

Workshop 1 — zadania

1

Plik `1.txt` zawiera macierz z trzema kolumnami — każda z nich to cena otwarcia pewnej firmy. Najczęściej wykorzystywanym przekształceniem w przypadku cen akcji jest wyznaczenie zwrotów logarytmicznych: jeśli x_1, x_2, \dots, x_n to ceny w kolejnych odstępach czasowych, to kolejne zwroty przedstawiamy za pomocą wzoru:

$$r_i = \ln \frac{x_{i+1}}{x_i}, \quad i = 1, 2, 3, \dots, n - 1.$$

Korzystając z funkcji `FunctionTransformer` z `sklearn.preprocessing` stwórz transformer kolumn, który zwracałby macierz zwrotów logarytmicznych.

Sprawdź implementację na macierzy z pliku `1.txt`. Przedstaw dane przed i po transformacji na wykresach.

Wskazówka: może zrobicie to lepiej ode mnie, no ale wspomnę, że skorzystałem z funkcji `apply_along_axis` z `numpy`.

Do zadań z danymi warto przygotować porządną EDE — pozwala to zorientować się, które zmienne są ważne, jakich transformacji należy użyć, etc.

2

Za oknem szklana pogoda? Problem klasyfikacji czeka! Pobierz [dane](#) odnośnie różnych klas szkła i rozpocznij prognozy. Wykorzystaj dowolny model do klasyfikacji dostępny w bibliotece scikit-learn (uwaga na typy zmiennych i ewentualne braki danych). Dokonaj ewaluacji modelu na kilku wybranych [metrykach](#). Zastanów się dlaczego niektóre metryki mogą zwracać lepsze wyniki od innych? Jak możesz poradzić sobie z tym problemem?

3

Dla danych z [kaggle](#) dotyczących klasyfikacji gwiazd zastosuj wybraną metodę redukcji wymiaru. Przedstaw dane o zredukowanym wymiarze na scatterplocie (najlepiej z zaznaczonymi typami). Co można wywnioskować?

4

Dobierz dwa modele klasyfikacyjne do danych z wcześniejszego zadania i porównaj jak sobie poradziły.