



PLIN037 Sémantika a počítače

Zuzana Nevěřilová
2024

Vyhodnocení

Vnitřní (Intrinsic):

- jaké skóre model dosahuje?
- porovnání s modely v „laboratorních“ podmínkách
- reproducibilita

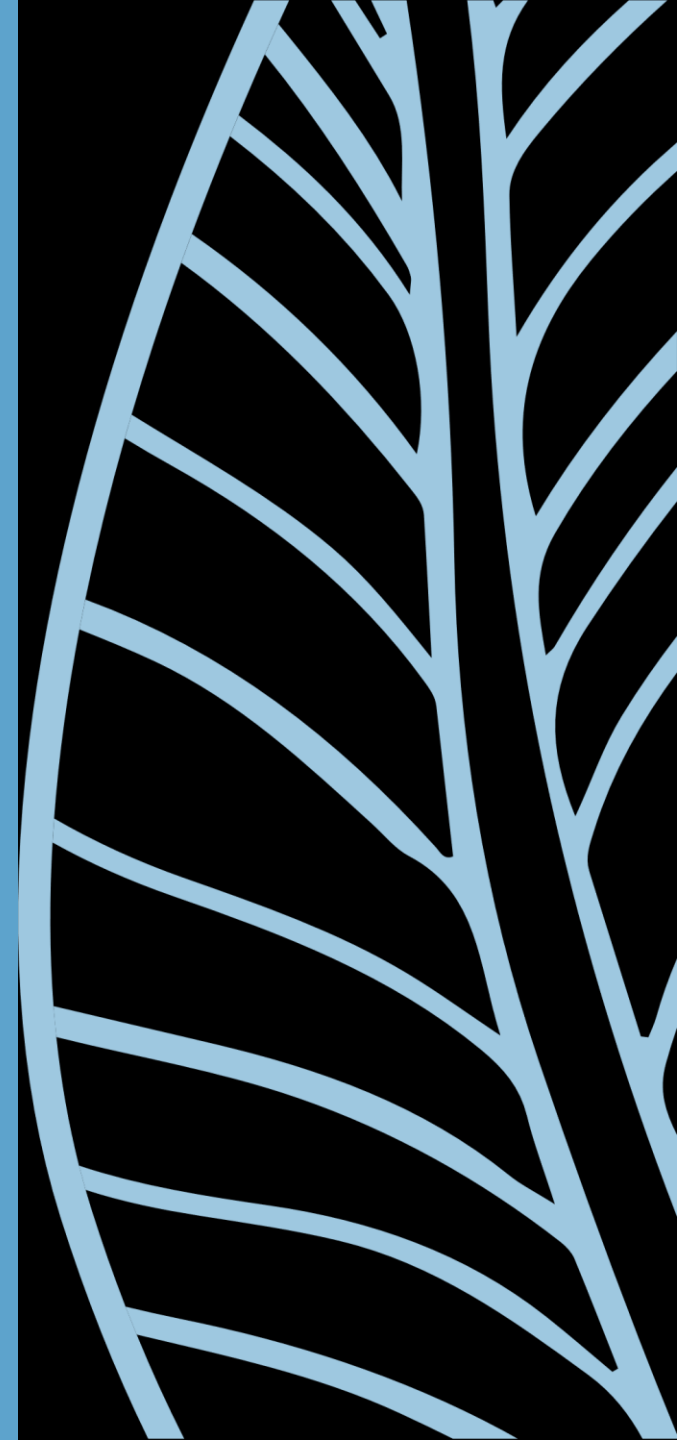
Vnější (Extrinsic):

- jak užitečný je model pro jinou úlohu?
- jak dobrý je model v praxi?
- porovnání s předchozím stavem
- nižší cena, vyšší bezpečnost, vyšší spokojenost uživatelů, ...

experiment

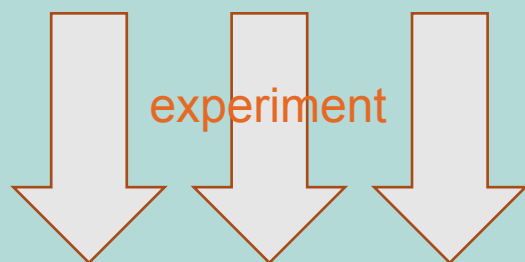


business



Kvantitativní a kvalitativní vyhodnocení

Pozorování → Hypotéza



Jak daleko je hypotéza od pozorování?

Proč a jaký to má dopad?

Spojitá data:

- Vzdálenost
- Podobnost (opačná hodnota)

Diskrétní data:

- Přesnost, pokrytí
- Jsou všechny kategorie vzdáleny stejně?

Statistika (připomenutí):

