

Regresja wieloraka liniowa. Trenowanie i ewaluacja modeli – część 2

Do wykonania tych ćwiczeń będziesz potrzebować wyników z poprzednich zajęć (Trenowanie i ewaluacja modeli – część 1).

1. Utwórz najprostszy model bazowy średniej rocznej wielkości opadów. Model bazowy odpowiada modelowi bez żadnej zmiennej objaśniającej. Porównaj uzyskany wynik obliczając średnią wartość ze średniej ilości opadów na każdej stacji.

2. Dla utworzonego modelu bazowego oblicz błąd średniokwadratowy (RMSE) pomiędzy wartościami obserwowanymi a przewidywanymi.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (predicted_i - observed_i)^2}{n}}$$

3. Zastosuj funkcję `predict()` na testowym zbiorze danych dla modelu bazowego. Oblicz błąd średniokwadratowy (RMSE) pomiędzy wartościami obserwowanymi a przewidywanymi.

4. Przedstaw na wykresach wartości średniej rocznej wielkości opadów dla zbioru treningowego oraz testowego z nałożonym modelem liniowym z ćw.1.

5. Obliczyliśmy na poprzednich zajęciach korelację pomiędzy zmiennymi. Utwórz model regresji liniowej definiując zmienną objaśniającą jako tę cechę, która posiada współczynnik korelacji Pearsona ze średnią roczną sumą opadów równy 0.76. Oblicz błąd średniokwadratowy pomiędzy wartościami obserwowanymi a przewidywanymi.

6. Zastosuj funkcję `predict()` na testowym zbiorze danych dla modelu liniowego z ćw.5. Oblicz błąd średniokwadratowy (RMSE) pomiędzy wartościami obserwowanymi a przewidywanymi.

7. Wyświetl na wykresie wartości średniej rocznej wielkości opadów dla zbioru treningowego i testowego z nałożonym modelem liniowym z ćw. 5.

8. Utwórz model liniowy z maksymalną ilością opadów jako zmienną objaśniającą. Oblicz błąd średniokwadratowy pomiędzy wartościami obserwowanymi a przewidywanymi.

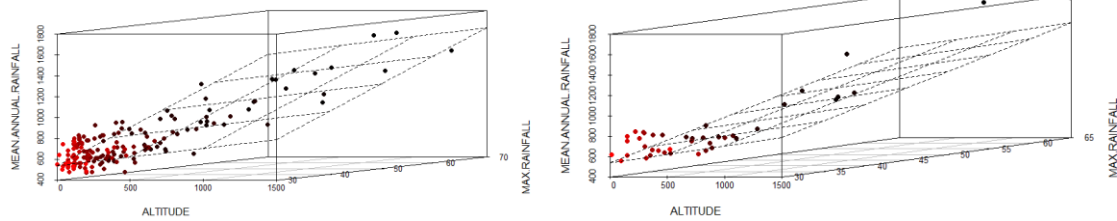
9. Zastosuj funkcję `predict()` na testowym zbiorze danych dla modelu liniowego z ćw.8 i oblicz RMSE pomiędzy wartościami obserwowanymi a przewidywanymi.

10. Wyświetl na wykresie wartości zmiennej objaśnianej dla obydwu zbiorów wraz z linią regresji modelu.

11. Utwórz model regresji liniowej wielokrotnej używając dwóch zmiennych objaśniających używanych do utworzenia modeli prostych regresji liniowych. Ile wynosi współczynnik R^2 i jak go zinterpretujesz? Oblicz RMSE pomiędzy wartościami obserwowanymi a przewidywanymi.

12. Wykonaj predykcję wartości zmiennej objaśnianej na podstawie modelu liniowej regresji wielokrotnej utworzonego w poprzednim ćwiczeniu. Oblicz RMSE pomiędzy wartościami obserwowanymi a przewidywanymi.

13. Przedstaw na wykresie 3D model regresji wielokrotnej używając funkcji `scatter.3d$plane3d` wraz z wartościami zmiennej objaśnianej oraz zmiennych objaśniających dla zbioru treningowego i testowego.



14. Przedstaw na wykresie wartości RMSE dla każdego z czterech modeli na zbiorze treningowym i testującym.