



Mašinsko učenje 2023

Zadatak 1

Sadržaj



Podsetnik - praktičan deo



Zadatak 1



SciPy Stack



Uputstva i saveti

Podsetnik - praktičan deo

Podsetnik - praktičan deo

- Praktičan deo predmeta nosi najviše 60* bodova.
- Sastoji se od:
 - 6 domaćih zadataka
 - Predmetnog projekta.
- Akcenat na timskom radu:
 - Svaki član tima mora dati svoj doprinos
 - Bodovi dodeljeni članovima istog tima mogu da se razlikuju.

* u posebnim slučajevima 60 bodova donosi i dodatnih 40.

Podsetnik - praktičan deo

- Opcije:
 - Samo projekat = najviše 25 bodova
 - 3 domaća zadatka (najviše 25) + projekat (najviše 25) = najviše 50 bodova
 - 4+ domaćih zadataka (najviše 35) + projekat (najviše 25) = najviše 60 bodova
 - Nagrada za najuspešnije = najviše 60 bodova.
 - Najuspešniji od najuspešnijih = 100 bodova.

Podsetnik - praktičan deo

- Kriterijumi:
 - Ostvareni rezultati i kako se do njih došlo:
 - Pristup problemima
 - Korišćeni algoritmi
 - Određivanje (hiper)parametara algoritama
 - Rad sa trening skupom podataka.
 - Prpratni izveštaji:
 - Sadržaj prpratnih izveštaja
 - Usklađenost izveštaja i izvornih kodova rešenja.
 - Diskusija:
 - Prezentovanje rešenja i odgovori na pitanja prilikom prezentovanja.

Podsetnik - praktičan deo

- Raspored domaćih zadataka:
 - 06.03. - 17.03. Jednostruka linearna regresija
 - 20.03. - 06.04. Višestruka regresija
 - 11.04. - 25.04. SVM
 - 03.05. - 12.05. Ansambl klasifikatora
 - 15.05. - 26.05. Klasterovanje
 - 29.05. - 04.06. PCA

Zadatok 1

Zadatak 1

- Jednostruka linearna regresija:
 - Upotrebom jednostruke linearne regresije prediktovati **Y** na osnovu **X**.
 - Zadatak je uspešno urađen ukoliko se na kompletnom testnom skupu podataka dobije **RMSE (Root Mean Square Error)** manji od 126.
 - Algoritmi mašinskog učenja se samostalno implementiraju - **zabranjena upotreba algoritama iz biblioteka**.
 - Rok za izradu zadatka je **17.03.2023. u 23:59h**.
 - Instalirane biblioteke za Zadatak 1:
 - NumPy
 - Pandas.
 - Sledeći termin vežbi (odbrana Zadatka 1 i predstavljanje Zadatka 2) je u nedelji **20.03. - 24.03.2023.**

Zadatak 1

- Koncepti vezani za Zadatak 1 (podsetiti se gradiva sa predavanja i ranijih predmeta):
 - Gradient Descent (Batch vs Stochastic)
 - Normal Equation
 - Outlier-i i ostale tačke visokog uticaja (high-leverage points)
 - Rad sa (trening) skupom podataka.

Zadatak 1

- Metrika se računa na osnovu sledeće formule:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

SciPy Stack

A thick, solid black horizontal bar spanning the entire width of the image, located at the bottom.

Zadatak 1

- Za izradu zadataka koristiti **Python 3.5.x**.
- Preporuka da se prilikom izrade zadataka oslonac bude SciPy Stack i njegove biblioteke:
 - NumPy
 - SciPy
 - Matplotlib
 - Jupyter
 - Pandas.

Zadatak 1

- Za Zadatak 1 na platformi su instalirane biblioteke (verzije date u Uputstvu):
 - **NumPy:**
 - [Docs](#)
 - [Stanford Tutorial](#)
 - **Pandas:**
 - [Docs](#)
 - [Tutorial](#)
 - [10 Minutes to Pandas](#).
- Za potrebe vizualizacije podataka i pisanja propratnog izveštaja od pomoći može biti biblioteka **Matplotlib**:
 - [Docs](#)
 - [Tutorial](#).

Uputstva i saveti

Zadatak 1

- Uputstvo za rad sa platformom i pisanje propratnog izveštaja se nalazi u:
 - **Files/Vežbe/Uputstvo.pdf.**
- Saveti za rešavanje zadataka:
 - Podsetiti se gradiva sa predavanja
 - Detaljno pročitati uputstvo za rad sa platformom i pisanje propratnog izveštaja
 - Vizualizacija podataka
 - Isprobati više pristupa - podeliti zaduženja tako da svaki član tima implementira jedan pristup. Nakon toga, zajedno analizirati implementirano i odabrati najbolji pristup koji će se evaluirati na platformi.
 - Ako se radi normalizacija podataka, obratiti pažnju kako će se računati RMSE metrika.

Zadatak 1

- Savet za implementaciju:
 - Metoda **fit(x, y)** za “fitovanje” trening podataka
 - Metoda **predict(x)** za predikciju vrednosti testnog skupa
 - Metoda **calculate_rmse(y_true, y_predict)** za računanje RMSE na osnovu date formule.