

Klasifikace podle minimální vzdálenosti – příklad

Bylo provedeno měření objemu hipokampu a objemu amygdaly u 3 pacientů s Alzheimerovou chorobou () a 3 kontrolních subjektů (). Naměřené hodnoty byly (v řádcích) zaznamenány do matic resp. (označení D – diseased, H – healthy):

Určete, zda testovací subjekt patří do skupiny pacientů či kontrolních subjektů pomocí klasifikace podle minimální vzdálenosti.

Řešení:

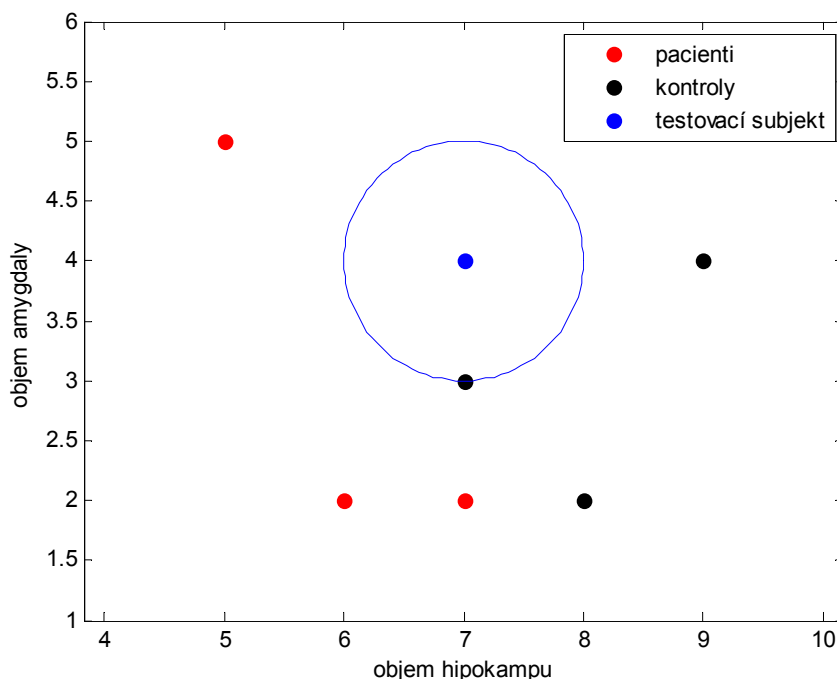
Postup:

- I) zvolení metriky pro výpočet vzdáleností 2 bodů
- II) zvolení metriky pro určení vzdálenosti mezi 2 množinami bodů

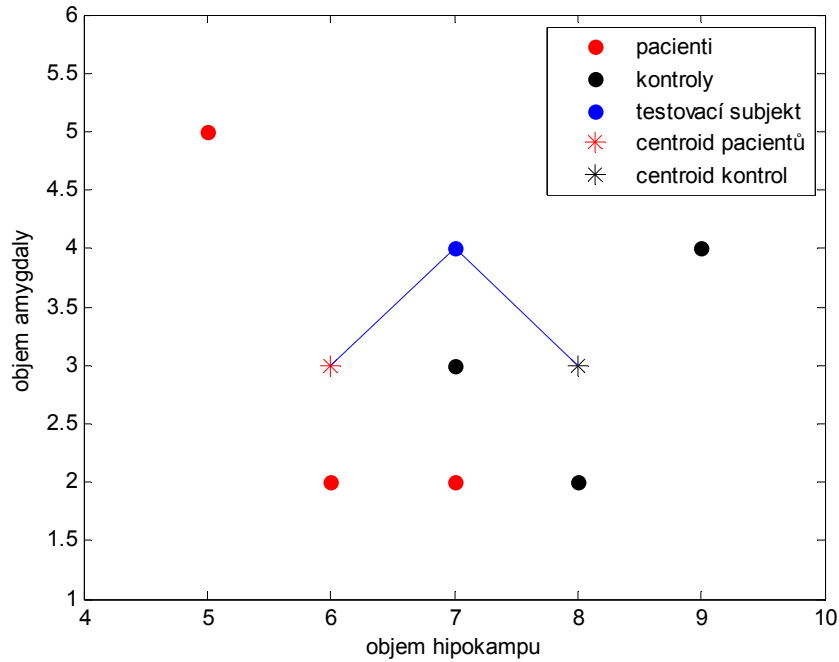
Předpoklad: budeme shlukovací algoritmy využívat jako neučící se algoritmy (klasifikátor natrénujeme na celé trénovací množině a pak už pouze klasifikujeme nové obrazy (např. už nepřepočítáváme centroid po zařazení každého nového subjektu či objektu apod.))