Nominální, ordinální, intervalová, poměrová data

Spojité data: vzdálenost a metrika

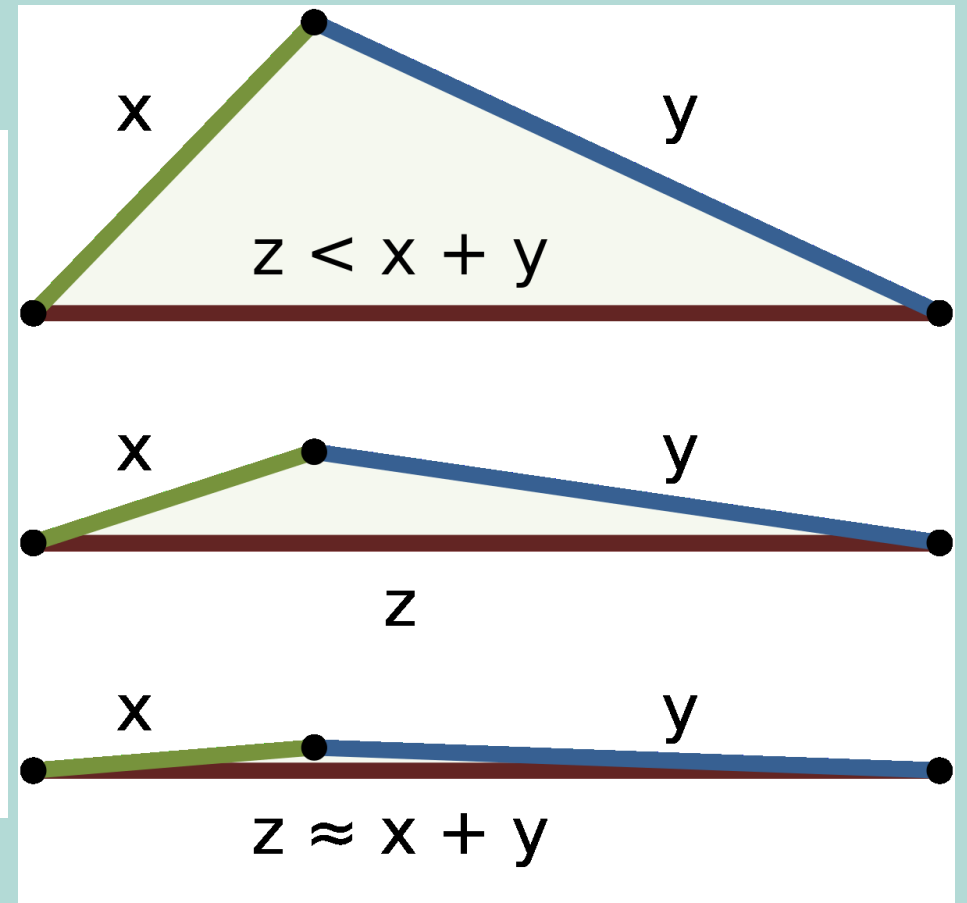
$D: X \times X \rightarrow \mathbb{R}, \forall x, y: D(x, y) \geq 0$

Metrika:

$D(x, y) \geq 0 \leftrightarrow x = y$ (**identita**)

$D(x, y) + D(y, z) \geq D(x, z)$ (**trojúhelníková nerovnost**)

$D(x, y) = D(y, x)$ (**symetrie**)



https://en.wikipedia.org/wiki/Triangle_inequality#/media/File:Triangleinequality.svg

Spojité data: vzdálenost bodů

Triviální diskrétní: $D(x, y) = 0 \leftrightarrow x = y,$
 $D(x, y) = 1 \leftrightarrow x \neq y$

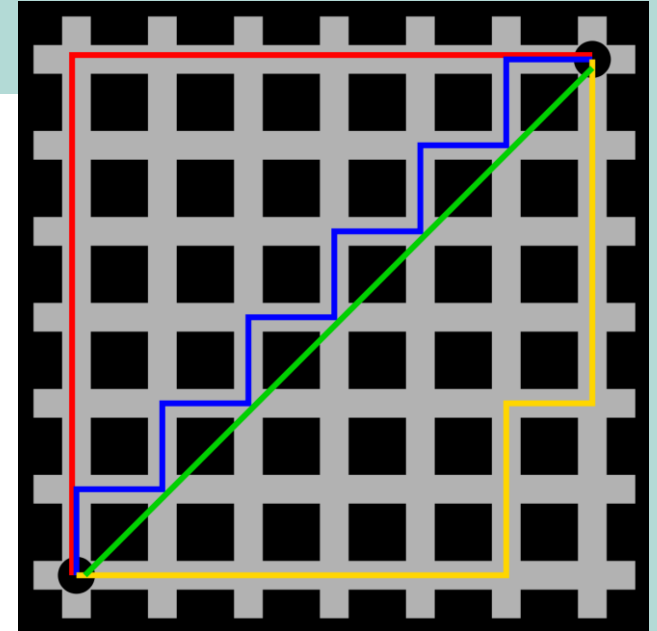
Euklidovská 1D: $D(x, y) = \sqrt{(x - y)^2}$

Euklidovská 2D: $D(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$

Euklidovská nD: $D(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$

Euklidovská kvadratická: $D(x, y) = (x_1 - y_1)^2 + (x_2 - y_2)^2$

Manhattanská: $D(x, y) = |x_1 - y_1| + |x_2 - y_2|$



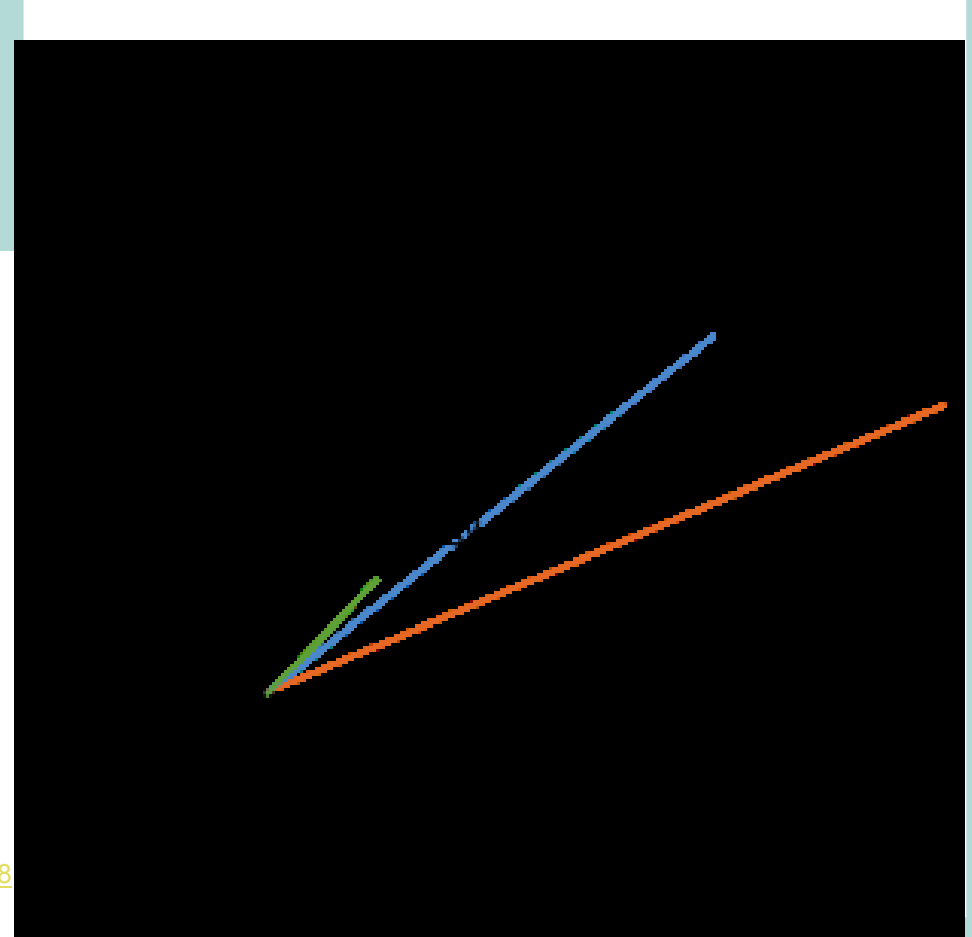
https://en.wikipedia.org/wiki/Taxicab_geometry

