



Návod na laboratorní úlohu

Support vector machine pro detekci a klasifikaci cévní mozkové příhody

Název studijního předmětu: Modeling and Simulations in Medicine

Vedoucí cvičení/experimentu: Ing. Tomáš Pokorný

Připravil: Ing. Tomáš Pokorný, Doc. Dr.-Ing. Jan Vrba, M.Sc.

Úvod:

Mikrovlnné zobrazování je metoda, kdy zobrazovanou oblast vystavíme mikrovlnnému záření o nízkých výkonech. Mikrovlnné pole je neionizující a tudíž pro člověka bezpečné. Mikrovlnné výkony využívané v lékařských diagnostických přístrojích jsou zpravidla velmi nízké, řádově jednotky až desítky mW. Mikrovlnné pole při průchodu zobrazovanou oblastí s biologickými tkáněmi interaguje - je tlumeno a rozptylováno. Naměřené elektromagnetické pole v okolí zobrazované oblasti v sobě nese informaci o rozložení biologických tkání. Jednou z nejčastěji diskutovaných aplikací mikrovlnného zobrazování je detekce a klasifikace cévních mozkových příhod (CMP). Oblast postižená CMP vykazuje odlišné elektrické vlastnosti než oblast zdravá. V rámci tohoto cvičení budou připraveny fantomy zobrazované oblasti a pomocí prototypu mikrovlnného zobrazovacího systému budou naměřeny mikrovlnné signály. Ty budou sloužit pro trénování klasifikátoru založeného na support vector machines (SVM) a následně bude klasifikátor ohodnocen.

Algoritmus podpůrných vektorů = Support Vector Machine (SVM)

Algoritmy podpůrných vektorů, v angličtině Support Vector Machine (SVM) je široce rozšířenými jádrovými algoritmy pro řešení problémů v elektromagnetizmu. SVM je jádrovým (kernel) algoritmus, který provádí mapování vstupních dat do uměle vytvořených dimenzí, kde je možné lineární oddělení dat tzv. nadrovinou. Optimální nadrovinu definují podpůrné vektory (Support Vector), které leží nejbližší u hranice. Podpůrné vektory mají stejnou vzdálenost od oddělovací nadroviny a určují tak šířku oddělovacího pásma. Podpůrné vektory jsou trénovací vektory, které mají nejvyšší informační hodnotu pro klasifikační úlohu, ale zároveň představují nejobtížněji klasifikovatelné vzory, protože leží nejbliž u hranice. Metoda SVM se využívá v mnoha oborech, počínaje medicínským zobrazováním, molekulární biologií a dalšími.

