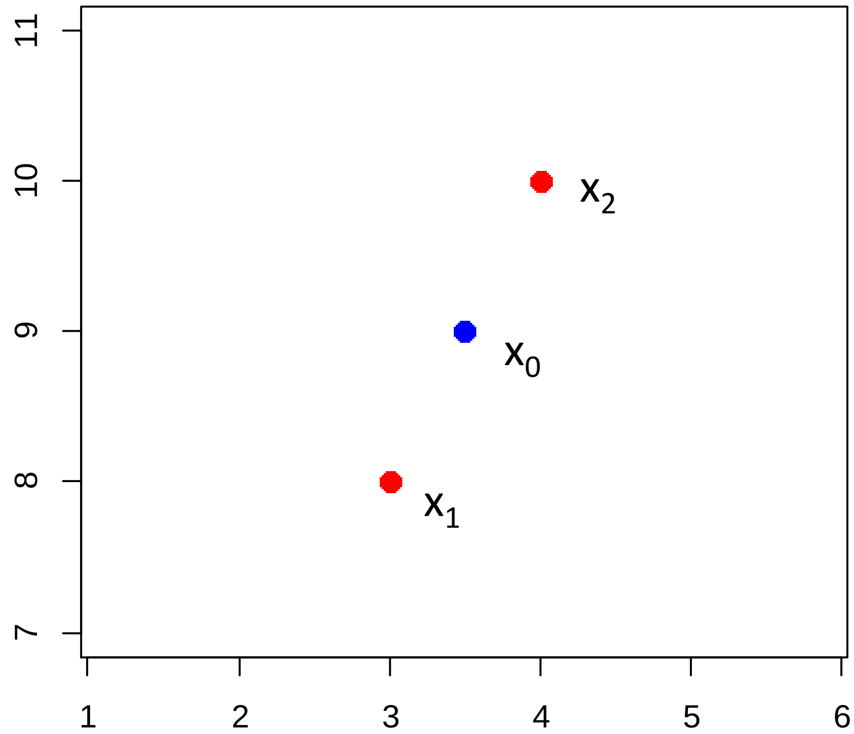
Vícerozměrné metody - cvičení

RNDr. Eva Koriťáková, Ph.D.

Příklad 1 – Vzdálenosti

- **Zadání:** Zjistěte, zda má subjekt $\mathbf{x}_0 = [3{,}5 \quad 9]$ kratší vzdálenost k subjektu $\mathbf{x}_1 = [3 \quad 8]$ či k subjektu $\mathbf{x}_2 = [4 \quad 10]$ pomocí Euklidovy, Hammingovy (manhattanské), Čebyševovy a Canberrské metriky.

- **Vizualizace:**



Euklidova metrika

$$D_E(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{i=1}^{n} (\mathrm{x}_{1i} - \mathrm{x}_{2i})^2}$$

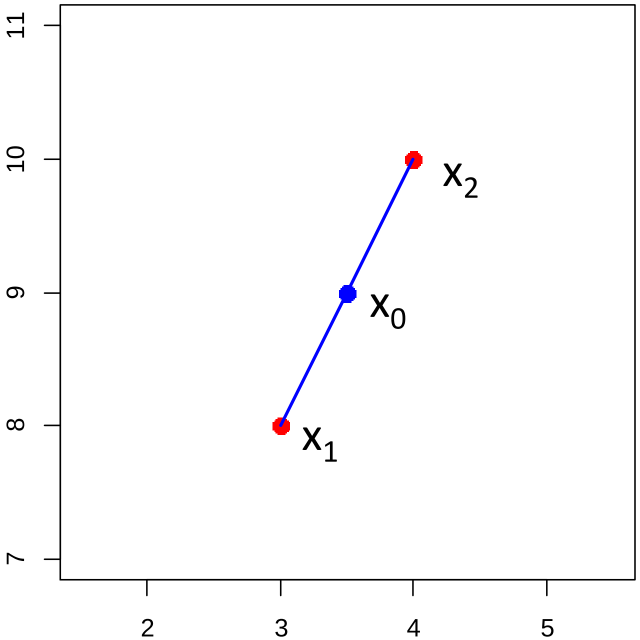
$$d_E(\mathbf{x}_1, \mathbf{x}_0) = \sqrt{(\mathrm{x}_{11} - \mathrm{x}_{01})^2 + (\mathrm{x}_{12} - \mathrm{x}_{02})^2} = \sqrt{(3 - 3{,}5)^2 + (8 - 9)^2} = \sqrt{0{,}25 + 1} = \\ = 1{,}12$$

