Zadanie 3

Data regression

Bc. Peter Tokovics

Obsah

[2 Postup 3](#_Toc129710855)

[2.1 Importovanie datasetu 3](#_Toc129710856)

[2.2 Parsovanie datasetu 3](#_Toc129710857)

[2.3 Trénovacie a testovacie data 4](#_Toc129710858)

[2.4 Regressor 4](#_Toc129710859)

[2.5 Predpovedanie 4](#_Toc129710860)

[2.6 Vykreslenie 4](#_Toc129710861)

[3 Výsledky 5](#_Toc129710862)

[3.1 Veľkosť testovacej vzorky: 0.2 5](#_Toc129710863)

[3.2 Veľkosť testovacej vzorky: 0.5 6](#_Toc129710864)

[3.3 Veľkosť testovacej vzorky: 0.8 7](#_Toc129710865)

[4 Vplyv zmeny veľkosti testovacej vzorky 8](#_Toc129710866)

# Postup

Main funkcia, ktorá volá jednotlivé funkcie.

Obrázok, na ktorom je text

Automaticky generovaný popis

## Importovanie datasetu

Na začiatku bolo potrebné importovať dataset, ktorý sme použili pri trénovaní.

Obrázok, na ktorom je text

Automaticky generovaný popis

## Parsovanie datasetu

Po tom čo sme si importovali dataset ho je potrebné rozparsovať na množiny X a Y.

Obrázok, na ktorom je text

Automaticky generovaný popis

## Trénovacie a testovacie data

Po parsovaní použijeme množiny X a Y pri vytváraní trénovacích a testovacích dát.

Obrázok, na ktorom je text

Automaticky generovaný popis

## Regressor

Teraz si vypočítané regressor pomocou trénovacích dát.

Obrázok, na ktorom je text

Automaticky generovaný popis

## Predpovedanie

Následne si pomocou vypočítaného regresora budeme vedieť predpovedať známku.

Obrázok, na ktorom je text

Automaticky generovaný popis

## Vykreslenie

Následne si vieme natrénované a predpovedané dáta vykresliť.