Spojité data: vektorová podobnost

	Math	Philosophy	Psychology
Theory	80	50	15
Harmony	45	60	20

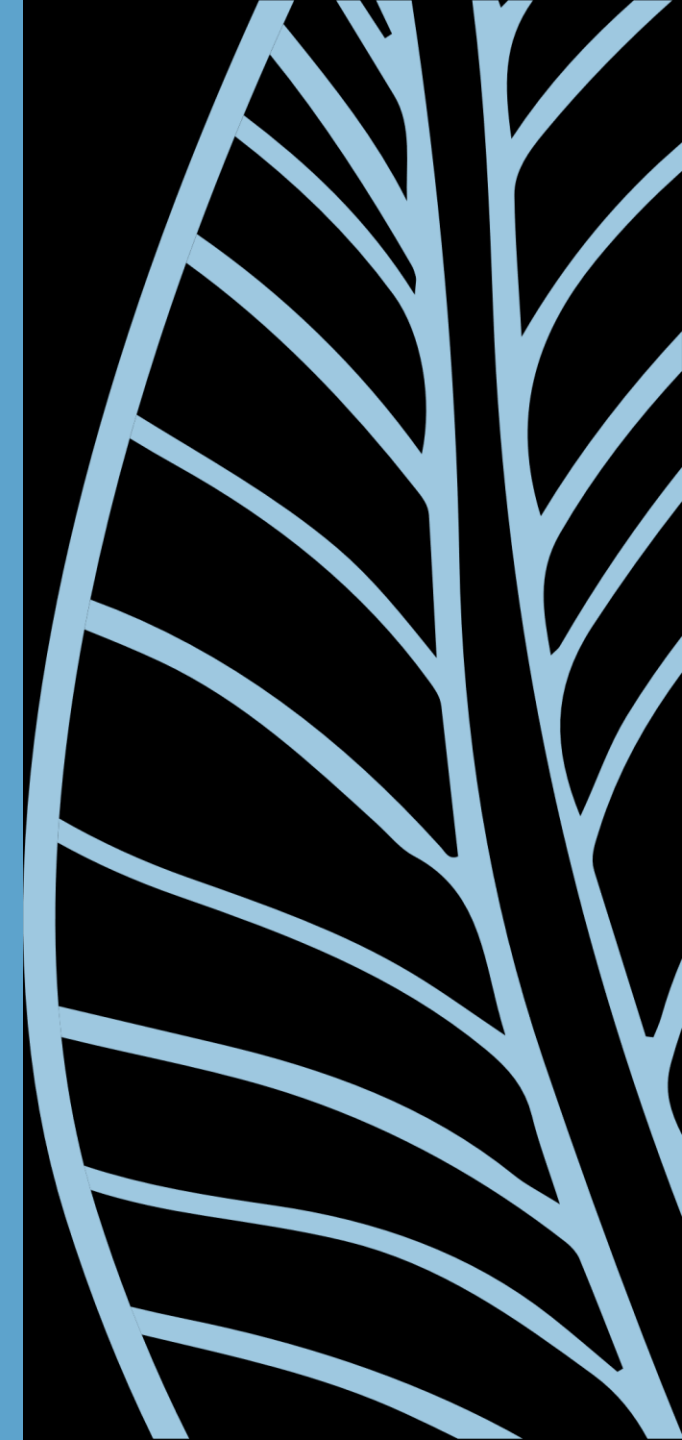
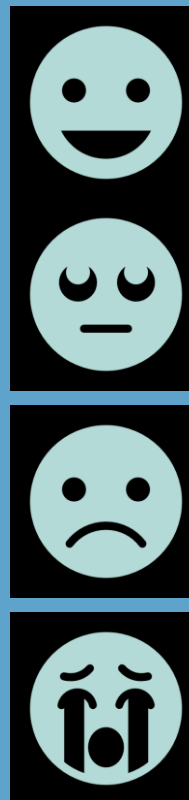
Kosinová podobnost: $\cos(\theta) = \frac{V \cdot W}{|V| \cdot |W|} = \frac{v_1 \cdot w_1 + v_2 \cdot w_2}{\sqrt{(v_1 + v_2)^2} \cdot \sqrt{(w_1 + w_2)^2}}$

<https://medium.com/swlh/euclidean-distance-and-cosine-similarity-which-one-to-use-and-when-28c97a18fe68>



Diskrétní data

- **Když model predikuje hodnotu, kterou jsme chtěli.**
- **Když model nepredikuje hodnotu, kterou jsme chtěli.**
 - Je to stejně špatně, jako by predikoval náhodnou hodnotu?
 - Je možnost, že je predikce méně špatně než jiná predikce?
 - Je možnost, že je predikce i tak užitečná?
 - Je možnost, že predikce je horší než náhodná hodnota?
 - Bylo by lepší, kdyby model řekl „nevím“, pokud si není jistý ((sebe)jistota = confidence)?