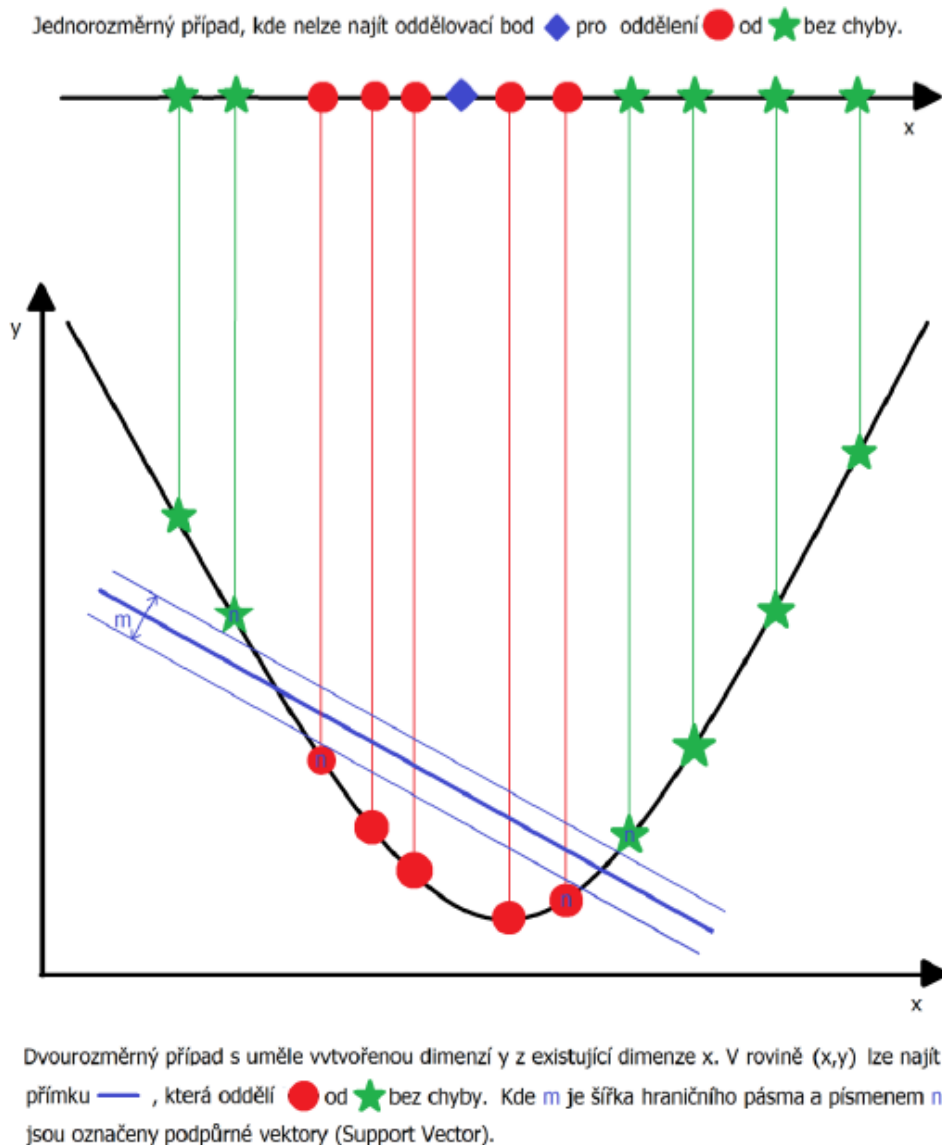
Princip trénování SVM spočívá ve vyhledání nejhůře klasifikovatelného vzoru ve vstupních datech. Takový vzor je nejvíce vzdálen od nadroviny a leží na její nesprávné straně. Parametry nadroviny se aktualizují tak, aby byl vzor na správné straně. Vzor pak tvoří jeden z nosných vektorů. Pro optimální nadrovinu platí, že musí být umístěna v co největším odstupě (angl. Maximal margin) od krajních bodů, nazývaných podpůrné vektory (angl. Support Vectors). Algoritmus se tedy snaží vytvořit z trénovacích dat co nejširší oddělovací pásmo. Širší oddělovací pásmo poskytuje lepší generalizaci klasifikátoru pro budoucí neznámé případy, a tudíž i velkou robustnost SVM. Šířka oddělovacího pásma závisí na výběru nosných vektorů a i jejich malá změna může šířku oddělovacího pásma ovlivnit, proto je důležitá co nejlepší volba trénovacích dat.

Je možné vybrat několik typů jader, které jsou základem SVM algoritmu. Pro lineární SVM bez mapování dat do vyšších dimenzí můžeme zvolit lineární jádro. Pro nelineární data lze zvolit jádro RBF (Radial basis function) nebo polynomiální jádro. Parametr γ používá jen jádro RBF a určuje nám, jak daleko od nadroviny budou mít body vliv na hranici. Při vysoké



hodnotě γ ovlivňují nadrovinu jen body, které jsou k ní blízko a vzdálené body nemají na nadrovinu vliv. Nadrovinu se může začít zakřivovat kolem jednotlivých bodů. Při nízké hodnotě γ nadrovinu ovlivňují i vzdálené body. Parametr C určuje kompromis mezi rovností nadroviny a počtem správně klasifikovaných tréninkových bodů. Pro vyšší hodnoty C se bude nadrovinu zakřivovat kolem jednotlivých bodů tak, aby všechny body zařadila do správné oblasti. Tímto však klasifikátor ztrácí schopnost zobecňovat. Při nižších hodnotách C bude nadrovinu sice rovná, algoritmus bude více obecný, ale může dojít k nesprávné klasifikaci některých bodů.

Nejlepší hodnoty parametrů γ a C nejsou známy předem. Musíme provést vyhledání těchto parametrů a k tomu se používá tzv. cross-validace.



Obrázek 1: Princip Support Vector Machine

Cíl:

Seznámení s SVM a jeho parametry a modalitami. Aplikace SVM na naměřená data. Vyhodnocení úspěšnosti detekce přítomnosti CMP a rozlišení varianty CMP.

