Cvičení 7

Pojmy potřebné pro následující cvičení: metrický prostor, metrika, vzdálenost, podobnost, ultrametrika, pseudometrika, nepodobnost, normování (normalizace), převod vzdálenosti na podobnost. Použité metriky: Minkowského (Manhattan, Euklid, Čebyšev), Hamming, LCS, editační, nejkratší cesta (shortest path) a použité podobnosti: kosinova, Jaccardova, Růžičkova.

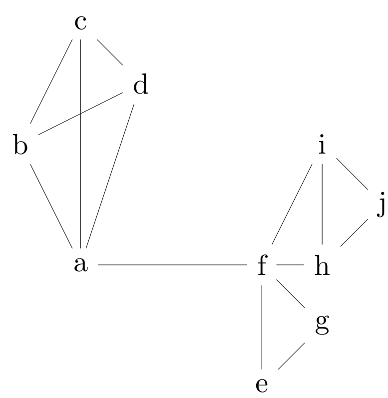
Příklad 1: Pokud je možné, tak použijte dále uvedené metody pro výpočet vzdálenosti a podobnosti mezi dvojici slov, mezi dvojici množin, mezi dvojici vektorů nebo mezi dvojici binárních vektorů. Manhatenovská vzdálenost, euklidovská vzdálenost, Čebyševova vzdálenost, Hammingova vzdálenost, editační vzdálenost (Levenshtein), longest common subsequence, Tanimotova vzdálenost, Jaccardova vzdálenost, kosinova podobnost, Jaccardova podobnost a Růžičkova podobnost.

a) $(1, 2, 3)$ - $(2, 2, 4)$
b) $(1, 2, 3)$ - $(2, 2, 40)$
c) $(1, 2, 3, 2)$- $(2, 2, 4, 9)$
d) lesnictví - lesněnky,
e) lesněnky - žvýkačky,
f) logika - geolog,
g) {pes, ves, nes, les, děs, rez, bez, fez, jez, lez, vez} - { les, lez, lis, los, nos, voz},
h) $\{a, b, c, d\} - \{a, c, e, f, g\}$
i) 10101 - 01010,
j) 011110 - 111111,
k) ACCGATTA - ACAGATAA.

Příklad 2: Určete podobnost vrcholů grafu na základě Jaccardovy podobnosti množin sousedů ($N_v^+ = \{u \in V; (u, v) \in E\} \cup \{v\}$), tj. $s'_J(u, v) = \frac{|N_u^+ \cap N_v^+|}{|N_u^+ \cup N_v^+|}$ a cosinovy podobnosti mezi vektory odpovídající vrcholům z matice sousednosti $s'_c(u, v) = \frac{\sum_{\forall i} A(u,i) \cdot A(v,i)}{\|A(v,)\|_2 \cdot \|A(u,)\|_2}$. Jak na základě této podobností vytvoříte matici nepodobností či dokonce vzdáleností? Určete vzdálenost pomocí nejkratších cest.

a) $V = \{1, 2, 3, 4, 5, 6\}$, $E = \{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (2, 5), (3, 6), (4, 5), (5, 6)\}$
$s'_J(1, 2) =$ $d'_J(1, 2) =$ $s'_c(1, 2) =$ $d'_c(1, 2) =$ $d_{sp}(1, 2) =$
$s'_J(1, 5) =$ $d_J(1, 5) =$ $s'_c(1, 5) =$ $d'_c(1, 5) =$ $d_{sp}(1, 2) =$

b) $s'_J(a, b) =$ $d'_J(a, b) =$
$s'_c(a, b) =$ $d'_c(a, b) =$
$d_{sp}(a, b) =$
$s'_J(a, h) =$ $d'_J(a, h) =$
$s'_c(a, h) =$ $d'_c(a, h) =$
$d_{sp}(a, h) =$



Příklad 3: Je dána množina prvků $A \subseteq \mathbb{R}^2$. Určete a zakreslete, které prvky z množiny $A = \{(0,0),(1,0),(-1,-1),(\sqrt{2}/2,-\sqrt{2}/2),(0.5,-0.5),(-0.2,0.8),(0.33,0.99)\}$ patří do:
uzavřené koule $K_m(0,1) =$, otevřené koule $Bm(0,1) =$, sféry $S_m(0,1) =$
uzavřené koule $K_e(0,1) =$, otevřené koule $B_e(0,1) =$, sféry $S_e(0,1) =$
uzavřené koule $K_c(0,1) =$, otevřené koule $B_c(0,1) =$, sféry $S_c(0,1) =$

Příklad 4: Je dána množina slov A={kles, kos, led, lem, len, lep, les, lesy, lis, los, lup, nes, nos, pes, ples, přes, ves, vez}. Určete, které prvky z množiny A patří do:
uzavřené koule $K_L(les, 1) =$
uzavřené koule $K_{LCS}(les, 1) =$
uzavřené koule $K_{Hamming}(les, 1) =$
sféry $S_L(los, 1) =$
sféry $S_{LCS}(pes, 1) =$
sféry $S_{Hamming}(led, 1) =$

Příklad 5: Na množině $A = \{a, b, c, d, e, f\}$ jsou dány následující relace:

a) ekvivalence ρ_1, která indukuje rozklad $A/\rho_1 = \{\{a, b, c, d\}, \{e, f\}\}$,
b) ekvivalence ρ_2, která indukuje rozklad $A/\rho_2 = \{A\}$,
c) tolerance ρ_3, která indukuje pokrytí $A/\rho_3 = \{\{b, d, e, f\}, \{a, b, c, d\}\}$,
d) $\rho_4 = \{(a, b),(b, c),(c, d),(d, e),(e, f),(f, a)\}$.

Nakreslete grafy, které reprezentují relace ρ_i a rozhodněte, zda lze použít délku nejkratší cesty v těchto grafech jako metriku (tj. $d_{SP}(x, y) = |SP(x, y)|$). Pokud je vzniklý prostor (A, d_i) metrický či ultrametrický (rozhodněte), tak vypočtěte průměr tohoto prostoru ($diam_i(A)$).

Příklad 6: Určete objemy dvou hyperkrychlí a hyperkoule v n-dimenzionálním euklei- dovském prostoru. Délka strany menší hyperkrychle je $a_1 = 1$, délka strany větší hyperkrychle je $a_2 = 2$ a poloměr hyperkoule je $r = 1$.

a) Objem menší hyperkrychle: $VC_n(1) =$
b) Objem větší hyperkrychle: $VC_n(2) =$
c) Objem hyperkoule: $VB_n(r) =$
d) Poměr objemu hyperkoule (která může být do hyperkrychle vepsaná) k objemu hyperkrychle (pro n=2,3,4, ...):
e) Vzdálenost středu menší hyperkrychle od stěny: $d(S, st) =$
f) Délka hlavní diagonály v menší hyperkrychli: $d(A_1, A_n) =$
g) $\lim_{n\to\infty} \frac{VB_n(R)-VB_n(R(1-\epsilon))}{VB_n(R)} =$