Diskrétní data: Editační vzdálenost řetězců

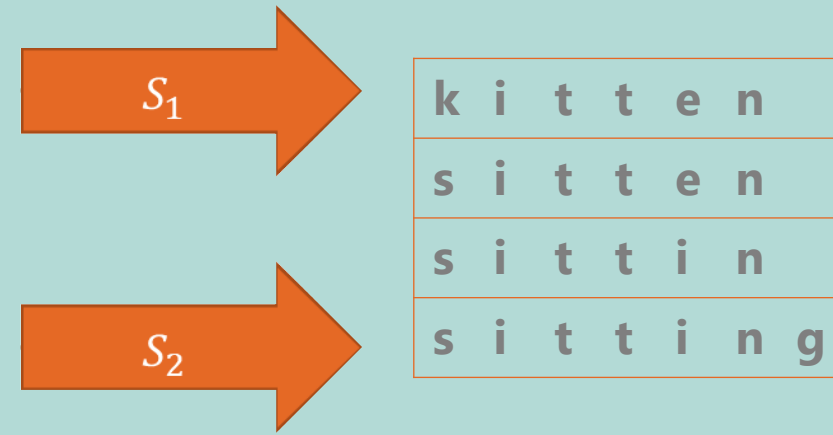
Kolik nejméně transformací potřebujeme, abychom vytvořili S_2 z S_1 ?

- Přidat znak
- Odebrat znak
- Nahradit znak
- (vyměnit dva znaky)

Levenshteinova vzdálenost (Levenshtein distance)

Damerau-Levenshteinova vzdálenost

**Hammingova vzdálenost (Hamming distance)
non-matching characters**



Diskrétní data: vzdálenost množin

Jaccardův koeficient (Jaccard index)

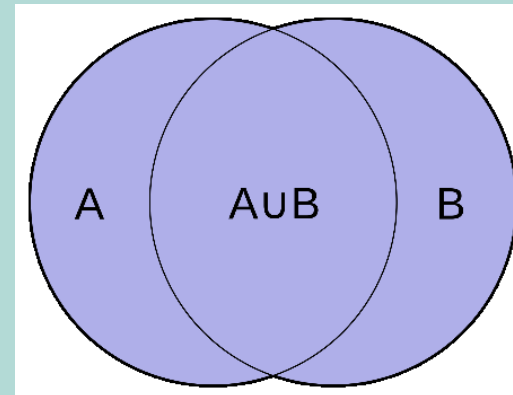
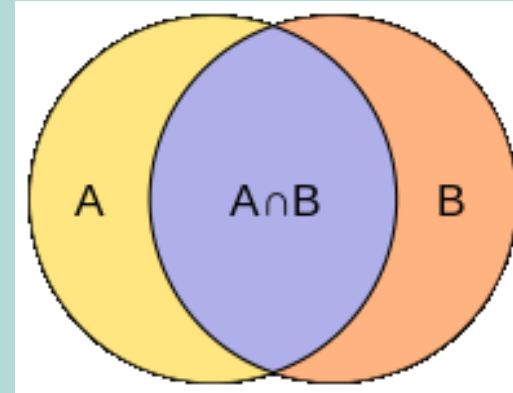
$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Jaccardova vzdálenost (Jaccard distance)

$$d(A, B) = 1 - J(A, B)$$

Sørensenův–Diceův koeficient (Sørensen–Dice coefficient)

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$



Diskrétní data: Editační vzdálenost stromů

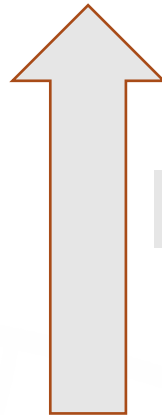
Operace:

Přidání uzlu n

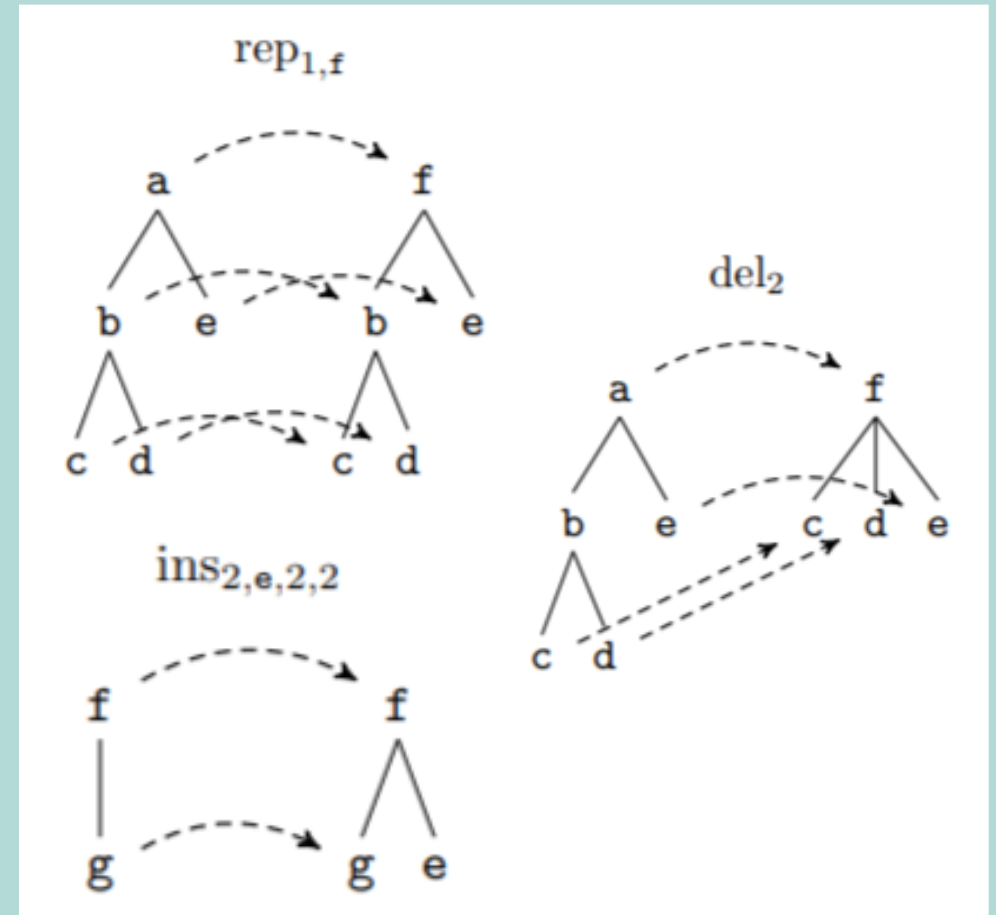
Odebrání uzlu n
(+ přepojení podstromu n na rodičovský uzel n)

