



T | Y | D | E | N | A | V

TÝDEN AKADEMIE VĚD ČESKÉ REPUBLIKY

1–7/11/2021

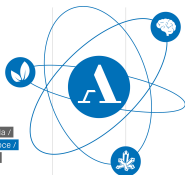
WWW.TYDENAVCR.CZ



/ věda /

/ v roeb /

/ 21



Statistické uvažování a jeho role při klinickém rozhodování

Jan Kalina

Ústav informatiky AV ČR, v.v.i.



“Data don't make any sense,
we will have to resort to statistics.”

Biomedicínský výzkum

- Začíná se medicínskou hypotézou
- Návrh experimentu
 - Spolupráce se statistikem na návrhu experimentu, volbě metod, analýze dat, interpretace
- Výsledky jsou zjištěny statisticky, platí v průměru (pro průměrného pacienta)

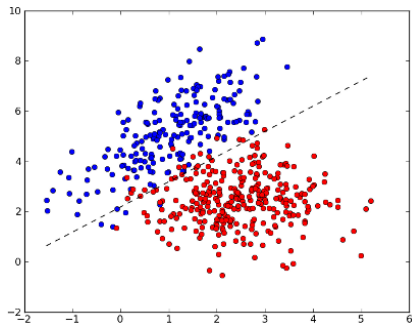
Příklad #1: Studie genových expresí

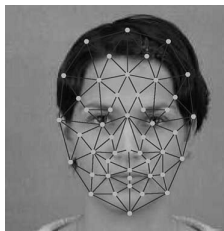
- Centrum biomedicínské informatiky
- Diagnostika kardiovaskulární onemocnění

Tabulka (zlogaritmovaných) hodnot genových expresí:

	Gene	24 pacientů s mrtvicí				24 kontrolních osob			
		# 1	#2	...	#24	# 1	#2	...	#24
1	ADORA3	5.82	6.04	...	5.99	5.71	6.12	...	6.09
2	CPD	3.53	4.08	...	2.32	4.21	5.01	...	4.66
3	ECHDC3	2.50	2.71	...	3.17	2.99	3.52	...	3.01
4	VNN3	3.38	3.03	...	4.59	4.56	3.98	...	4.70
5	IL18RAP	4.03	4.91	...	5.81	5.12	5.01	...	5.23
6	ERLIN1	5.76	4.38	...	4.90	6.49	5.02	...	6.18
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38 590	PHACTR1	5.21	4.99	...	5.06	5.15	5.53	...	5.20

- Cílem je sestavit klasifikační pravidlo do dvou skupin
- Vysoce dimenzionální data vyžadují specifické metody
- Redukce dimenzionality (selekce proměnných)

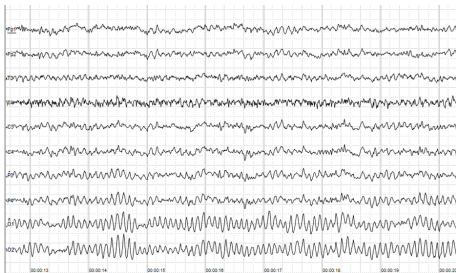
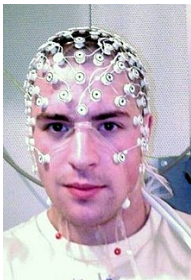


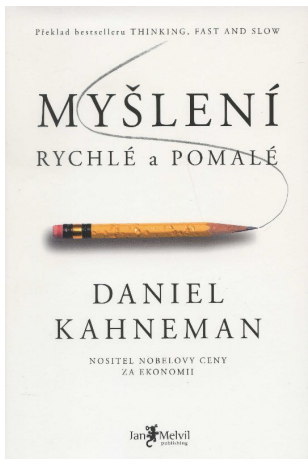


- Cíl: automatické určení diagnózy pro 10 genetických syndromů
- Je třeba začít automatickou lokalizací obličejů
- Deformace šablony
- Jak tuto deformaci měřit?
- Klasifikace genetických onemocnění: 147 obrázků pacientů
- Systém pro podporu rozhodování: klasifikační správnost 75 %

Příklad #3: Výzkum mozku

- Data z Fakultní nemocnice Hradec Králové
 - EEG signály (elektroencefalografie)
 - 28 pacientů s Alzheimerovou demencí
 - 146 kontrolních osob
 - Data naměřena v 19 kanálech a 46 frekvencích (rozklad signálu na vlny o různých frekvencích)
- 1 Kalina J., Kukul J., Vyšata O. (2021): Multivariate tests for comparing EEG signals (with multiple comparisons). V přípravě.





- Kritika lidského rozhodování
- Podmíněné pravděpodobnosti, Bayesův vzorec

- Diagnostický test pro danou nemoc
- 2 % celé populace mají danou nemoc
- Člověk s danou nemocí má pravděpodobnost pozitivního testu 95 %
- Člověk, který danou nemoc nemá, má pravděpodobnost negativního testu 90 %
- Jdu na test a výsledek je pozitivní
- Jaká je pravděpodobnost, že mám danou nemoc?

