

Wprowadzenie do uczenia maszynowego

Lista 3

Janusz Szwabiński

1. **Implementacja prostej regresji liniowej:** Zaimplementuj prostą regresję liniową (jedna zmienna niezależna) od podstaw, używając metody najmniejszych kwadratów (Ordinary Least Squares - OLS). Wygeneruj syntetyczny zbiór danych, dopasuj do niego model i zwizualizuj linię regresji.
2. **Regresja liniowa z scikit-learn:** Użyj klasy `LinearRegression` z biblioteki `scikit-learn` do rozwiązania problemu z zadania 1. Porównaj uzyskane współczynniki z własną implementacją. Następnie, użyj publicznie dostępnego zbioru danych (np. `Boston Housing`) do zbudowania modelu regresji wielokrotnej.
3. **Metryki oceny regresji:** Oblicz i zinterpretuj podstawowe metryki oceny modelu regresji: średni błąd kwadratowy (MSE), pierwiastek średniego błędu kwadratowego (RMSE), średni błąd bezwzględny (MAE) oraz współczynnik determinacji (R^2).
4. **Regularyzacja - Ridge i Lasso:** Wprowadź szum do danych z zadania 1. Zbuduj model regresji liniowej, regresji grzbietowej (**Ridge**) i regresji lasso (**Lasso**). Porównaj współczynniki modeli i wyjaśnij, jak regularyzacja wpływa na wagę cech.
5. **Wprowadzenie do regresji logistycznej:** Zaimplementuj funkcję sigmoidalną w Pythonie. Wyjaśnij jej rolę w regresji logistycznej i zwizualizuj jej przebieg. Omów, dlaczego jest ona używana w problemach klasyfikacji binarnej.
6. **Implementacja regresji logistycznej z scikit-learn:** Wczytaj publiczny zbiór danych do klasyfikacji (np. `Titanic` lub `Breast Cancer`). Podziel dane na zbiór treningowy i testowy. Wytrenuj model regresji logistycznej i dokonaj predykcji.

7. **Metryki oceny klasyfikacji:** Dla modelu z zadania 6, oblicz i zinterpretuj następujące metryki: dokładność (**accuracy**), precyzję (**precision**), czułość (**recall**) oraz macierz pomyłek (**confusion matrix**).