Přejmenování uzlu n

Spočítat editační vzdálenost stromů je NP těžké.



rekurze



Porovnání s „pravdou“: binární data

Binární klasifikace: rozpoznání jmen osob (Named Entity Recognition)

April	Jackson	comes	from	Jackson	,	Mississippi	.
1	1	0	0	0	0	0	0
0	1	0	0	1	0	0	0

Kolik jmen model rozpoznal?

Kolik jmen model nenašel?

Kolik slov, která nebyla jmény, model označil jako jména?

Binární data

April	Jackson	comes	from	Jackson	,	Mississippi	.
1	1	0	0	0	0	0	0
0	1	0	0	1	0	0	0



Pravda	Predikce	Souhlasí?
1	0	no
1	1	yes
0	0	yes
0	0	yes
0	1	no
0	0	yes
0	0	yes
0	0	yes

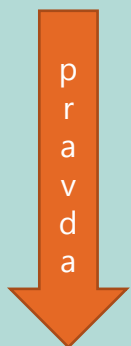
Matice záměn (Confusion matrix)

	1	0
1	1	1
0	1	5



Matrice Záměn Confusion Matrix

TP	1
TN	5
FP	1
FN	1



	1	0
1	1	1
0	1	5

← predikce

Typy chyb

Přesnost (Precision): $P = \frac{TP}{TP+FP}$

Pokrytí (Recall): $R = \frac{TP}{TP+FN}$

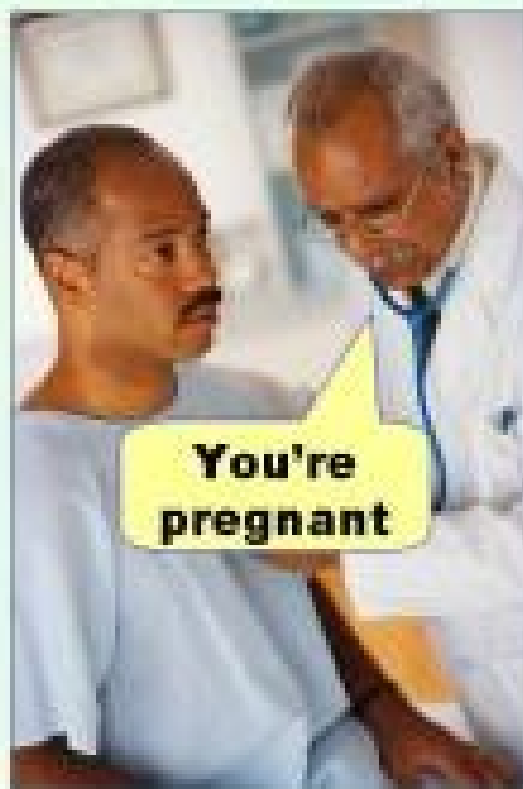
(Míra F1) F1 score: $F_1 = \frac{2PR}{P+R}$

Připomínka

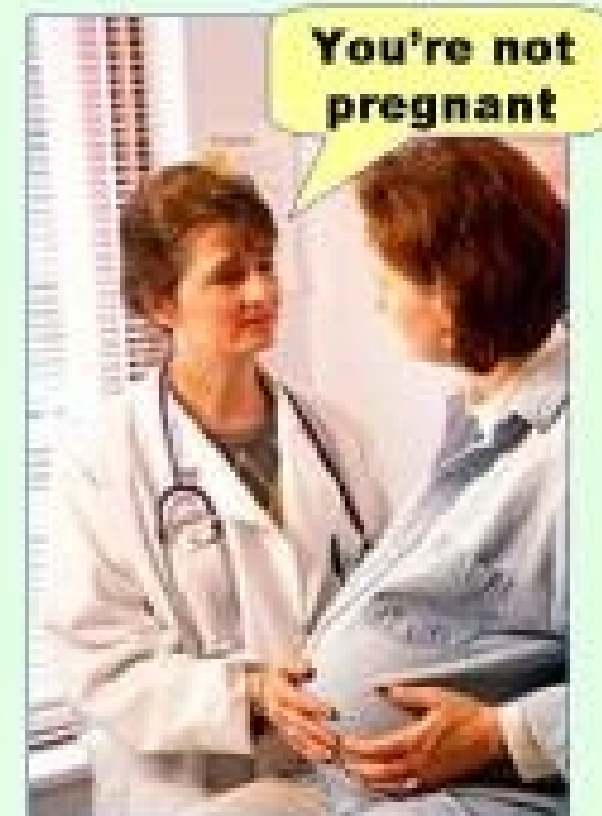
Sørensenův–Diceův koeficient:

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

Type I error
(false positive)



Type II error
(false negative)



<https://www.playinglean.com/blogs/playing-lean-blog/experiment-cards-under-the-magnifier-false-negatives-and-false-positives>

Od binární k n-ární klasifikaci

Matice záměn neobsahuje kladné a záporné hodnoty.

p
r
a
v
d
a

	Kočka	Pes	Myš
Kočka	1	1	1
Pes	2	4	2
Myš	1	7	3

predikce

TP, TN, FP, FN je třeba spočítat pro všechny třídy odděleně.