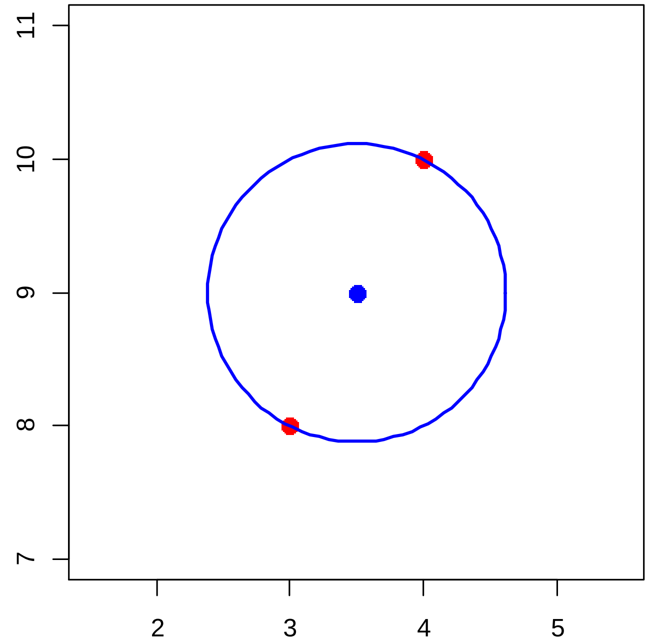
$$d_E(\mathbf{x}_2, \mathbf{x}_0) = \sqrt{(\mathrm{x}_{21} - \mathrm{x}_{01})^2 + (\mathrm{x}_{22} - \mathrm{x}_{02})^2} = \sqrt{(4 - 3{,}5)^2 + (10 - 9)^2} = \sqrt{0{,}25 + 1} = \\ = 1{,}12$$

$\mathbf{x}_0 = [3{,}5 \quad 9]$

$\mathbf{x}_1 = [3 \quad 8]$

$\mathbf{x}_2 = [4 \quad 10]$





Závěr:
Vzdálenost je stejná.

Hammingova (manhattanská) m.

$$D_H(\mathbf{x}_1, \mathbf{x}_2) = \sum_{i=1}^{n} |\mathrm{x}_{1i} - \mathrm{x}_{2i}|$$

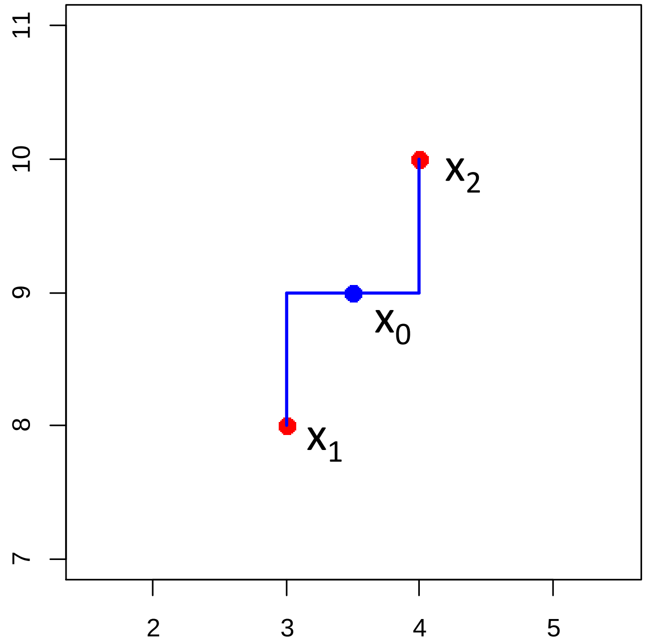
$$d_H(\mathbf{x}_1, \mathbf{x}_0) = |\mathrm{x}_{11} - \mathrm{x}_{01}| + |\mathrm{x}_{12} - \mathrm{x}_{02}| = |3 - 3{,}5| + |8 - 9| = 0{,}5 + 1 = 1{,}5$$

