

Lista 8

Python: regresja liniowa i nieliniowa

Wszystkie zadania dotyczące modelowania proszę wykonywać w bibliotece statsmodels.

1. Wczytaj dane data_health, oraz zapoznaj się z plikiem readMe. Przygotuj dane do dalszej analizy.
2. Wyznacz współczynniki korelacji, pomiędzy zmiennymi ciągłymi oraz sprawdź ilu pacjentów choruje na każdą z chorób.
3. Używając biblioteki matplotlib, stwórz wykres obrazujący zależność pomiędzy poziomem hemoglobiny, a wiekiem.
4. Używając biblioteki matplotlib, stwórz wykres obrazujący rozkład wieku.
5. Stwórz model, który pozwoli na przewidzenie wieku na podstawie poziomu hemoglobiny i GFR. Przedstaw i zinterpretuj otrzymany model.
6. Wybierz losowo 15 wierszy z kolumn Hb i GFR, następnie dokonaj predykcji wieku. Sprawdź czy przewidziane wartości są zbliżone do prawdziwych.
7. Stwórz model, który jak najlepiej przewidzi płeć pacjenta.
8. Wybierz losowo 10 wierszy z danych i dokonaj na nich predykcji.