

Класификация на гъби

Проекта

Да се предвиди дали дадени гъби са отровни или не.

Данни

Имате набор от данни с информация за различни видове гъби, както и дали те са отровни или не.

Задание

Създайте модел на изкуствена невронна мрежа на **PyTorch**, който да предвижда дали определени гъби са отровни.

Проектът да бъде написан в Google Colaboratory notebook и да съдържа:

- Описание на избрания модел.
- Код за четене и обработка (pre-processing) на входните данни за обучение на мрежата.
- Код за обучение и тестване на създадения модел по дадени входни данни за обучение и тестване.
- Сравнение на различни модели и параметри за обучение като брой и големина на скрити слоеве, брой епохи, скорост на обучение / learning rate, както и таблица, сравняваща различните резултати и показваща каква е точността (accuracy) на различните модели.
- Графики на загубата и точността на моделите.
- Заключение, кой модел и параметри работят най-добре за решаване на проблема.
- Да се създаде изходен файл **result.csv**, който да съдържа предвидените стойности от тестовия набор от данни.

Точността (accuracy) на моделите да се изчислява като процент - колко от предвижданията съвпадат с етикетите в набора от данни за валидация. Полученият резултат да се превърне в десетична дроб между 0 и 1 (0.5 означава 50% точност, 0.95 – 95% точност).

Набори от данни (datasets)

Дадени са два файла:

- Набор от данни за обучение (training dataset) http://ml.monov.eu/mushrooms/train_data.csv, който съдържа общо 21 колони: колона “**id**” с идентификатор на гъбите, колона “**poisonous**” с информация дали гъбата е отровна или не (**y** – гъбата е отровна, **n** – гъбата не е отровна) и 19 колони с различни параметри (свойства / features) на гъбите. Примерни данни:

id	poisonous	cap-diameter	ring-type	habitat	season
134	y	4.99		f	h	a
962	n	19.53		f	d	a
1327	y	3.84		f	p	a

- Набор от данни за тестване (testing dataset) http://ml.monov.eu/mushrooms/test_data.csv, който съдържа 20 колони със същите полета като данните за обучение, без полето “**poisonous**”. Примерни данни:

id	cap-diameter	cap-shape	ring-type	habitat	season
78	4.93	s		f	d	a
1368	8.97	x		p	d	a

Резултатът **result.csv** да съдържа 2 колони със следните полета:

id – идентификатор на гъбата от тестовите данни

poisonous – да съдържа дали съответната гъба е отровна или не (**y** или **n**)

Примерен резултат:

id	poisonous
128	y
2032	n

Точността на генерирания резултат (**result.csv**) може да бъде тествана на следния адрес:
<https://ml.monov.eu/mushrooms/test/>.