1.1 Metoda k nejbližších sousedů + Euklidova metrika:

Znázornění klasifikace podle nejbližšího souseda pomocí Euklidovy metriky je uvedeno na *Obrázku 1*.



Obrázek 1. Znázornění klasifikace podle nejbližšího souseda pomocí Euklidovy metriky. Modře je vyznačena množina bodů v obrazovém prostoru, které mají od testovacího subjektu stejnou vzdálenost.



Obrázek 2. Ilustrace klasifikace testovacího subjektu pomocí centroidové metody, přičemž vzdálenosti testovacího subjektu od centroidů skupin jsou počítány pomocí Euklidovy metriky. Je patrné, že nelze rozhodnout, do jaké třídy máme testovací subjekt zařadit., protože jeho Euklidova vzdálenost od centroidu pacientů je stejná jako od centroidu kontrol.

Centroidová metoda s využitím medoidu:

Medoid (odvozen vizuálně – spočítal by se tak, že by se našel nejbližší bod k centroidu u dané skupiny nebo jako bod s nejmenší sumou vzdáleností od ostatních bodů)

medoid pro pacienty:

medoid pro kontroly:

Protože _____, testovací subjekt bude zařazen do třídy kontrolních subjektů.

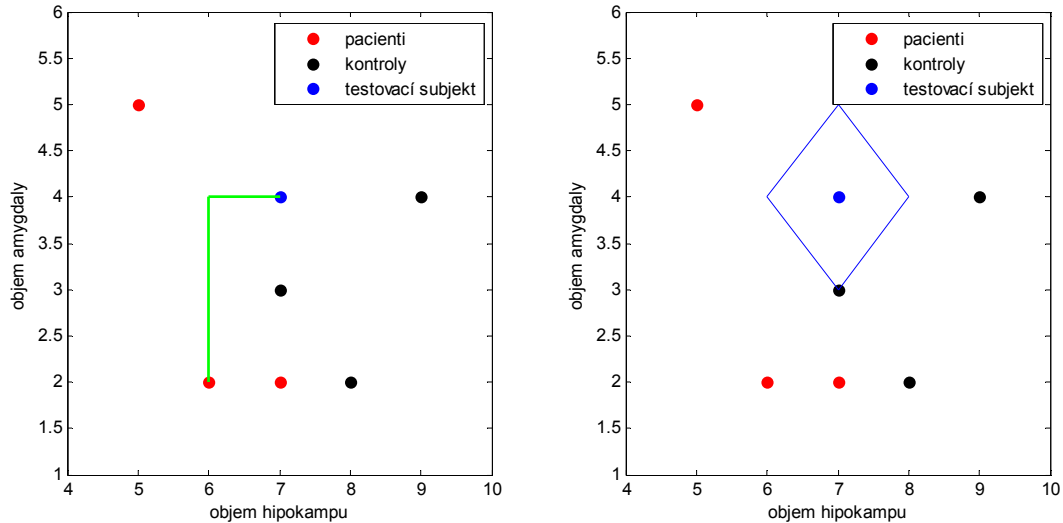
2.1 Metoda k nejbližších sousedů + Hammingova (manhattanská) metrika:

Znázornění výpočtu vzdálenosti dvou bodů pomocí Hammingovy (manhattanské) metriky je uvedeno na Obrázku 3.

Výpočet vzdáleností testovacího (nového) subjektu od všech subjektů z obou skupin:

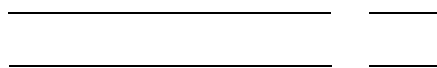
Seřazení vzdáleností:

- pro : nejbližší sused bodu je bod , protože je nejmenší → testovací subjekt bude zařazen do třídy kontrolních subjektů
- pro : nelze rozhodnout
- pro : subjekt zařazen do třídy kontrolních subjektů, protože mezi 3 nejbližšími susedy je 1 pacient a 2 kontrolní subjekty
- pro : nelze rozhodnout
- pro : nelze rozhodnout
- pro : nelze rozhodnout



Obrázek 3. Ilustrace výpočtu vzdálenosti dvou bodů pomocí Hammingovy (manhattanské) metriky (vlevo) a znázornění klasifikace podle nejbližšího suseda (vpravo). Modře je vyznačena množina bodů v obrazovém prostoru, které mají od testovacího subjektu stejnou vzdálenost.