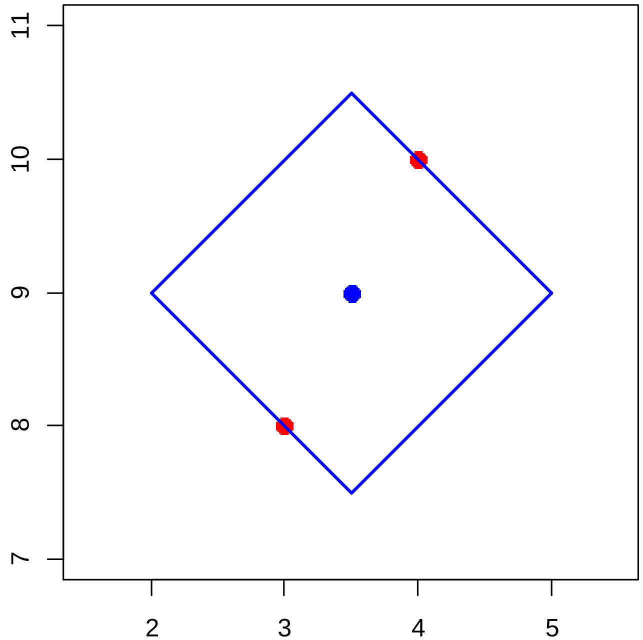
$$d_H(\mathbf{x}_2, \mathbf{x}_0) = |\mathrm{x}_{21} - \mathrm{x}_{01}| + |\mathrm{x}_{22} - \mathrm{x}_{02}| = |4 - 3{,}5| + |10 - 9| = 0{,}5 + 1 = 1{,}5$$

$\mathbf{x}_0 = [3{,}5 \quad 9]$

$\mathbf{x}_1 = [3 \quad 8]$

$\mathbf{x}_2 = [4 \quad 10]$





Závěr:
Vzdálenost je stejná.

Čebyševova metrika

$$D_C(\mathbf{x}_1, \mathbf{x}_2) = \max_{\forall i} |x_{1i} - x_{2i}|$$

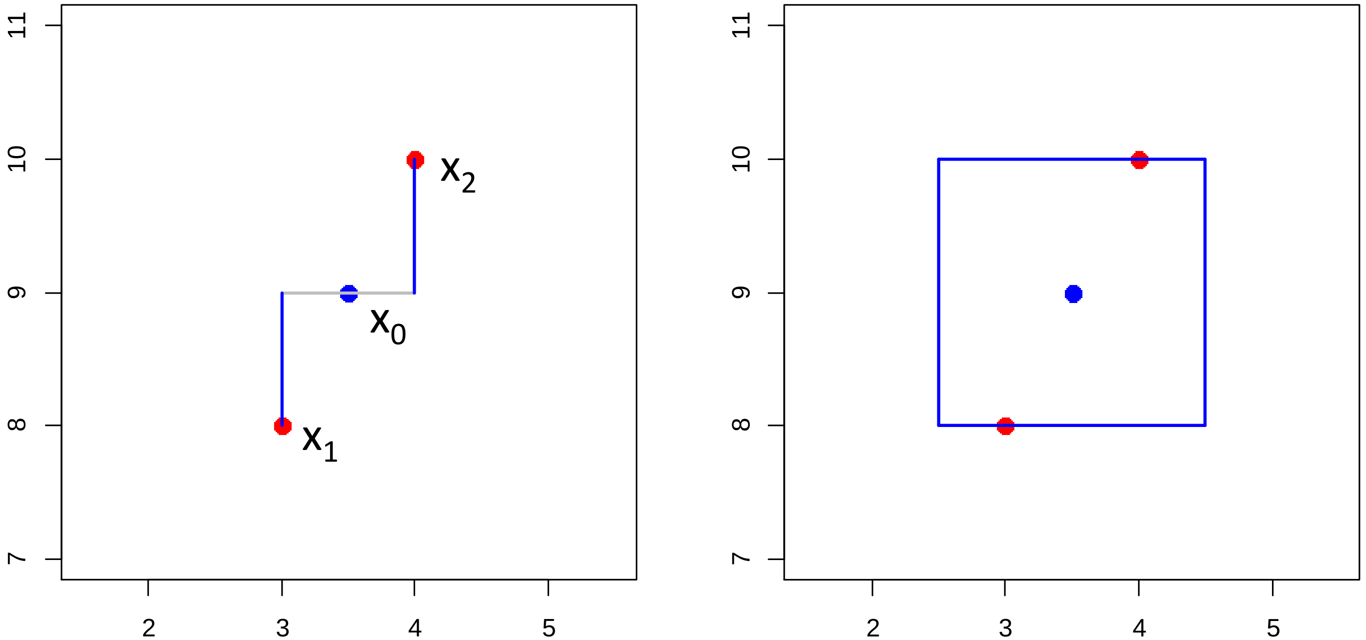
$$d_C(\mathbf{x}_1, \mathbf{x}_0) = \max(|x_{11} - x_{01}|; |x_{12} - x_{02}|) = \max(|3 - 3{,}5|; |8 - 9|) = \max(0{,}5; 1) = 1$$