$$TP = 1$$

$$TN = 4 + 2 + 7 + 3 = 16$$

$$FP = 1 + 1 = 2$$

$$FN = 2 + 1 = 2$$

Od binární k n-ární klasifikaci

- Kočka – nekočka
- Pes – nepes
- Myš – nemyš

	Kočka	Pes	Myš
Kočka	1	1	1
Pes	2	4	2
Myš	1	7	3

$$\text{Precision: } P = \frac{TP}{TP+FP}$$

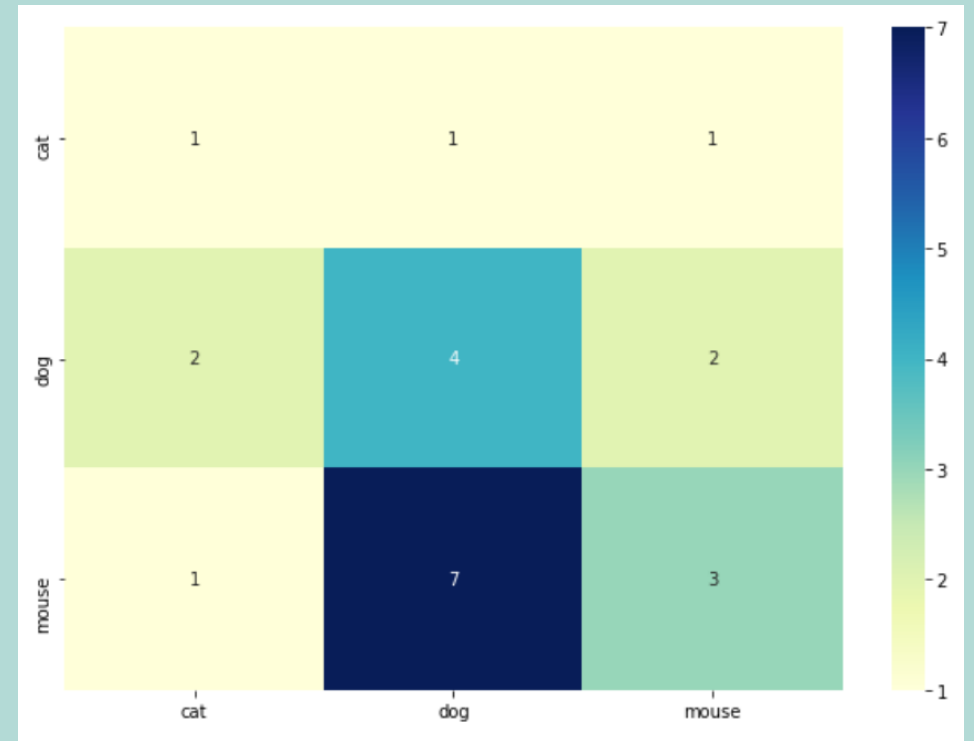
$$\text{Recall: } R = \frac{TP}{TP+FN}$$

$$\text{F1 score: } F_1 = \frac{2PR}{P+R}$$

	Kočka	Pes	Myš
TP	1	4	3
TN	16	6	8
FP	3	8	3
FN	2	4	8
P	0.25	0.33	0.5
R	0.33	0.5	0.27
F1	0.29	0.4	0.35

Vizualizace matice záměn

	Kočka	Pes	Myš
Kočka	1	1	1
Pes	2	4	2
Myš	1	7	3



Porovnání s „pravdou“?

Co je „pravda“ (ground truth, golden standard) a kde ji získat?

Manuální anotace:

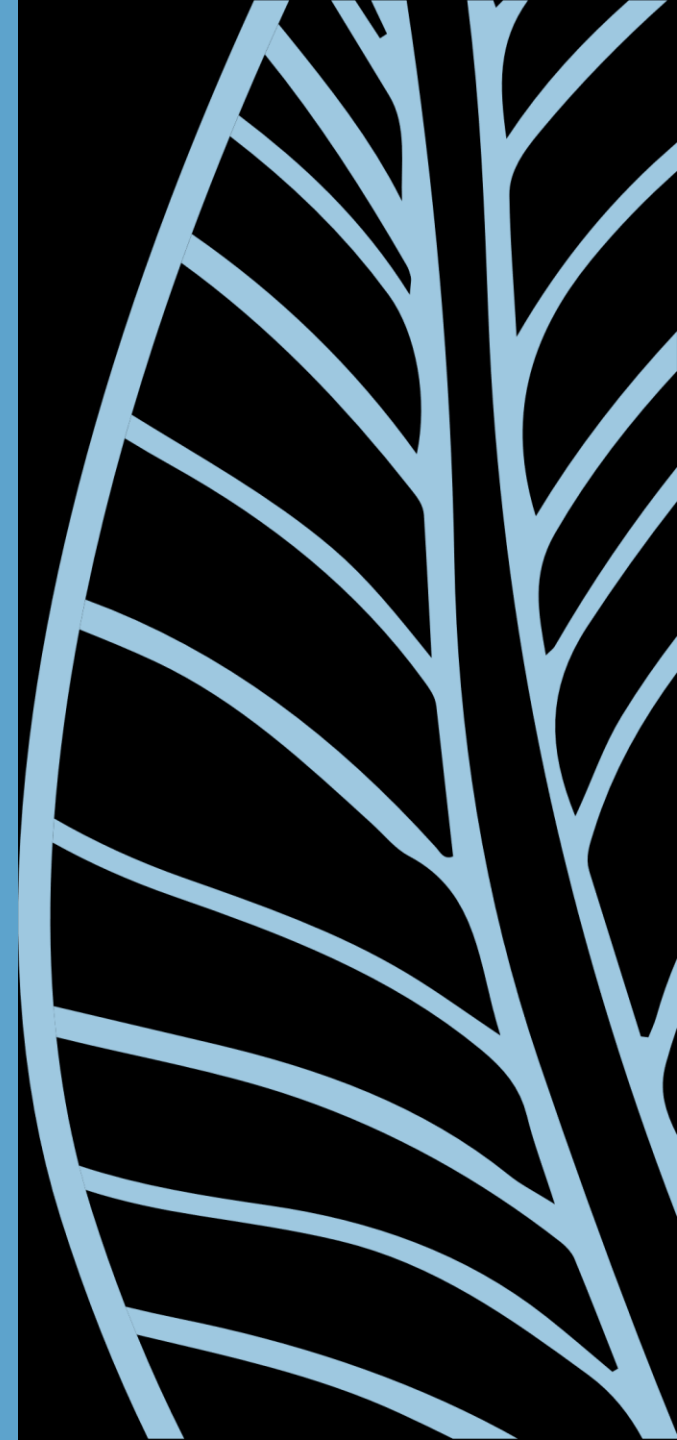
- Anotační manuál, školení anotátorů
- Vícenásobná anotace
- Vyhodnocení anotátorů

Shoda a neshoda:

- Jak rozhodnout v případě neshody?
- Náhodná shoda?