2.2 Metoda průměrné vazby + Hammingova (manhattanská) metrika:

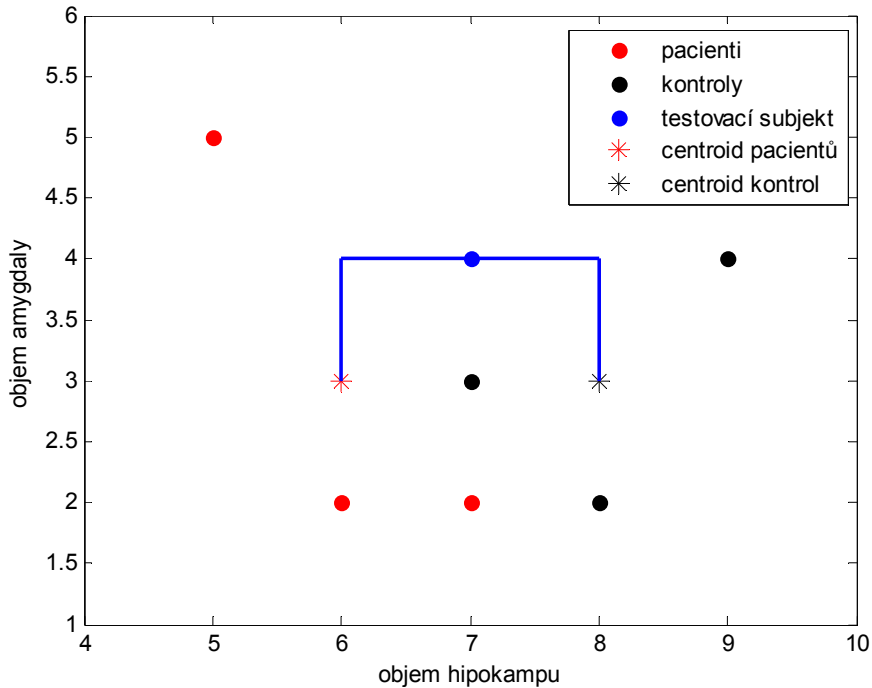


Protože , testovací subjekt bude zařazen do třídy kontrolních subjektů.

2.3 Centroidová metoda + Hammingova (manhattanská) metrika:

Protože , nelze rozhodnout, do které třídy bude testovací subjekt zařazen.

Znázornění klasifikace testovacího subjektu pomocí centroidové metody, přičemž vzdálenosti testovacího subjektu od centroidů skupin jsou počítány pomocí Hammingovy (manhattanské) metriky, je na Obrázku 4.



Obrázek 4. Ilustrace klasifikace testovacího subjektu pomocí centroidové metody, přičemž vzdálenosti testovacího subjektu od centroidů skupin jsou počítány pomocí Hammingovy (manhattanské) metriky. Je patrné, že nelze rozhodnout, do jaké třídy máme testovací subjekt zařadit., protože jeho Hammingova (manhattanská) vzdálenost od centroidu pacientů je stejná jako od centroidu kontrol.

Centroidová metoda s využitím medoidu:

Protože , testovací subjekt bude zařazen do třídy kontrolních subjektů.

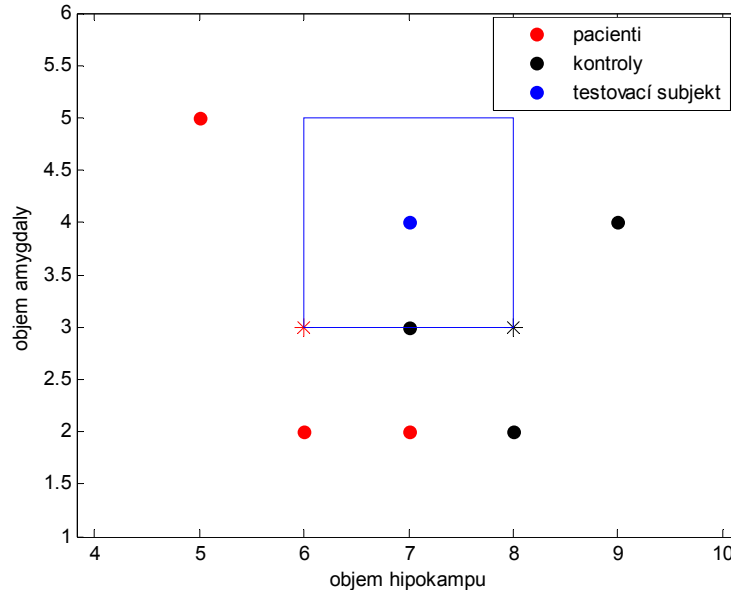
3.1 Metoda k nejbližších sousedů + Čebyševova metrika:

Znázornění klasifikace podle nejbližšího souseda pomocí Čebyševovy metriky je uvedeno na Obrázku 5.