$$d_C(\mathbf{x}_2, \mathbf{x}_0) = \max(|x_{21} - x_{01}|; |x_{22} - x_{02}|) = \max(|4 - 3{,}5|; |10 - 9|) = \max(0{,}5; 1) = 1$$

$\mathbf{x}_0 = [3{,}5 \quad 9]$

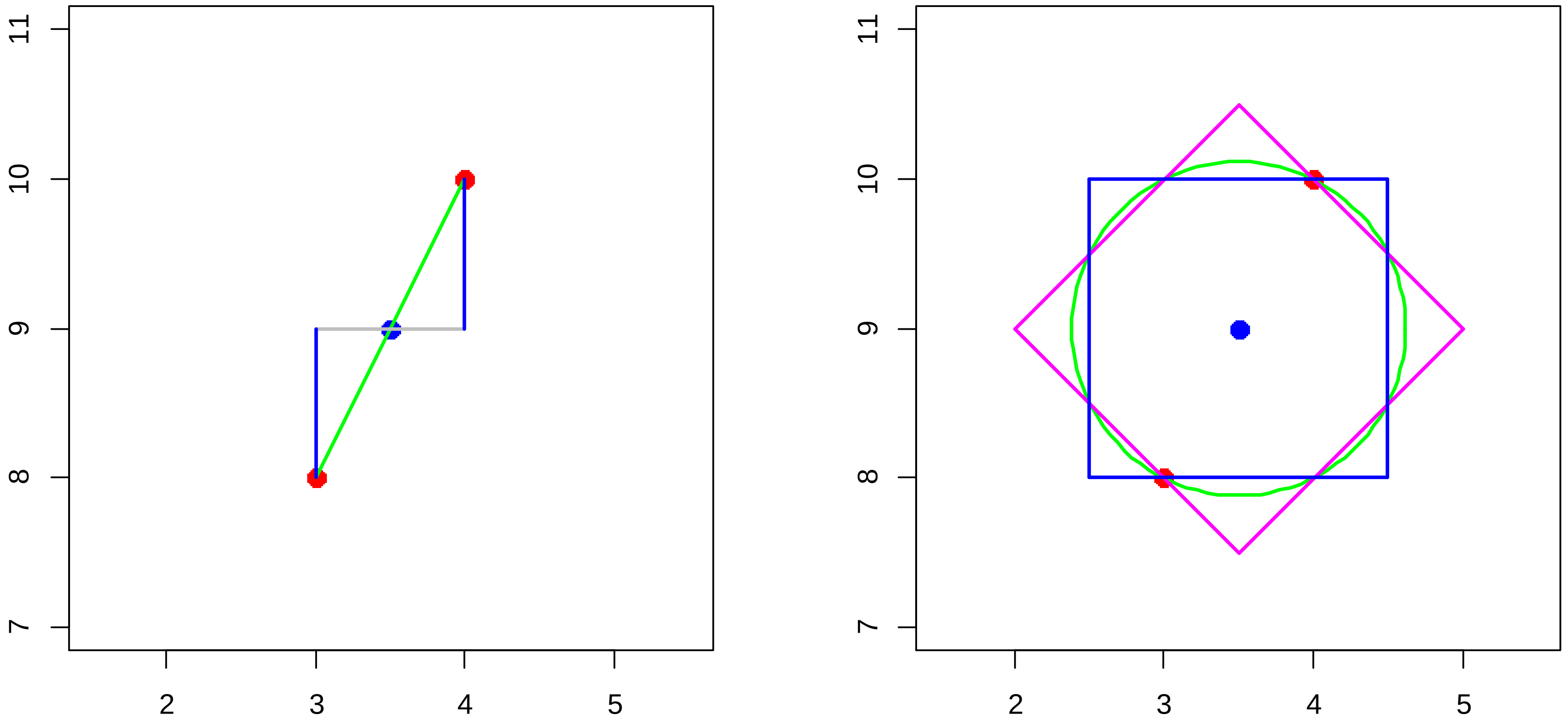
$\mathbf{x}_1 = [3 \quad 8]$

$\mathbf{x}_2 = [4 \quad 10]$



Závěr:
Vzdálenost je stejná.

Srovnání metrik



Canberrská metrika

$$D_{CA}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{i=1}^{n} \frac{|\mathrm{x}_{1i} - \mathrm{x}_{2i}|}{|\mathrm{x}_{1i}| + |\mathrm{x}_{2i}|}$$

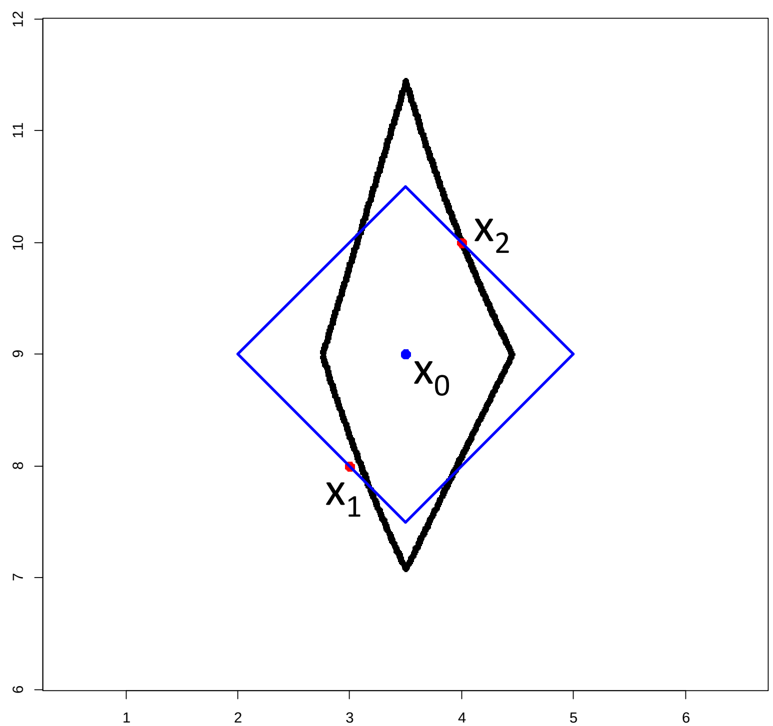
$$d_{CA}(\mathbf{x}_1, \mathbf{x}_0) = \frac{|\mathrm{x}_{11} - \mathrm{x}_{01}|}{|\mathrm{x}_{11}| + |\mathrm{x}_{01}|} + \frac{|\mathrm{x}_{12} - \mathrm{x}_{02}|}{|\mathrm{x}_{12}| + |\mathrm{x}_{02}|} = \frac{|3 - 3{,}5|}{|3| + |3{,}5|} + \frac{|8 - 9|}{|8| + |9|} = \frac{0{,}5}{6{,}5} + \frac{1}{17} = 0{,}14$$

$$d_{CA}(\mathbf{x}_2, \mathbf{x}_0) = \frac{|\mathrm{x}_{21} - \mathrm{x}_{01}|}{|\mathrm{x}_{21}| + |\mathrm{x}_{01}|} + \frac{|\mathrm{x}_{22} - \mathrm{x}_{02}|}{|\mathrm{x}_{22}| + |\mathrm{x}_{02}|} = \frac{|4 - 3{,}5|}{|4| + |3{,}5|} + \frac{|10 - 9|}{|10| + |9|} = \frac{0{,}5}{7{,}5} + \frac{1}{19} = 0{,}12$$

$\mathbf{x}_0 = [3{,}5 \quad 9]$

$\mathbf{x}_1 = [3 \quad 8]$

$\mathbf{x}_2 = [4 \quad 10]$



Závěr:
Subjekt $\mathbf{x}_0$ má kratší vzdálenost od subjektu $\mathbf{x}_2$ než od subjektu $\mathbf{x}_1$.