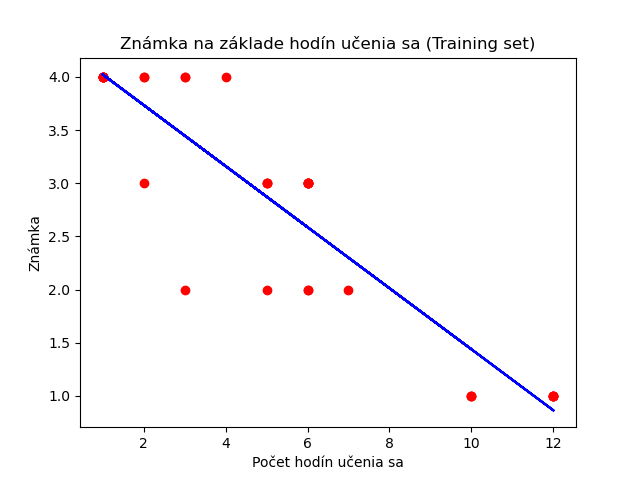
Obrázok, na ktorom je text

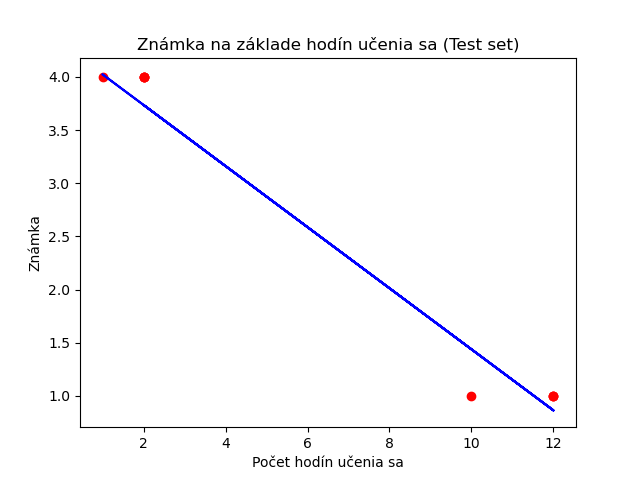
Automaticky generovaný popis

# Výsledky

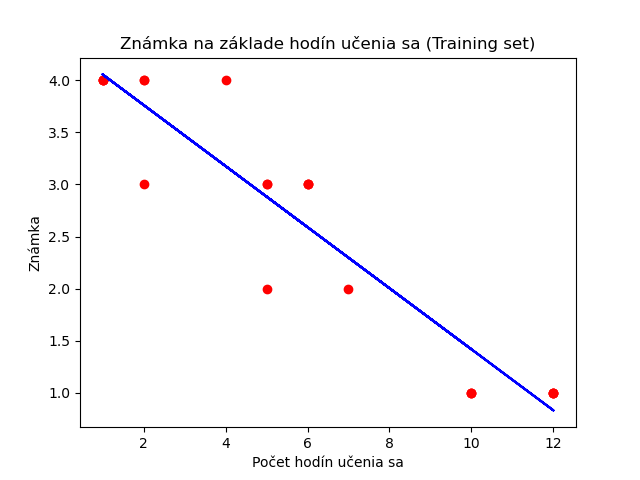
Boli vytvorené rôzne výsledky, pri ktorých bola menená veľkosť testovacej vzorky.

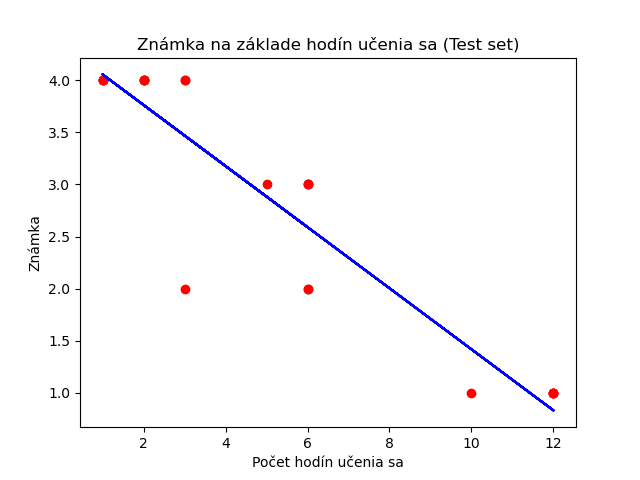
## Veľkosť testovacej vzorky: 0.2



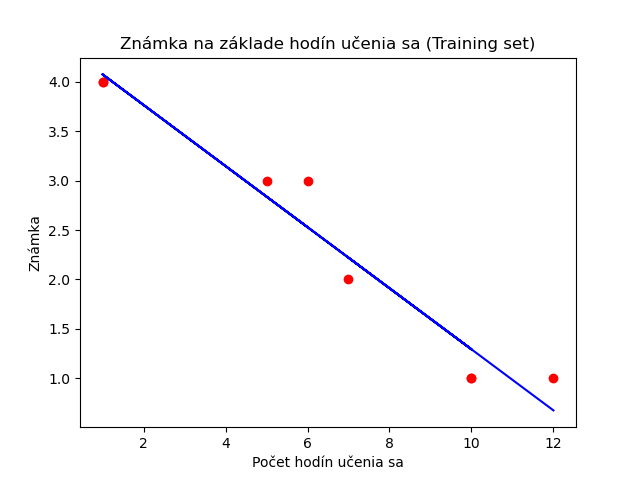


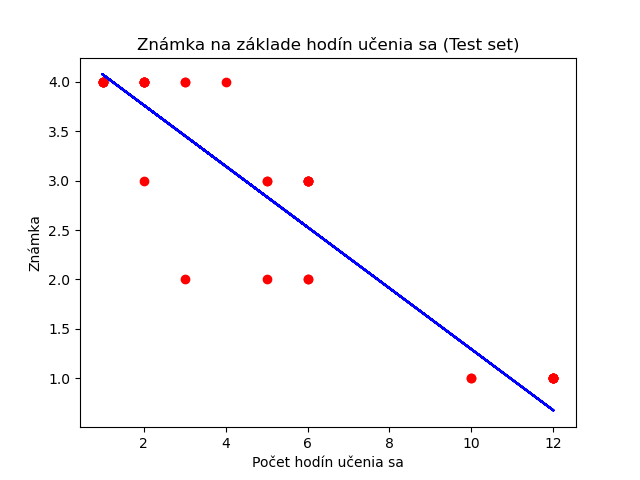
## Veľkosť testovacej vzorky: 0.5





## Veľkosť testovacej vzorky: 0.8





# Vplyv zmeny veľkosti testovacej vzorky

Tým že zväčšujeme veľkosť testovacej vzorky tým zmenšujeme aj veľkosť našich dát, na základe, ktorých budeme odhadovať. Tým sa stávajú naše testovacia dáta menej presné. Pomer veľkosti testovací a trénovacích dát je asi najlepší v prvom príklade a to 0.2.