

## Úkol 1 – Explorace dat a lineární regrese – Arrhythmia data

### Cíl

Cílem prvního úkolu je seznámit se s daty z Arrhythmia studie a pomocí lineární regrese zjistit, jak se projeví změna tepové frekvence v délce EKG segmentů.

### Data

Arrhythmia data byla vytvořena pro účely klasifikace normálního průběhu EKG a 15 arytmií. Informace o pacientech spolu s průběhy 12-ti svodového EKG byly popsány pomocí 279 příznaků. Mezi nimi je prvních 15, které představují obecné patientské charakteristiky, a jejich popis je tabulce níže. Zbytek jsou parametry z jednotlivých EKG svodů popsané pomocí délek segmentů a jejich amplitud, a pak specifické jevy typické pro arytmiie. Popis dat pro svod DI v tabulce níže.

### Explorace dat

Data z point studie obsahují 279 příznaků. V rámci explorace dat prozkoumejte všechny příznaky, které se zabývají popisem pacientů (příznaky 1-15) a příznaky popisující délky segmentů v prvním svodu DI (příznaky 16-27) a jejich amplitudy (příznaky 160-169). Seznamte s typem dat v jednotlivých příznacích, rozsahem jejich hodnot, jejich rozdělením, případně počtem chybějících a odlehlých hodnot. Zvažte, jestli některé z příznaků není potřeba transformovat pro potřeby dalších analýz. Prozkoumejte závislosti mezi jednotlivými příznaky a zajímavé závislosti vizualizujte.

### Modelování pomocí lineární regrese

Při zvyšování tepové frekvence se délka trvání segmentů nemění stejným způsobem. Pomocí lineárních modelů prozkoumejte, jak se s měnící se tepovou frekvencí mění délka trvání QRS komplexu, P-R interval, QT interval, T interval a P interval. Zvolte jeden nebo více parametrů, které nejlépe vysvětlují změny v tepové frekvenci a modelujte tuto závislost pomocí lineárního modelu. Výsledný model zhodnoťte z hlediska přesnosti modelu a jeho statistické významnosti.

### Požadované kroky analýzy

Průzkumová analýza. [4 body]

- Dimenze dat (počet příznaků, počet instancí)?
- Chybějící a odlehlé hodnoty, transformace dat
- Vizualizace

Modelování vztahu mezi koncentrací kreatininu a tělesné hmotnosti [6 bodů]

- Formální zápis modelu
- Interpretace koeficientů modelu
- Přesnost modelu
- Statistická významnost

Výsledky upravte do formy zprávy, která bude obsahovat stručný **úvod**, popis metod, které jste použili, v sekci **metody**, výsledky jejich aplikace na data v sekci **výsledky** a závěry, které jste zjistili interpretací výsledků v sekci **závěr**.

1	Age	věk pacienta v letech
2	Sex	pohlaví (0 – muž/ 1 – žena)
3	Height	výška v cm
4	Weight	váha v kg
5	QRS_duration	průměrná délka QRSv ms
6	PR_interval	Průměrná doba mezi začátkem vlny P a vlny Q v ms
7	QT_interval	Průměrná doba mezi začátkem vlny Q a vlny T v ms
8	T_interval	Průměrná délka T vlny v ms
9	P_interval	Průměrná délka P vlny v ms

#### Úhly elektrické osy srdeční v intervalech

10	QRS	
11	T	
12	P	
13	QRST	
14	J	

15	Heart_rate	Počet tepů za minutu
----	------------	----------------------

#### Kanál DI:

Průměrná délka v ms

16	Q_wave	
17	R_wave	
18	S_wave	
19	R_small_wave	small peak just after R
20	S_small_wave	

21	Počet odchylek	
----	----------------	--

22	Existence_of_ragged_R_wave	Existence nepravidelné R vlny
23	Existence_of_diphasic_derivation_of_R_wave	Existence difázické R vlny
24	Existence_of_ragged_P_wave	
25	Existence_of_diphasic_derivation_of_P_wave	
26	Existence_of_ragged_T_wave	
27	Existence_of_diphasic_derivation_of_T_wave	

#### Kanál DII:

28 ... 39 (podobné 16 ... 27 kanálu DI)

atd.

Kanál DI

Amplituda \* 0.1 mV

160	JJ_wave
161	Q_wave
162	R_wave
163	S_wave
164	R_small_wave
165	S_small_wave
166	P_wave
167	T_wave

168	QRSA	Suma plochy všech segmentů dělená 10, (plocha= šířka * výška / 2 )
169	QRSTA	QRSTA = QRSA + 0.5 * šířka T vlny * 0.1 * výška T vlny

Kanál DII:

170 ... 179

atd.