Asociační koeficienty

		$\mathbf{x}_j$	
		false/0	true/1
$\mathbf{x}_i$	false/0	D	C
	true/1	B	A

Jaccardův – Tanimotův asociační koeficient

$$S_{JT}(\mathbf{x}, \mathbf{y}) = \frac{A}{A + B + C}$$

Sokalův – Michenerův asociační koeficient

$$S_{SM}(\mathbf{x}, \mathbf{y}) = \frac{A + D}{A + B + C + D}$$

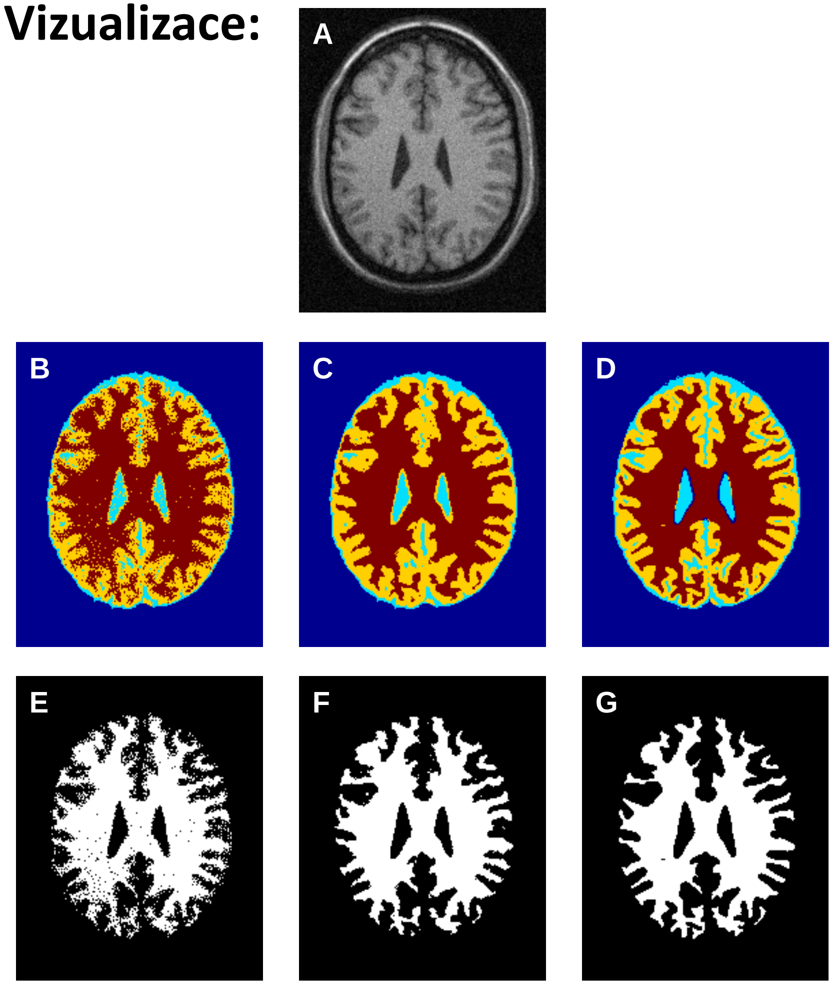
Příklad 2 – Podobnosti

Zadání: Byla provedena segmentace bílé hmoty v obrazu mozku z magnetické rezonance pomocí dvou segmentačních metod (viz *Obrázek 1*). Chceme výsledky segmentace srovnat s maskou bílé hmoty, která byla získána z atlasu mozku. Zajímá nás tedy překryv s maskou, na základě čehož budeme moci usoudit, která metoda segmentuje obraz lépe.

Obrázek 1 Vizualizace segmentace bílé hmoty pomocí dvou segmentačních metod a jejich srovnání s atlasem mozku.

