Sprawozdanie

Laboratorium 2

Zadanie Płatki śniadaniowe

GRUPA WCY21IJ1N1

Opracowali:

Karczewski Paweł

Tarkowski Adam

# Treść zadania

Plik o nazwie „MED-lab-2-Zad-Platki sniadaniowe.zip” zawiera dane dotyczące płatków śniadaniowych. Proszę dokonać analizy zawartości zbioru i na gruncie regresji logistycznej należy:

1. przeanalizować dane i zastanowić się nad tym „Co chcemy zbadać i dlaczego?”, tzn. na jakie pytania chcemy sobie odpowiedzieć;
2. w powyższym kontekście, na gruncie regresji logistycznej, należy zaproponować model lub modele badające wybrane zależności;
3. proponowane modele należy uzasadnić oraz wskazać sposób ich praktycznego użycia do predykcji;
4. na podstawie opracowanych modeli i przeprowadzonych obliczeń sformułować własne wnioski;
5. wyniki swoich analiz zawierających: postawione pytania „badawcze”, wzory, udokumentowane wyniki obliczeń, wnioski, itp. proszę zawrzeć w postaci sprawozdania. Obliczenia można przeprowadzić w dowolnie wybranym narzędziu. Do sprawozdania proszę dodać jako załączniki wszystkie pliki z obliczeniami

# Analiza danych

Dane posiadają informacje o płatkach śniadaniowych:

1. Nazwę płatków
2. Wartości odżywcze (kalorie, cukry, węglowodany, proteiny, tłuszcz, sód, błonnik, potas).
3. Nazwę producenta.
4. Numer półki

Z rzeczy, które można zbadać to wybór półki dla określonych wartości odżywczych.

# Zaproponowany model

Chcąc używać funkcji logistycznej możemy zaprezentować funkcję logitową dla każdego modelu w taki sposób:

Gdzie:

W innej wersji model wygląda tak:

Gdzie:

Szukamy współczynników równania:

Aby obliczyć te współczynniki rozwiążemy pewne zadanie optymalizacyjne:

Wyznaczamy takie:

Dla którego spełnione jest równanie:

Aby wyznaczyć optymalne parametry skorzystać możemy z metody największego spadku opartej o gradienty, która zastosowała zastosowana przy obliczeniach ograniczając ją ilością iteracji.

# Wyniki

Przy typowaniu półki nr 2, po wyznaczeniu współczynników z grupy danych trenujących i przetestowaniu ich mieliśmy takie wyniki:

A screenshot of a computer

Description automatically generated

Gdzie dokładność to wartość trafionych wyników czyli gdzie prawdopodobieństwo było większe niż 50%. Zestawione zostało to z wynikami regresji liniowej o tej samej metodyce liczenia trafień.

Znalezione parametry w kolejności jaka zastała podana w Analizie danych

-0.5683772747111845,

-0.5736320431001749,

0.12674826668438507,

-0.4458982051573689,

-0.6864230312349382,

-0.9934698720679568,

0.2667906839079196

Przy półce nr 1:

A screenshot of a computer code

Description automatically generated

Przy półce nr 3:

A screenshot of a computer code

Description automatically generated

# Podsumowanie

Nauczony model świetnie sprawdza się do sprawdzania półki nr 2. Widocznie półka nr 2 z jakiś powodów jest ważniejsza od innych i tam wrzucane są wartości o wyróżniającym się składzie. Pozostałe półki, mimo, że można wyznaczyć model regresji to nie spisuje się tak samo dobrze.