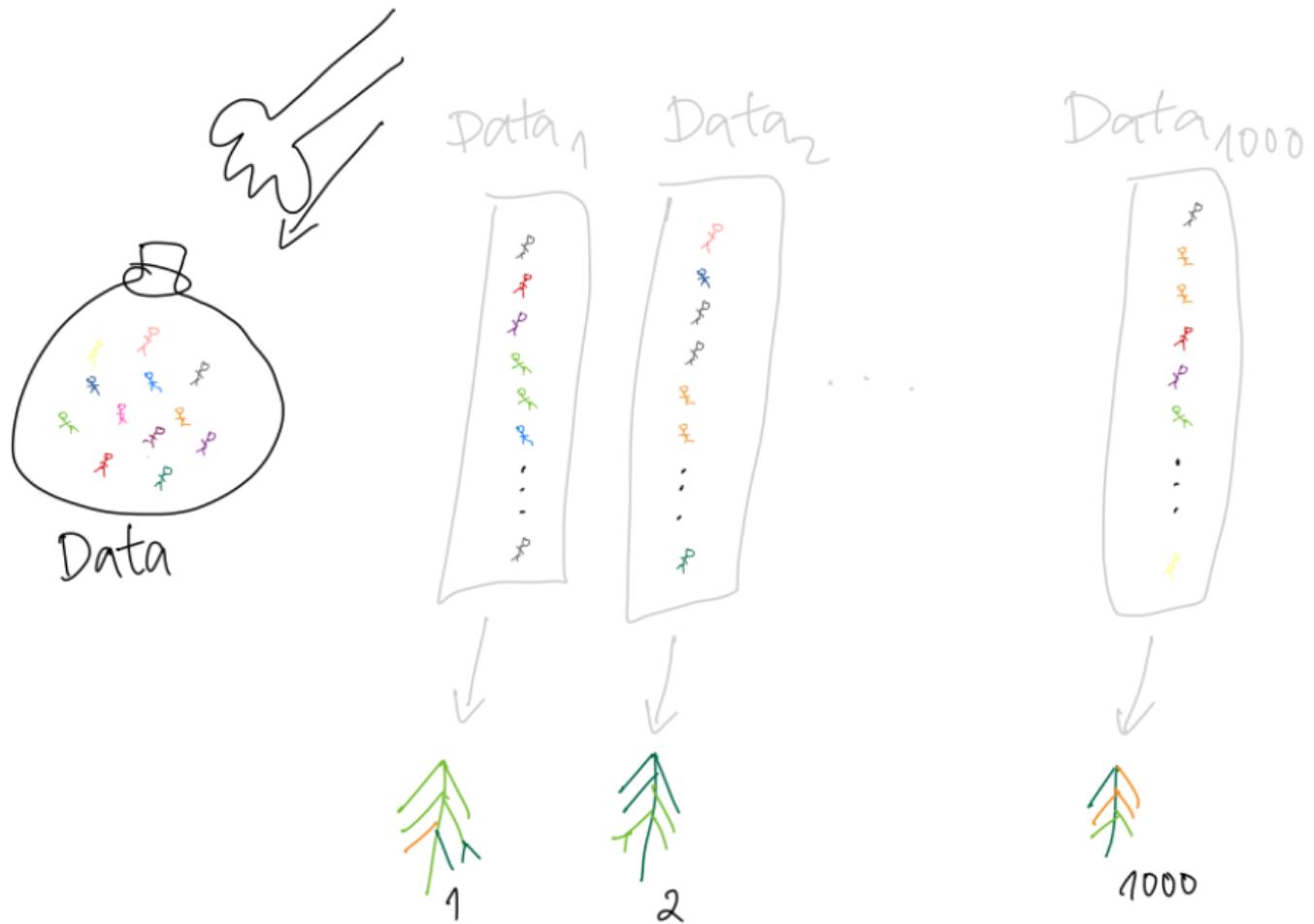
Necháme narásť mnoho stromov.

A urobíme priemer predpovedí.

Tie, ktoré lepšie predpovedajú urobíme dôležitejšie.



Náhodný les.



Les - veľa podobných stromov



Hm, nie až taký nahodný, že?

Náhodný les

Stromy sú rôzne lebo sú narastené z rôznych vzoriek.

Náhodný les

Stromy sú rôzne lebo sú narastené z rôznych vzoriek.