A) původní obraz mozku z magnetické rezonance;
B) segmentovaný obraz pomocí metody *k*-průměrů;
C) segmentovaný obraz pomocí metody *k*-nejbližších sousedů (*k*-NN);
D) obraz segmentovaný na základě atlasu mozku;
E) obraz bílé hmoty vzniklý prahováním obrazu B (tzn. na základě metody *k*-průměrů);
F) obraz bílé hmoty mozkové vzniklý prahováním obrazu C (tzn. na základě metody *k*-NN);
G) obraz bílé hmoty mozkové vzniklý prahováním obrazu D (tzn. na základě atlasu mozku).

Vizualizace:



V obrazech B až D tmavě červená barva značí bílou hmotu, žlutá značí šedou hmotu, světle modrá značí mozkomíšní mok a tmavě modrá značí pozadí.

Příklad 2 – Podobnosti – řešení

Počty voxelů označených jako bílá hmota pomocí segmentačních metod a jejich srovnání s maskou sumarizujeme:

Tabulka 1. Sumarizace počtu voxelů označených a neoznačených jako bílá hmota na základě segmentace **metodou *k*-průměrů** a na základě masky.

		$\mathbf{y}$		
		0	1	Celkem
$\mathbf{x}_1$	0	28 453 (D_1)	477 (C_1)	28 930
	1	1406 (B_1)	8 941 (A_1)	10 347
	Celkem	29 859	9 418	39 277 (N)

Tabulka 2. Sumarizace počtu voxelů označených a neoznačených jako bílá hmota na základě segmentace **metodou *k*-nejbližších sousedů** a na základě masky.

		$\mathbf{y}$		
		0	1	Celkem
$\mathbf{x}_2$	0	29 046 (D_2)	284 (C_2)	29 330
	1	813 (B_2)	9 134 (A_2)	9 947
	Celkem	29 859	9 418	39 277 (N)

$\mathbf{x}_1$ - vektor počtu voxelů neoznačených jako bílá hmota (0) a počtu voxelů označených jako bílá hmota (1) na základě segmentace metodou *k*-průměrů;
$\mathbf{x}_2$ - vektor počtu voxelů neoznačených a označených jako bílá hmota na základě segmentace metodou *k*-nejbližších sousedů;
$\mathbf{y}$ je vektor počtu voxelů neoznačených a označených jako bílá hmota na základě masky.

Příklad 2 – Jaccardův-Tanimotův asociační koef.

Tabulka 1. Metoda *k*-průměrů

		y		
		0	1	Celkem
$\mathbf{x}_1$	0	28 453 (D_1)	477 (C_1)	28 930
	1	1406 (B_1)	8 941 (A_1)	10 347
	Celkem	29 859	9 418	39 277 (N)

Tabulka 2. Metoda *k*-nejbližších sousedů

		y		
		0	1	Celkem
$\mathbf{x}_2$	0	29 046 (D_2)	284 (C_2)	29 330
	1	813 (B_2)	9 134 (A_2)	9 947
	Celkem	29 859	9 418	39 277 (N)

$\mathbf{x}_1$ - vektor počtu voxelů neoznačených jako bílá hmota (0) a počtu voxelů označených jako bílá hmota (1) na základě segmentace metodou *k*-průměrů;
$\mathbf{x}_2$ - vektor počtu voxelů neoznačených a označených jako bílá hmota na základě segmentace metodou *k*-nejbližších sousedů;
$\mathbf{y}$ je vektor počtu voxelů neoznačených a označených jako bílá hmota na základě masky.

Výpočet podobnosti mezi osegmentovanými obrazy a maskou:

Jaccardův-Tanimotův asociační koeficient:

$$s_{JT}(\mathbf{x}_1,\mathbf{y}) = \frac{\mathrm{A}_1}{\mathrm{A}_1+\mathrm{B}_1+\mathrm{C}_1} = \frac{8\,941}{8\,941+1\,406+477} = 0{,}826$$

$$s_{JT}(\mathbf{x}_2,\mathbf{y}) = \frac{\mathrm{A}_2}{\mathrm{A}_2+\mathrm{B}_2+\mathrm{C}_2} = \frac{9\,134}{9\,134+813+284} = 0{,}893$$

Větší podobnost s maskou má obraz segmentovaný metodou *k*-nejbližších sousedů, metoda *k*-nejbližších sousedů tedy osegmentovala obraz lépe než metoda *k*-průměrů.

