Pracownia 1. Prowadzenie prostych eksperymentów klasyfikacji

Zadanie 1.1.

Wykorzystując bibliotekę `scikit-learn` wygeneruj syntetyczny problem zadania klasyfikacji, spełniający następujące wymagania: - problem składa się z dwóch atrybutów, - oba atrybuty problemu są informatywne, a więc nie ma cech redundantnych ani zbędnych, - problem jest dychotomią, - szum etykiet (błędnych przypisań klasy) stanowi 5% ogółu wzorców, - próbka problemu składa się z trzystu wzorców równomiernie rozłożonych po jego klasach.

Wygenerowany problem zapisz w formie pliku CSV, którego ostatnia kolumna zawiera etykiety oraz wygeneruj **scatterplot** prezentujący rozkład jego obiektów w przestrzeni cech. Kolor obiektów uzależnij od klasy, do której należą.

Dokumentacja:
- `make_classification`[1]
- `scatterplot`[2]

Przypilnuj, aby przy każdym uruchomieniu skryptu był generowany dokładnie ten sam zbiór danych, a nie tylko zbiór danych spełniający te same wymagania.

Zadanie 1.2.

Zrealizuj w skrypcie następujące podpunkty:

1. Podziel wygenerowany w pierwszym zadaniu zbiór danych na część testową i uczącą, przyjmując 30% do testowania i 70% do uczenia.
2. Zainicjalizuj gaussowski, naiwny klasyfikator Bayesa ze standardowymi hiperparametrami i wyucz (dopasuj) go na podstawie zbioru uczącego.
3. Wyznacz macierz wsparć dla zbioru testowego wyuczonego klasyfikatora.
4. Na podstawie macierzy wsparć wyznacz predykcję klasyfikatora dla zbioru testowego.
5. Na podstawie wyznaczonej predykcji wylicz wartość metryki accuracy klasyfikatora.
6. Na podzielonej na dwie części ilustracji, w formie scatterplota, przedstaw wsparcia klasyfikatora wyznaczone na zbiorze testowym (wzorce w dziedzinie wsparć). Dla lewej ilustracji przyjmij kolory dla etykiet rzeczywistych, dla prawej – etykiet będących wynikiem predykcji.

Dokumentacja:
- `train_test_split`[3]
- `accuracy_score`[4]

Zadanie 1.3.

W tym zadaniu musisz zrealizować poprawny eksperyment z wykorzystaniem stratyfikowanej walidacji krzyżowej. Pamiętaj, że wszystkie zadania podczas kolejnych pracowni będziesz wykonywać zgodnie z tą metodyką, nawet, jeśli nie jest to wskazane wprost.

1. Przygotuj obiekt stratyfikowanej walidacji krzyżowej z pięcioma foldami.
2. Przygotuj zmienną, w której będziesz przechowywać wyniki eksperymentu. Pięciofoldowa walidacja krzyżowa generuje dla każdego algorytmu pięć wyników.
3. W każdej pętli walidacji krzyżowej:
   - zainicjalizuj klasyfikator bazowy (gaussowski naiwny klasyfikator Bayesa),
   - zbuduj model klasyfikatora (wykorzysując zbiór uczący),
   - wyznacz predykcję (wykorzystując zbiór testowy),
   - oszacuj jakość modelu (metryką accuracy),
   - zapisz wynik w odpowiednim polu wektora przygotowanego w punkcie 2.
4. Wylicz wartość średnią i odchylenie standardowe uzyskanych wyników.

Dokumentacja:
- StratifiedKFold[5]

[1] <https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_classification.html>
[2] <https://matplotlib.org/3.1.1/api/_as_gen/matplotlib.pyplot.scatter.html>
[3] <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html>
[4] <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html>
[5] <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html>