Fleiss κ , Cohen κ

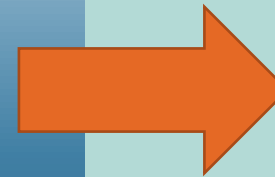


Robustní evaluace

Predikce + Anotace
Metoda vyhodnocení

Začátek dat

Konec dat



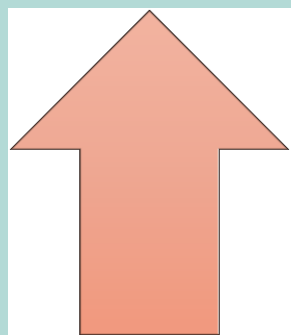
Zamíchat data

Opakovat měření

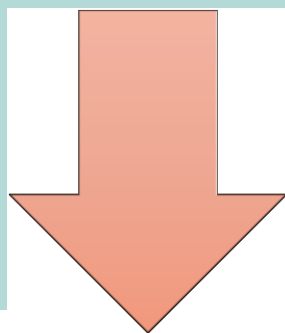
Křížová validace

Kompromis mezi přesností a pokrytím

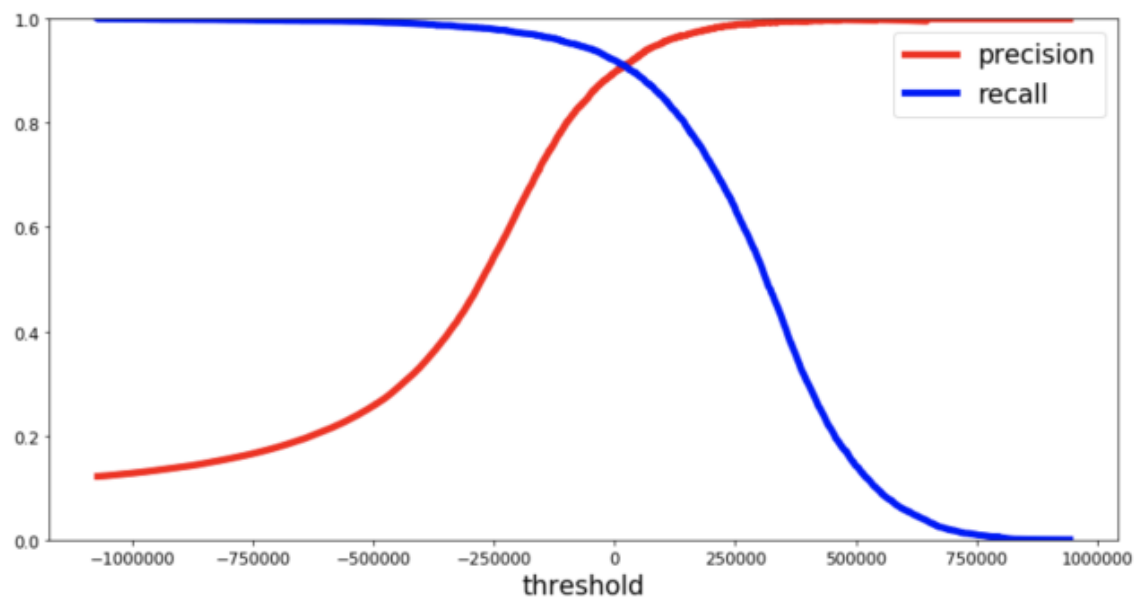
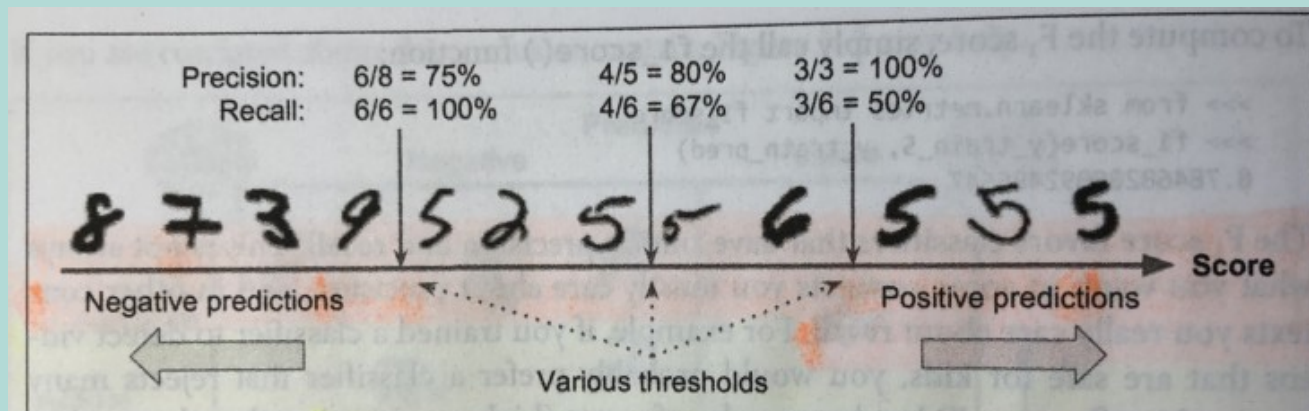
Precision-Recall Tradeoff