Úkoly:

1. Příprava kapalných fantomů hlavy a cévních mozkových příhod
2. Ověření elektrických vlastností fantomů a shodu s tabulkovými hodnotami
3. Kalibrace mikrovlnného zobrazovacího systému
4. Nastavení měřicí sekvence
5. Měření pro prázdný systém a pro systém s fantom CMP postupně umístěnými na jednu z předpřipravených pozic
6. Analýza naměřených dat pomocí Support Vector Machines v prostředí MATLAB

Použité zařízení a přístroje:

- Sonda a SW pro měření dielektrických vlastností kapalin, Dielectric Assesment Kit, SPEAG, Švýcarsko
- Vektorový analyzátor obvodů, ZNB8, Rohde&Schwarz, SRN
- Přepínací mikrovlnná matice, ZN-Z84, Rohde&Schwarz, SRN
- Mikrovlnný zobrazovací systém
- 6ti portová automatická kalibrační sada, ZN-152, Rohde&Schwarz, SRN
- PC a MATLAB

Postup:

1) Příprava kapalných fantomů hlavy a cévních mozkových příhod

Podle dostupné literatury vypočítejte průměrné dielektrické vlastnosti lidské hlavy, ischemické a hemorhagické cévní mozkové příhody (CMP) na frekvenci vhodné k mikrovlnnému zobrazování hlavy [1, 2]

Lidská hlava $\epsilon_r =$

$\sigma =$

Hemorhagická CMP $\epsilon_r =$

$\sigma =$

Ischemická CMP $\epsilon_r =$

$\sigma =$

$$\sigma =$$

Použijte destilovanou vodu, alkohol (izopropanol) a NaCl pro vytvoření roztoků s dielektrickými parametry tkání. Výslednou kapalinou naplňte kulové formy představující ischemickou a hemorhagickou cévní mozkovou příhodu.

2) Ověření elektrických vlastností fantomů a shodu s tabulkovými hodnotami

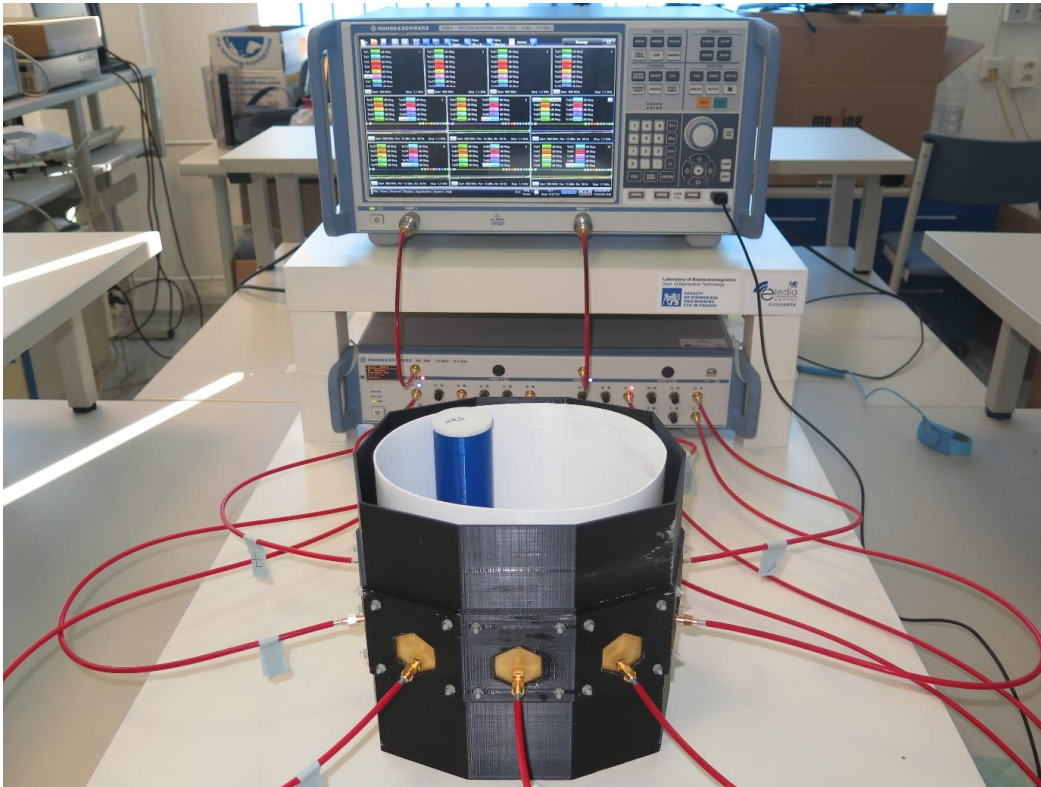
Systémem DAK (Dielectric Assessment Kit) s vektorovým analyzátozem změřte dielektrické parametry připravených roztoků a v případě potřeby parametry doladíte tak, aby relativní chyba dielektrických parametrů byla menší než 10 %.