- Diagnostický test pro danou nemoc
- 2 % celé populace mají danou nemoc
- Člověk s danou nemocí má pravděpodobnost pozitivního testu 95 %
- Člověk, který danou nemoc nemá, má pravděpodobnost negativního testu 90 %
- Jdu na test a výsledek je pozitivní
- Jaká je pravděpodobnost, že mám danou nemoc? 16 %

- Diagnostický test pro danou nemoc
- 2 % celé populace mají danou nemoc
- Člověk s danou nemocí má pravděpodobnost pozitivního testu 95 %
- Člověk, který danou nemoc nemá, má pravděpodobnost negativního testu ~~90%~~ 60 %
- Jdu na test a výsledek je pozitivní
- Jaká je pravděpodobnost, že mám danou nemoc? ~~16%~~

- Diagnostický test pro danou nemoc
- 2 % celé populace mají danou nemoc
- Člověk s danou nemocí má pravděpodobnost pozitivního testu 95 %
- Člověk, který danou nemoc nemá, má pravděpodobnost negativního testu ~~90 %~~ 60 %
- Jdu na test a výsledek je pozitivní
- Jaká je pravděpodobnost, že mám danou nemoc? ~~16 %~~ 5 %

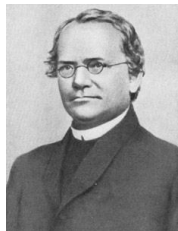
- Teorie pravděpodobnosti: vznik pro hazardní hry (17. století)
- Thomas Bayes (1702–1761): podmíněné pravděpodobnosti
- Adolphe Quetelet (1796–1874): průměrný člověk
- Experimenty v biologii (Gregor Mendel, 1822–1884)
- Teorie navrhování experimentů: Fisher (1936)
- Klasifikační metody: Fisher (od 1936)



Bayes



Quetelet



Mendel

- 1 Kalina J., Soukup L. (2019): Průkopníci statistiky ve vědách o člověku v 19. století. Informační bulletin České statistické společnosti 30(3), 1–15.

- Nízká kvalita dat
- Kontaminace dat (šum, chyby měření, odlehlé hodnoty)
- Vstupní data jsou opravdu surová (neupravená)?
- Je třeba používat mnohorozměrné metody
- Jak najít nejvhodnější metodu pro analýzu konkrétních dat
- Statistická analýza vyžaduje zkušenosti
- Modely jsou vždy jen přibližné, statistické metody mají své předpoklady a svá omezení
- Výsledky je potřeba interpretovat
- Výsledky platí v průměru; korelace versus kauzalita



- Rozhodování: klasifikační úloha za nejistoty
 - Nástroje umělé inteligence
 - Klinická medicína: diagnóza, terapie, prognóza
 - Rozličné typy a formáty dat
 - Trénování \implies validace na nezávislých datech
 - Uživatelská přívětivost
- 1 Kalina J. (2021): Mental health clinical decision support exploiting big data. In: Research Anthology on Mental Health Stigma, Education, and Treatment. IGI Global, 341–359.
 - 2 Kalina J., Tobišková N. (2021): Ethical aspects of information-based medicine (with a focus on mental health). In: Ethical Implications of Reshaping Healthcare with Emerging Technologies. IGI Global, 42–70.

- 1 Onari M.A., Yousefi S., Rabieepour M., Alizadeh A., Rezaee M.J. (2021): A medical decision support system for predicting the severity level of COVID-19. *Complex Intelligence Systems* 2021, v tisku.

Predikce závažnosti onemocnění COVID-19 podle rizikových faktorů jednotlivců.

- 2 Jahn B., Sroczynski G., Bicher M. (2021): Targeted COVID-19 vaccination (TAV-COVID) considering limited vaccination capacities—An agent-based modeling evaluation. *Vaccines* 9 (5), 434.

Výběr jednotlivců pro očkování při omezených očkovacích možnostech, strategie pro prioritizaci pacientů pro rakouskou populaci.

- 3 Aggarwal L., Goswami P., Sachdeva S. (2021): Multi-criterion intelligent decision support system for COVID-19. *Applied Soft Computing* 101, 107056.

Prioritizace pacientů, plánování očkování, příprava na zhoršení epidemiologické situace.

- 4 Karaarslan E., Aydın D. (2021): An artificial intelligence-based decision support and resource management system for COVID-19 pandemic. *Data Science for COVID-19*, Academic Press, San Diego, 25–49.

Epidemiologický manažerský systém (EMS). Management karantény (monitorování pohybu nakažených) nebo zdrojů (nemocniční lůžka, ventilátory).

Medicína založená na důkazech

- Statistika vyhodnocuje experimenty v biomedicínském výzkumu
- (Statistická) evidence
- Výsledky formulovány pro průměrného pacienta

Postupy (medicína, veřejné zdravotnictví) založené na informacích

- Informační technologie (umělá inteligence)
 - Systémy pro podporu rozhodování (využití statistiky, strojového učení)
 - Statistická gramotnost
 - Synergie umělé inteligence a lékaře
- 1 Kalina J. (2022): Pandemic-driven innovations contribute to the development of information-based medicine. In: Digital Innovation for Healthcare in COVID-19 pandemic—Strategies and Solutions. Elsevier, kap. 15. V tisku.

TÝDEN VĚD

TÝDEN AKADEMIE VĚD
ČESKÉ REPUBLIKY

1–7/11/2021

WWW.TYDENAVCR.CZ



7. věda /
7. v roce /
7.21

