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:

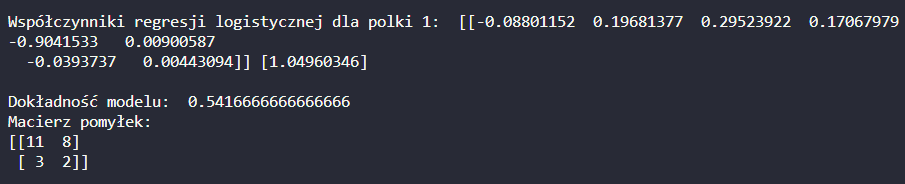
Gdzie:

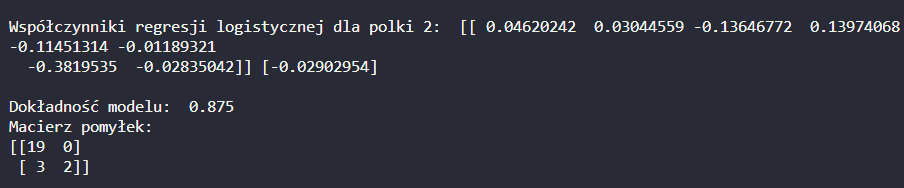
# Użycie modelu

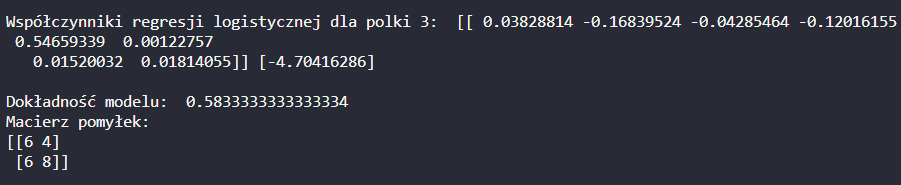
Niestety sposób wyliczenia liniowych współczynników modelu nie jest możliwye używając metod analitycznych, ale w pythonie w bibliotece scikit-learn mamy klasę LogisticRegression, która za pomocą metod numerycznych posłuży nam do tworzenia modelu (wyliczenia współczynników). A potem weryfikacji jego działania. Załadujemy go danymi treningowymi z wszystkimi wartościami odżywczymi.

Wyniki obliczeń, które są robione za pomocą pliku “Calculations.py”:

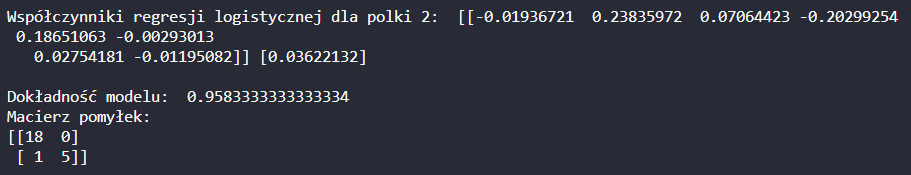








Oczywiście dane treningowe są losowo wybierane więc współczynniki i dokładności modelu mogą być inne, np. dla półki nr 2 wartość dokładności modelu zbliża się już nawet do całości:



Reszta półek ma przeważnie wartości powyżej 0,5 ale nie są to wartości tak widowsikowe jak przy półce nr 2.

# 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.