**Uczenie maszynowe**

Zasady ćwiczeń:

1. Format rozwiązania: notebook (.ipynb). Widzimy kod, rozwiązanie, a jak trzeba grafikę.
2. Wykonane skrypty wkładamy do Zadania na platformie Moodle (czas jest ograniczony).
3. Na zajęciach należy zreferować rozwiązania (proszę, żeby nie było sytuacji, że nikt nie referuje).

Lab: **5** Schemat uczenia + metryki

Jeśli metodykę któregoś z etapów pracy można uogólnić, taki etap proszę zakodować w postaci metody  
(z ew. parametrami).

1. Zakodować schemat uczenia
   1. Hold-out: (b/duże zbiory danych) jednokrotny podział na część treningową i testową. Wyliczane są jednokrotnie metryki na zbiorze testowym.
   2. k-fold CV (Cross-Validation): (średnie zbiory danych) k-krotna walidacja krzyżowa czyli k-krotny podział zbioru (domyślnie k=10) i uczenie/testowanie wg. schematu: (k-1)-części do uczenia, jedna wyłączona część do testowania. Obliczane metryki są uśredniane po wszystkich iteracjach.
   3. LOO (Leave-One-Out): (b/małe zbiory danych) walidacja krzyżowa, w której k=(liczba danych). Oznacza to wyłączenie w każdej iteracji jednego przykładu do testowania i uczenie na pozostałych. Metryki obliczone w każdej iteracji są uśredniane.
2. W metodzie Hold-out i k-fold CV powinna być możliwość sparametryzowania:
   1. metody przydziału do zbioru testowego: losowo lub wg. naturalnej kolejności danych
   2. klas objętych schematem: wszystkie lub wybrane (np. wg. nazwy/indeksu)
3. W metodzie LOO powinna być możliwość sparametryzowania klas objętych schematem: wszystkie lub wybrane (np. wg. nazwy/indeksu).
4. Zademonstrować działanie przygotowanego schematu (podział, liczność zbiorów) na dowolnym zbiorze danych przeznaczonym do klasyfikacji np. Iris (https://archive.ics.uci.edu/ml/machine-learning-databases/iris/). Sama metoda uczenia/testowania może być pusta.
5. W pewnym klasyfikatorze binarnym dla etykiet przewidziano:

. Narysuj krzywą ROC i oblicz AUC.

1. Dla dowolnych danych i dowolnego klasyfikatora wyliczyć i narysować macierz błędu.
2. Zakodować w Pythonie własną metodę do rysowania krzywej ROC i wyliczania puktu optymalnego (punkt na krzywej ROC najbliższy: czułość=100%, specyficzność=100%). Przetestować na dowolnym klasyfikatorze lub wymyślonym zestawie:

etykiety\_prawdziwe (y)/etykiety\_przewidywane (ypred).

1. Dla dowolnych danych i dowolnego problemu regresji wyliczyć typowe metryki: MSE (Mean Squared Error) i MAE (Mean Absolute Error).

Przykładowe archiwa danych:

www.kaggle.com/datasets

archive.ics.uci.edu/ml

datasetsearch.research.google.com