

Laboratorium 2

Metoda najmniejszych kwadratów

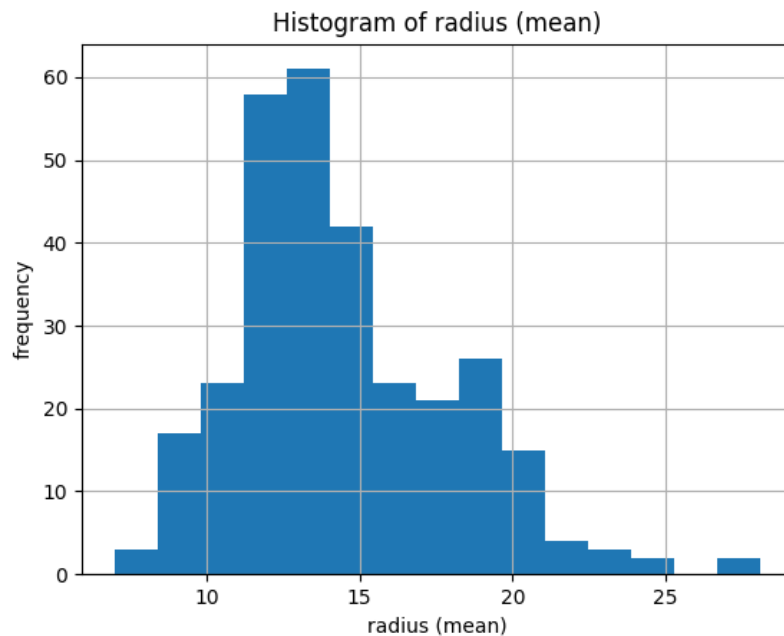
Mateusz Król

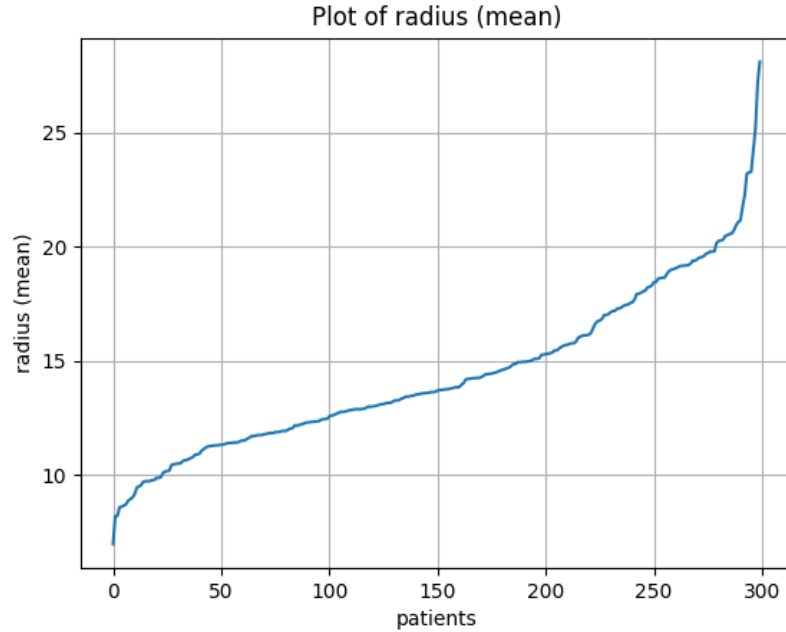
12/03/2024 r.

Zadanie 1.

Zastosuj metodę najmniejszych kwadratów do predykcji, czy nowotwór jest złośliwy (ang. malignant) czy łagodny (ang. benign). Nowotwory złośliwe i łagodne mają różne charakterystyki wzrostu. Istotne cechy to m. in. promień i tekstura. Charakterystyki te wyznaczane są poprzez diagnostykę obrazową i biopsje.

Przykładowe dane:





Korzystając z A_{lin} - liniowej reprezentacji oraz A_{quad} - kwadratowej reprezentacji danych z pierwszego zbioru dostarczonego do zadania, jesteśmy w stanie wyznaczyć wektor wag w :

$$A^T \cdot A \cdot w = A^T \cdot b$$

W tym celu wykorzystuję funkcję `scipy.linalg.solve`, służącą do rozwiązywania układów równań liniowych, z biblioteki `SciPy`.

Współczynniki uwarunkowania macierzy liczę korzystając z funkcji `numpy.linalg.cond` z biblioteki `NumPy`. Wartości wyniosły:

- dla reprezentacji liniowej - $\text{cond}(A^T \cdot A) \approx 1.81 \cdot 10^{12}$
- dla reprezentacji kwadratowej - $\text{cond}(A^T \cdot A) \approx 9.06 \cdot 10^{17}$

Następnie wyznaczając wektor p na podstawie danych z drugiego zbioru dostarczonego do zadania i porównując go z wartościami prawdziwymi, jesteśmy w stanie oszacować dokładność metody dla różnych reprezentacji macierzowych danych:

Matrix representation	false-positives	false-negatives
linear	6	2
quadratic	15	5

Dokładność reprezentacji macierzowych:

- linear: $accuracy = \frac{260-6-2}{260} \approx 96.9\%$
- quadratic: $accuracy = \frac{260-15-5}{260} \approx 93.5\%$

Wnioski

Reprezentacja liniowa okazała się być dokładniejsza dla dostarczonych zbiorów danych - zwróciła mniejszą ilość wyników fałszywych niż reprezentacja kwadratowa. Wynika to zapewne z faktu, że dostarczyliśmy w reprezentacji kwadratowej mniej danych - zaledwie 4 kolumny.

Wykorzystanie w metodzie reprezentacji liniowej, zapewniło mniejszy współczynnik uwarunkowania.