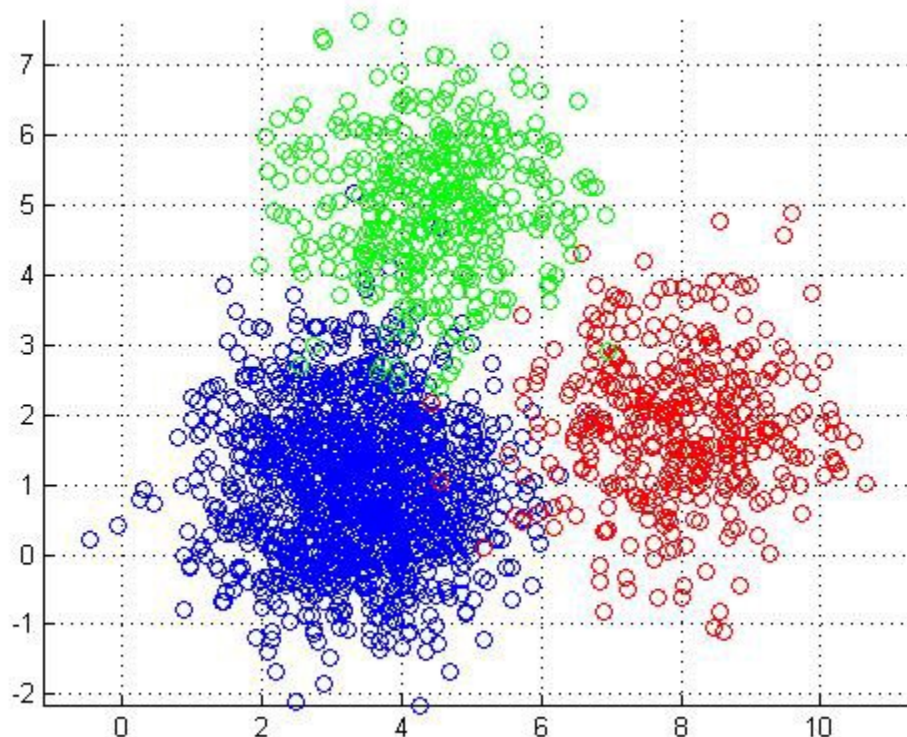


ИЗВЕШТАЈ ЗАДАТКА 5

Решавање класификационог проблема кластеровањем



Душан Бркић, Филип Живанац

4. цветња лета Господњег 2022.

Софтверско инжењерство и информационе технологије

Факултет техничких наука

Универзитет у Новом Саду

ЗАДАТАК

Кластеровати државе на основу њихових карактеристика у кластере који представљају географске регионе (регион): Африка, Америка, Азија и Европа. Задатак је успешно урађен уколико се на комплетном тестном скупу података добије в мера (енг. *v measure score*) већа од 0.40.

ПРИСТУП ПРОБЛЕМУ

За решење проблема смо на почетку торке које садрже празну вредност уклонили. Затим смо претворили сва класна обележја у бројеве.

ИСПРОБАНЕ МЕТОДЕ

УКЛАЊАЊЕ АУТЛАЈЕРА

Аутлајере смо ручно уклањали тако што смо приметили неколико да само неколико држава из Америке, Азије и Африке имају четвороцифрену плату за разлику од осталих које имају троцифрену. Такође само једна држава у Европи има троцифрену плату за разлику од осталих које имају четвороцифрену. Због малог обима скупа података ове промене донеле су значајно побољшање.

РЕЗУЛТАТИ

Модел гаусових мешавина

Пошто је у задатку неопходно да се проблем решава једино уз помоћ модела гаусових мешавина једини параметарр ког се трудимо да подесимо је насумично стање. Пошто превише вредности долази у обзир да би се ручно проверавало решење свих користили смо алгоритам ПСО за оптимизацију. ПСО у овом случају оптимизује функцију која прима као улаз вредност насумичног стања а враћа в скор као излаз.

ОДАБРАНО РЕШЕЊЕ

На основу резултата које нам је дао ПСО најбоље насумично стање је 38953 које даје резултат 0.730318666952313.