**Zadanie č. 2**

Strojové učenie a neurónové siete

## Zadanie

Algoritmy strojového učenia obsahujú veľa hyper parametrov, ktoré sa nastavujú bez exaktných pravidiel a s ohľadom na vstupné dáta. Predtým, než začnete aplikovať prvé algoritmy strojového videnia je preto potrebné poznať svoje dáta. V rámci ďalších zadaní sa budete snažiť, aby váš systém určoval fotografie rovnakých ovocí ako podobné a iných ovocí ako odlišné - preto je dobré vedieť, nakoľko sú si dáta podobné pixel po pixeli v rámci jednotlivých tried aj medzitriedne.

## Úloha č. 1

Cieľ: zistite, na koľko sú si podobné dáta v rámci jednotlivých tried

Vstup: uložené binárne dáta pre jednotlivé ovocia

Postup: načítajte dáta z predošlého zadania (uložené podľa jednotlivých tried). Porovnajte jednotlivé vzorky medzi sebou:

A) vzdialenosťou pixelov (L2,Manhattan)

C) vzhľadom na ich vzdialenosť od priemernej vzorky pre danú triedu

Analyzujte získané výsledky - ktorá trieda je najviac rozmanitá, ktorá najmenej (odporúčame štatistickou analýzou výsledkov).Prenesú sa tieto vlastnosti aj po vašom delení na trénovacie/validačné/testovacie množiny?

Výstup: zrozumiteľné zobrazenie výsledkov (odporúčame formou grafov a základných štatistík).Postup aj získané grafy dobre opíšte v dokumentácii.

## Riešenie

## Úloha č. 2

Cieľ: zistite, nakoľko sú si podobné dáta v rozdielnych triedach.

Vstup: uložené binárne dáta pre trénovaciu množinu (testovaciu, validačnú).

Postup: načítajte dáta z predošlého zadania (uložené do množín pre ďalšiu prácu). Zistite, ktoré vzorky (triedy) sú si v rámci tejto množiny navzájom najpodobnejšie. Využívať nato budete zhlukovacie algoritmy:

A) k-means clustering

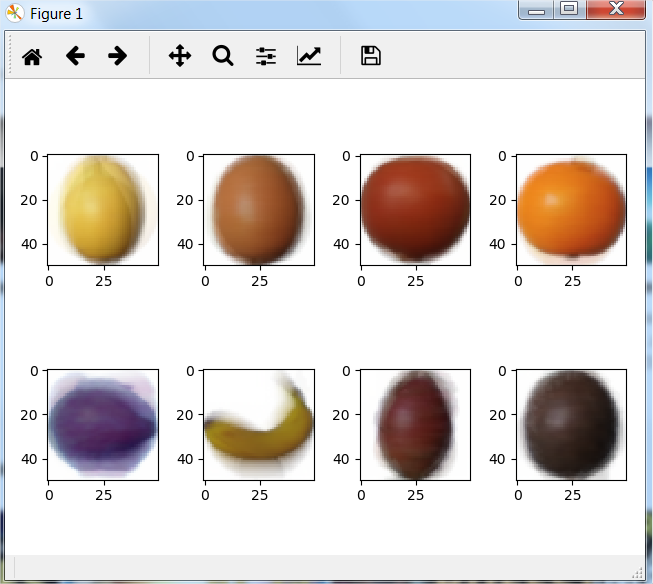
B) DBSCAN/ Chinese whispers

C) SOM (za dva bonusové body)

Ktoré vám vrátia skupiny(zhluky, clustre) navzájom podobných obrázkov v rámci vášho datasetu. Pre každý získaný zhluk nakoniec zobrazte centrum zhluku (priemerný obrázok) a vypíšte, aké je triedne zloženie pre daný zhluk(v prípade veľkého množstva zhlukov zobrazte len vhodnú podmnožinu).Pokúste sa analýzou týchto dát zodpovedať na otázku, ktoré ovocia sú si navzájom najpodobnejšie.

Výstup: zrozumiteľné zobrazenie výsledkov – príklady zobrazenia centier aj príslušníkov niekoľkých zhlukov a ku nim prislúchajúce triedne zloženie. Postup aj obrázky dobre opíšte v dokumentácii. Uveďte, ktoré ovocia sú si najviac podobné a ako ste dospeli ku takému záveru.

## Riešenie



Výstup z KMeans