Příklad 2 – Sokalův-Michenerův asociační koef.

Tabulka 1. Metoda *k*-průměrů

		y		
		0	1	Celkem
$\mathbf{x}_1$	0	28 453 (D_1)	477 (C_1)	28 930
	1	1406 (B_1)	8 941 (A_1)	10 347
	Celkem	29 859	9 418	39 277 (N)

Tabulka 2. Metoda *k*-nejbližších sousedů

		y		
		0	1	Celkem
$\mathbf{x}_2$	0	29 046 (D_2)	284 (C_2)	29 330
	1	813 (B_2)	9 134 (A_2)	9 947
	Celkem	29 859	9 418	39 277 (N)

$\mathbf{x}_1$ - vektor počtu voxelů neoznačených jako bílá hmota (0) a počtu voxelů označených jako bílá hmota (1) na základě segmentace metodou *k*-průměrů;
$\mathbf{x}_2$ - vektor počtu voxelů neoznačených a označených jako bílá hmota na základě segmentace metodou *k*-nejbližších sousedů;
$\mathbf{y}$ je vektor počtu voxelů neoznačených a označených jako bílá hmota na základě masky.

Výpočet podobnosti mezi osegmentovanými obrazy a maskou:

Sokalův-Michenerův (simple matching = jednoduchý srovnávací) asoc. koeficient:

$$s_{SM}(\mathbf{x}_1, \mathbf{y}) = \frac{A_1 + D_1}{A_1 + B_1 + C_1 + D_1} = \frac{8\,941 + 28\,453}{8\,941 + 1\,406 + 477 + 28\,453} = 0{,}952$$

$$s_{SM}(\mathbf{x}_2, \mathbf{y}) = \frac{A_2 + D_2}{A_2 + B_2 + C_2 + D_2} = \frac{9\,134 + 29\,046}{9\,134 + 813 + 284 + 29\,046} = 0{,}972$$

Větší podobnost s maskou má obraz segmentovaný metodou *k*-nejbližších sousedů, metoda *k*-nejbližších sousedů tedy osegmentovala obraz lépe než metoda *k*-průměrů.

Příklad 2 – Srovnání koeficientů

Jaccardův-Tanimotův asociační koeficient:

$$s_{JT}(\mathbf{x}_1, \mathbf{y}) = \frac{\mathrm{A}_1}{\mathrm{A}_1+\mathrm{B}_1+\mathrm{C}_1} = \frac{8\,941}{8\,941 + 1\,406 + 477} = 0{,}826$$

$$s_{JT}(\mathbf{x}_2, \mathbf{y}) = \frac{\mathrm{A}_2}{\mathrm{A}_2+\mathrm{B}_2+\mathrm{C}_2} = \frac{9\,134}{9\,134 + 813 + 284} = 0{,}893$$

Sokalův-Michenerův (simple matching = jednoduchý srovnávací) asoc. koeficient:

$$s_{SM}(\mathbf{x}_1, \mathbf{y}) = \frac{\mathrm{A}_1+\mathrm{D}_1}{\mathrm{A}_1+\mathrm{B}_1+\mathrm{C}_1+\mathrm{D}_1} = \frac{8\,941 + 28\,453}{8\,941 + 1\,406 + 477 + 28\,453} = 0{,}952$$

$$s_{SM}(\mathbf{x}_2, \mathbf{y}) = \frac{\mathrm{A}_2+\mathrm{D}_2}{\mathrm{A}_2+\mathrm{B}_2+\mathrm{C}_2+\mathrm{D}_2} = \frac{9\,134 + 29\,046}{9\,134 + 813 + 284 + 29\,046} = 0{,}972$$

Závěr:

- Oba koeficienty dokázaly určit, že větší podobnost s maskou má obraz segmentovaný metodou *k*-nejbližších sousedů, metoda *k*-nejbližších sousedů tedy osegmentovala obraz lépe než metoda *k*-průměrů.
- Protože Jaccardův-Tanimotův asociační koeficient nezahrnuje negativní shody (tzn. voxely pozadí), umožňuje lépe postihnout rozdíl v úspěšnosti mezi segmentačními metodami.