přesnost

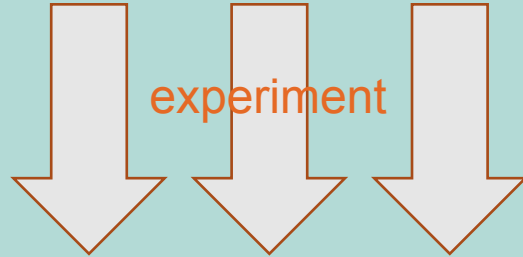


pokrytí



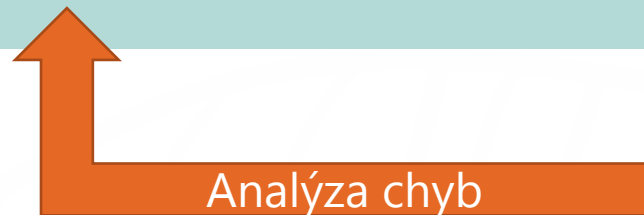
Kvantitativní a kvalitativní vyhodnocení

Pozorování → Hypotéza



Jak daleko je hypotéza od pozorování?

Proč a jaký to má dopad?



Kvalitativní vyhodnocení

Vyhodnocení typicky na menším vzorku dat

Otázky na celkový dopad modelu:

- Jaký je přínos modelu?
- Jaké pocity vzbuzují predikce?

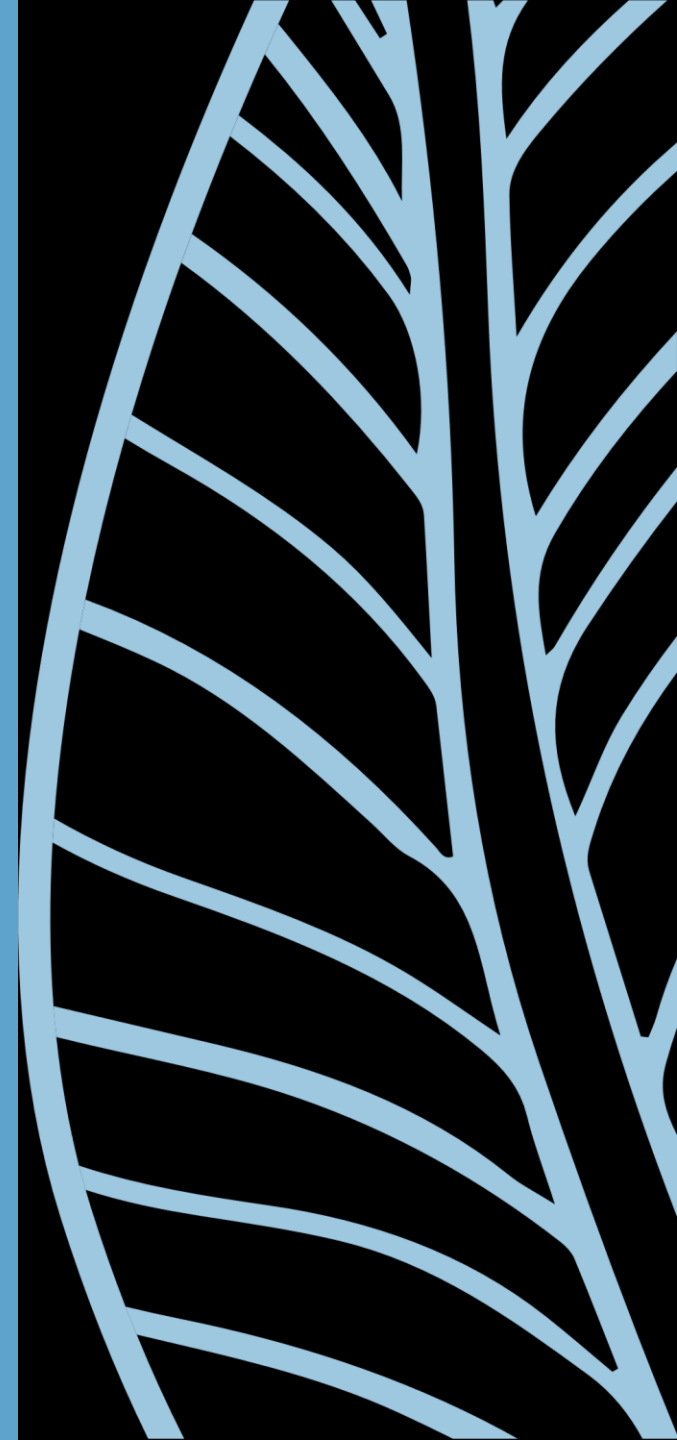
Respondenti:

- Pracovní pozice
- Zkušenosti
- ...

Dotazník

Evaluační hra (gamification,
Serious games)

Sledování činnosti uživatele





Literatura

- Gueter Josmy Faure: **Euclidean Distance and Cosine Similarity. Which One to Use and When?** Medium.com, September 2020. <https://medium.com/swlh/euclidean-distance-and-cosine-similarity-which-one-to-use-and-when-28c97a18fe68>
- Benjamin Paassen: **Revisiting the tree edit distance and its backtracing: A tutorial.** Computer Science, Mathematics ArXiv. 2018. <https://arxiv.org/abs/1805.06869v3>
- Joydwip Mohajon: **Confusion Matrix for Your Multi-Class Machine Learning Model: A beginner's guide on how to calculate Precision, Recall, F1-score for a multi-class classification problem.** Towards Data Science. 2020. <https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826>
- Wikipedia contributors. (2020, November 10). **Precision and recall.** In *Wikipedia, The Free Encyclopedia*. Retrieved 18:00, December 9, 2020, from https://en.wikipedia.org/w/index.php?title=Precision_and_recall&oldid=988053867
- Niklas Donges: **Evaluation Metrics for Classification.** 3. April 2018. machinelearning-blog.com. <https://machinelearning-blog.com/2018/04/03/evaluation-metrics-for-classification/>