



Ročník: 19

Kolo: 2

Úloha: 3

Úloha o rozdeľovaní zoznamu dát

Boom umelej inteligencie a strojového učenia zasiahol aj obyvateľov Korytnáčkova. Nejaký čas sa nadchýnali možnosťami jazykových modelov (ako napr. TouchGPT), ale potom začali uvažovať, ako to využiť vo svoj prospech.

Jednou z aplikácií, ktorú začali vyvíjať je predikcia ochorenia pacientov na základe popisu ich stavu. Jadrom aplikácie je počítačový model, ktorý na základe popisu stavu pacienta dokáže predikovať, či pacient je pozitívny alebo negatívny na nejaké ochorenie. Aby sme takýto model mohli použiť, musíme ho najskôr natréňovať. To sa robí tak, že mu dáme niekoľko príkladov, na ktorých sa bude učiť. Na nejakých iných príkladoch potom overíme, ako dobre sa model naučil predikovať ochorenie.

Príklady majú tvar dvojprvkových zoznamov, kde prvý prvok je popis stavu pacienta a druhý informácia, či je pacient pozitívny (1) alebo negatívny (0). Príklady môžu vyzeráť takto:

```
['bolí ho hlava, teplota 38, cíti sa unavený, bolia ho kĺby a svaly, pracuje v kolektíve', 1]  
['začal navštevovať fitness, teplotu nemá, cíti sa unavený, bolia ho svaly', 0]
```

Dostupné príklady najskôr náhodne rozdelíme do dvoch zoznamov – tréningového a testovacieho. Tréningový zoznam typicky obsahuje 70 % až 80 % zo všetkých príkladov a testovací zvyšok, teda 20 % až 30 % príkladov. Dôležité je, aby pomer pozitívnych a negatívnych pacientov v týchto zoznamoch bol rovnaký ako pomer pozitívnych a negatívnych pacientov medzi všetkými príkladmi. Tu však programátori narazili na problém ako realizovať rozdelenie. Vedel by si im pomôcť?

Vytvor funkciu **rozdel_priklady**, ktorá zadané príklady rozdelí podľa zadaného počtu percent do tréningového a testovacieho zoznamu. Výsledné rozdelenie funkcia vráti. Testovacie príklady zadáme ako zoznam príkladov a výsledok očakávame ako dva zoznamy príkladov. Ak by sa príklady vzhľadom na svoj počet nedali presne rozdeliť, akceptujeme aj rozdelenie s toleranciou +- jeden príklad.

Svoje riešenie ulož do súboru **strojove_ucenie.py**.