3) Nastavení měřicí sekvence

Využijte předpřipravený ovládací skript pro měření s-parametrů mikrovlnného zobrazovacího systému. Inicializujte spojení MATLABu s vektorovým analyzátozem obvodů a přepínací maticí. Nastavte porty na přepínací matici, kam jsou připojeny jednotlivé antény mikrovlnného systému. Nastavte frekvenci měření, šířku pásma a další parametry. Navrhněte alespoň rozdílných 10 pozic, na které bude umístěn fantom cévní mozkové příhody a ovládací skript patřičně upravte.

4) Měření pro prázdný systém a pro systém s fantomem CMP

Do mikrovlnného systému napustíte připravenou kapalinu napodobující lidskou hlavu. Na předem připravené pozice postupně vložte fantomy cévních mozkových příhod. Provedte alespoň 10 měření pro ischemickou CMP a 10 měření pro hemorhagickou CMP, které budou umístěné na různé pozice uvnitř mikrovlnného zobrazovacího systému. Provedte 10 měření systému bez přítomnosti cévní mozkové příhody. Nakonec provedte několik měření pro oba typy CMP na náhodných pozicích. Tyto data budou sloužit ke klasifikaci neznámých dat. Jednotlivé měření uložte do souborů vhodných pro načtení do programu MATLAB.



Obrázek 2: Měřicí sestava skládající se z vektorového analyzátoru obvodů, přepínací matice a mikrovlnného zobrazovacího systému.

5) Analýza naměřených dat pomocí Support Vector Machines v prostředí MATLAB

- Úprava dat

Pro každou S-matici vypočítejte modul a fázi a data transformujte na 1 řádek. Je vhodné zmenšit množství dat oříznutím duplicitních hodnot v S-matici (Př. $s_{12} = s_{21}$). Ze všech měření vytvořte matici trénovacích dat (TrainData), která bude obsahovat na každém řádku jedno měření. K matici trénovacích dat vytvořte ekvivalentní matici tříd (TrainDataClass), obsahující jeden sloupec s označením dat. Měření s hemoragickou CMP označte číslem 0, měření s ischemickou CMP označte -1 a měření bez CMP označte číslem 0.

- Trénování algoritmu

Pro vytvoření natrénovaného modelu, podle kterého budeme klasifikovat data použijeme matici trénovacích dat (TrainData), matici tříd (TrainDataClass) odpovídající trénovacím datům a případně další nastavení pro konkrétní algoritmus.

Návod: $Model = FceAlgoritmu(TrénovacíData, MaticeTříd, ...)$

- Načtení neznámých dat

Postupujeme jako při tvorbě trénovacích dat. Matice třídy pro neznámých dat použijeme až k ověření přesnosti klasifikace.

- **Klasifikace dat**

Klasifikaci neznámých dat provedeme funkcí *predict* do které vstupuje natrénovaný model a neznámá data. Výstupem jsou klasifikované třídy.

Návod: $[KlasifikovanéTřídy] = predict(NatrénovanýModel, NeznámáData);$

- **Hodnocení úspěšnosti klasifikace**

Jednou z často používaných metod pro hodnocení klasifikace algoritmu je konfuzní matice (matice změn). Matice obsahuje ve sloupcích skutečnou hodnotu předpovídaného znaku a v řádcích předpověď klasifikátoru. Buňky matice obsahují četnosti klasifikace dané kombinace.

Návod: *confusionchart*

Determined Stroke Type	Hemorrhagic	54 50.0%	15 13.9%	78.3% 21.7%
	Ischemic	0 0.0%	39 36.1%	100% 0.0%
		100% 0.0%	72.2% 27.8%	86.1% 13.9%
		Hemorrhagic	Ischemic	
		True Stroke Type		

Obrázek 3: Konfuzní matice pro klasifikaci ischemické a hemoragické cévní mozkové příhody.

Autorské dílo