



Трећи пројектни задатак из предмета Интелигентни системи школске 2024/25. године

Основне информације

Циљ пројектног задатка је реализација модела алгоритма линеарне регресије за предвиђање цена некретнина и реализација модела за класификацију колача према њиховом саставу користећи стабло одлучивања.

1. Линеарна регресија

Неопходно је реализовати модел алгоритма линеарне регресије са више променљивих на примеру предвиђања цена некретнина. Скуп података који се користи (*house_prices_train.csv*) је дат на предметном сајту. Атрибути скупа података су следећи:

- *year_built* - година изградње.
- *area* - укупна површина свих соба у кући изражена у квадратним стопама.
- *bath_no* - број купатила.
- *bedroom_no* - број спаваћих соба.
- *price* - цена изражена у доларима.

Решење треба да прати следеће кораке:

- учитавање скупа података и приказ првих и последњих пет редова у табели.
- приказ концизних информација о садржају табеле и статистичких информација о свим атрибутима.
- графички приказ зависности излазног атрибута од сваког улазног атрибута расејавајући тачке по Декартовом координатном систему.
- одабир атрибута који учествују у тренирању модела.
- извршавање додатних трансформација над подацима табеле, по потреби.
- реализација алгоритма градијентног спуста за одређивање параметара модела без коришћења уграђених модела.
- реализација модела алгоритма линеарне регресије користећи неки од уграђених модела.
- приказ свих добијених параметара модела, вредности функције грешке и прецизности (*score*) модела за оба реализована приступа. Раздвојити скуп података на скупове за тренирање и валидацију.



2. Стабло одлучивања

Неопходно је реализовати модел за класификацију колача према њиховом саставу користећи стабло одлучивања. Скуп података који се користи (*cakes_train.csv*) је дат на предметном сајту. Сваки ред у скупу података представља рецепт за један од датих колача. Атрибути скупа података су следећи:

- *flour* - количина брашна коришћена за колаче изражена у грамима.
- *eggs* - количина јаја коришћена за колаче (сматрати да једно јаје има масу од 63 грама).
- *sugar* - количина шећера коришћена за колаче изражена у грамима.
- *milk* - количина млека коришћена за колаче изражена у грамима.
- *butter* - количина путера коришћена за колаче изражена у грамима.
- *baking_powder* - количина прашка за пециво коришћена за колаче изражена у грамима.
- *type* - тип колача који може бити *muffin* или *cupcake*.

Имати у виду да рецепти у скупу не производе исту количину колача: неки рецепти су за мању количину, неки за већу количину колача.

Решење треба да прати следеће кораке:

- учитавање скупа података и приказ првих и последњих пет редова у табели.
- приказ концизних информација о садржају табеле и статистичких информација о свим атрибутима.
- графички приказ зависности излазног атрибута од сваког улазног атрибута расејавајући тачке по Декартовом координатном систему.
- графички приказ корелационе матрице свих атрибута.
- одабир атрибута који учествују у тренирању модела.
- извршавање додатних трансформација над подацима табеле, по потреби.
- реализација модела алгоритма стабла одлучивања користећи неки од уграђених модела.
- приказ добијеног стабла одлучивања, вредности функције грешке и прецизности (*score*) модела. Раздвојити скуп података на скупове за тренирање и валидацију.



Универзитет у Београду - Електротехнички факултет
Катедра за рачунарску технику и информатику

Напомене

Електронску верзију решења овог домаћег задатка послати као ZIP архиву на следећем [линку](#). Одбрана домаћег задатка планирана је у сваком испитном року ове школске године.

Домаћи задатак из предмета *Интелигентни системи* се ради самостално и није обавезан за полагање испита (на испиту се може заменити са теоријским питањима из целокупног градива или рачунским задацима из области машинског учења). Овај домаћи задатак се може бранити само у за то предвиђеним терминима. Домаћи задатак вреди максимално 15 поена.

На усменој обрани кандидат мора самостално да покрене своје решење. Кандидат мора да поседује потребан ниво знања о задатку, мора да буде свестан недостатака приложеног решења и могућности да те недостатке реши. Кандидат мора тачно да одговори и на одређен број питања која се баве тематиком домаћег задатка.

Евентуална питања послати асистентима на мејл, али као једну поруку, а не две одвојене (другог асистента обавезно ставити у копију - СС поруке).

aki@etf.rs

jocke@etf.rs