Môže urobiť stromy rôznorodé **nasilu**.

Náhodný les

Stromy sú rôzne lebo sú narastené z rôznych vzoriek.

Môže urobiť stromy rôznorodé **nasilu**.

Nepoužijeme všetky informácie na predpovedanie. Napr. si z nich náhodne vyberieme len \sqrt{p} namesto všetk p .

Náhodný les

Stromy sú rôzne lebo sú narastené z rôznych vzoriek.

Môže urobiť stromy rôznorodé **nasilu**.

Nepoužijeme všetky informácie na predpovedanie. Napr. si z nich náhodne vyberieme len \sqrt{p} namesto všetk p .

Koľko stromov? Nuž dosť na to aby sa predikcia už pridaním dalších nezlepšovala.



Random forest.

Náhodný les - veľa rôznych stromov



Efekt premennej

Hm, ako vyčíslime efekt premennej v tak komplikovanej veci ako je náhodný les?

Efekt premennej

Hm, ako vyčíslime efekt premennej v tak komplikovanej veci ako je náhodný les?

Nech nás zaujíma *vek*.

- **Partial dependence** - nastavíme $vek = 30$ a pozrieme sa na priemerné predpovede

Efekt premennej

Hm, ako vyčíslime efekt premennej v tak komplikovanej veci ako je náhodný les?

Nech nás zaujíma *vek*.

- **Partial dependence** - nastavíme $vek = 30$ a pozrieme sa na priemerné predpovede
- **Partial effects** - nastavíme ostatné premenné na priemerné hodnoty.

Efekt premennej

Hm, ako vyčíslime efekt premennej v tak komplikovanej veci ako je náhodný les?

Nech nás zaujíma *vek*.

- **Partial dependence** - nastavíme $vek = 30$ a pozrieme sa na priemerné predpovede
- **Partial effects** - nastavíme ostatné premenné na priemerné hodnoty.
- **Importance** - Ako sa nám zmení predikčná chyba ak náhodne zamiešame *vek*.

Efekt premennej

Hm, ako vyčíslime efekt premennej v tak komplikovanej veci ako je náhodný les?

Nech nás zaujíma *vek*.

- **Partial dependence** - nastavíme *vek* = 30 a pozrieme sa na priemerné predpovede
- **Partial effects** - nastavíme ostatné premenné na priemerné hodnoty.
- **Importance** - Ako sa nám zmení predikčná chyba ak náhodne zamiešame *vek*.
- **Minimal depth** - Ako často a ako hlboko sa používa *vek* pri raste stromov?

História

- Salzberg and Heath 1993
- Tin Kam Ho 1995
- Breiman 2006

Účinnosť liečby



Detekcia SPAMu



Výkon športovcov



Výnos akcií



Otázky?

- `lukas.laffers@umb.sk`
- Zoom/Teams
- osobne [FPV UMB], F226A

Táto prezentácia bude na [<https://www.lukaslaffers.com/misc>].

štúdijný program: **Matematika (Bc)**

štúdijný program: **Matematika v analýze dát a vo financiách (Mgr)**

Prečo matematika?

- → uplatnenie
- analytické myslenie
- precíznosť
- jazykom prír. vied
- zábava



Prečo na UMB?

- →moderný program
- práca so softvérom
- špičková veda
- rodinné prostredie
- Banská Bystrica



Uplatnenie

- data science
- škola/veda
- fin. matematika
- štatistika
- programovanie



Čo sa naučíte?

- rozmanité oblasti matematiky
- pracovať analyticky
- dátovú analýzu
- základy programovania

Naši absolventi

- Národná Banka Slovenska
- analytické útvary (ČSOB, VÚB)
- poisťovne
- softvérové firmy
- technologické firmy (Dell, IBM)
- vládne think-tanky (Inštitút sociálnej politiky)

Katedra matematiky

- v akreditácii hodnotenie A
- špičkový vedecký tím
- spolupráca so zahr. inštitúciami (Oxford, Londýn, Sydney)

Ďakujem za pozornosť

- www.umbmath.sk
- lukas.laffers@umb.sk
- [f](https://www.facebook.com/umbmath) [@ umbmath](https://www.instagram.com/umbmath)

Ďakujem za pozornosť!



KATEDRA MATEMATIKY
UNIVERZITA MATEJA BELA



UMBmath
umbmath