

Zadanie č.1

Cieľom prvého zadania je naučiť študentov, ako pracovať s dátami tak, aby bolo možné s nimi pracovať aj neskôr. V tejto úlohe použijeme databázu [fruit-360](#). Databáza obsahuje farebné obrázky ovocia. V databáze je 81 tried, z ktorej každá obsahuje cca 300 snímok. Najprv je potrebné si tieto data stiahnuť k sebe na lokálny počítač. Trénovacie dáta obsahujú cca 42 000 obrazov, zatiaľ čo testovacie dáta obsahujú len 14.000 znakov.

1. Úloha

Nahliadnite do dát a zistite, či stiahnuté data sú zmysluplné. Každý z obrázkov by mal obsahovať ovocie. Zobrazte príklady obrázkov, ktoré ste si práve stiahli. (Z každého druhu ovocia aspoň 4 obrázky).

2. Úloha

Pripravte dáta do formátu s ktorým bude možná jednoduchšia manipulácia. Pre každú triedu vytvorte samostatný súbor, ktorý bude možné spracovať separátne. Neskôr tieto súbory spojíme do jedného manažovateľného súboru. Celý dataset konvertujeme do poľa (index, x, y, ch), ktoré bude obsahovať float hodnoty, pričom dáta budú normalizované. Tzn. priemer bude približne 0 a štatistická odchýlka bude $\sim 0,5$ aby sme uľahčili tréningovanie.

Súbory, ktoré nebude možné prečítať preskočíme.

3. Úloha

Verifikujte, či dáta po transformácii sú stále korektné. Zobrazte obrázky a priradte k nim aj ich názov.

4. Úloha

Dáta majú dobré rozloženie ak každá trieda obsahuje približne rovnaký počet obrázkov. Overte, či nami vybrané dáta sú rovnomerne rozložené.

5. Úloha

Spojte pripravené do jedného súboru. V závislosti od vášho počítača, tak aby bolo možné súbor načítať do pamäte, nastavte počet trénovacích vzoriek. Označenia tried zakódujte číslicami indexmi. Tiež vytvorte validačnú množinu. Takto pripravené dáta náhodne preusporiadajte.

6. Úloha

Overte, či dáta sú stále korektné aj po preusporiadaní.

7. Úloha

Dáta uložte tak, aby ich bolo možné znova rýchlo načítať.

Súčasťou zadania je aj vypracovaná dokumentácia. Cieľom dokumentácie je dokázať, že ste zadanie splnili.