Výpočet vzdáleností testovacího (nového) subjektu od všech subjektů z obou skupin:

Seřazení vzdáleností:

- pro : nejbližší souseď bodu je bod , protože je nejmenší → testovací subjekt bude zařazen do třídy kontrolních subjektů
- pro : nelze rozhodnout
- pro : nelze rozhodnout
- pro : nelze rozhodnout
- pro : nelze rozhodnout
- pro : nelze rozhodnout



Obrázek 5. Znázornění klasifikace podle nejbližšího souseda pomocí Čebyševovy metriky. Modře je vyznačena množina bodů v obrazovém prostoru, které mají od testovacího subjektu stejnou vzdálenost.

3.2 Metoda průměrné vazby + Čebyševova metrika:

Protože _____, testovací subjekt bude zařazen do třídy testovacích subjektů.

3.3 Centroidová metoda + Čebyševova metrika:

Protože _____, nelze rozhodnout, do které třídy bude testovací subjekt zařazen.

Centroidová metoda s využitím medoidu:

Protože _____, testovací subjekt bude zařazen do třídy kontrolních subjektů.

4.1 Metoda k nejbližších sousedů + Canberrská metrika:

Výpočet vzdáleností testovacího (nového) subjektu od všech subjektů z obou skupin:

_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____

Seřazení vzdáleností:

- pro _____ : nejbližší souseď bodu _____ je bod _____, protože _____ je nejmenší → testovací subjekt bude zařazen do třídy kontrolních subjektů
- pro _____ : subjekt zařazen do třídy kontrolních subjektů, protože mezi 2 nejbližšími sousedy jsou 2 kontrolní subjekty a žádný pacient.
- pro _____ : subjekt zařazen do třídy kontrolních subjektů, protože mezi 3 nejbližšími sousedy je 1 pacient a 2 kontrolní subjekty
- pro _____ : nelze rozhodnout
- pro _____ : subjekt zařazen do třídy kontrolních subjektů, protože mezi 5 nejbližšími sousedy jsou 2 pacienti a 3 kontrolní subjekty
- pro _____ : nelze rozhodnout

4.2 Metoda průměrné vazby + Canberrská metrika:

Protože _____, testovací subjekt bude zařazen do třídy kontrolních subjektů.

4.3 Centroidová metoda + Canberrská metrika:

Protože _____, testovací subjekt bude zařazen do třídy kontrolních subjektů.

Centroidová metoda s využitím medoidu:

Protože _____, testovací subjekt bude zařazen do třídy kontrolních subjektů.

5.1 Metoda k nejbližších souseďů + Mahalanobisova metrika:

Nejprve je potřeba vypočítat výběrové kovarianční matice pro třídu pacientů a kontrol, tzn. _____ a _____ (výpočet výběrových kovariančních matic lze nalézt ve

Cvičení 1) a její inverzi její inverzi _____ a _____.

Výpočet vzdáleností testovacího (nového) subjektu od všech subjektů z obou skupin:

_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

Seřazení vzdáleností:

- pro : nejbližší sused bodu je bod , protože je nejmenší → testovací subjekt bude zařazen do třídy kontrolních subjektů
- pro : nelze rozhodnout
- pro : subjekt zařazen do třídy kontrolních subjektů, protože mezi 3 nejbližšími susedy je 1 pacient a 2 kontrolní subjekty
- pro : nelze rozhodnout
- pro : subjekt zařazen do třídy kontrolních subjektů, protože mezi 5 nejbližšími susedy jsou 2 pacienti a 3 kontrolní subjekty
- pro : nelze rozhodnout

5.2 Metoda průměrné vazby + Mahalanobisova metrika:

Protože , testovací subjekt bude zařazen do třídy kontrolních subjektů.

5.3 Centroidová metoda + Mahalanobisova metrika:

Protože , testovací subjekt bude zařazen do třídy kontrolních subjektů.

Centroidová metoda s využitím medoidu:

Protože , testovací subjekt bude zařazen do třídy kontrolních subjektů.

Výsledky uspořádáme do tabulky:

metrika	Euklidova	Hammingova	Čebyševova	Canberrská	Mahalanobisova
NN	H	H	H	H	H
3-NN	H	H	-	H	H
5-NN	-	-	-	H	H
GA	H	H	H	H	H
CE-centroid	-	-	-	H	H
CE-medoid	H	H	H	H	H

Je patrné, že výsledek klasifikace se může lišit při použití různých metrik vzdálenosti. Testovací subjekt sice nebyl v žádném ze sledovaných případů zařazen mezi pacienty, ale v mnoha případech nešlo o výsledné klasifikaci rozhodnout, což se v praxi nejčastěji řeší tak, že je testovaný subjekt zařazen do náhodně zvolené třídy.