

Klasyfikacja i regresja: Granica decyzyjna

21 listopada 2018

Opis pliku z zadaniami

Wszystkie zadania na zajęciach będą przekazywane w postaci plików `.pdf`, sformatowanych podobnie do tego dokumentu. Zadania będą różnego rodzaju. Za każdym razem będą one odpowiednio oznaczone:

- Zadania do wykonania na zajęciach oznaczone są symbolem $\triangle$ – nie są one punktowane, ale należy je wykonać w czasie zajęć.
- Punktowane zadania do wykonania na zajęciach oznaczone są symbolem $\diamond$ – należy je wykonać na zajęciach i zaprezentować prowadzącemu.
- Zadania do wykonania w domu oznaczone są symbolem $\star$ – są one punktowane, należy je dostarczyć w sposób podany przez prowadzącego i w wyznaczonym terminie (zwykle przed kolejnymi zajęciami).

1 Granica decyzyjna w NKB – analiza

10p.◇

Treść

Celem ćwiczenia jest odpowiedzenie na pytanie: Jaką postać ma granica decyzyjna w naiwnym klasyfikatorze bayesowskim (NKB)?

1. Jaka może to być funkcja i jak może wyglądać?
2. Jak można empirycznie sprawdzić postać takiej funkcji: czy możemy rozwiązać ten problem implementacyjnie?

(a) Zaimplementuj generator sztucznych danych:

- Przyjmij liczbę cech równą 2, czyli $\mathbf{x} = (x_1, x_2)$. Ponadto załóżmy, że $\mathbf{x} \in \mathcal{X} = [-1, 1]^2$
- Wylosuj z rozkładu jednostajnego punkty w przestrzeni $\mathcal{X}$.
- Dla każdego punktu przyjmij klasę

$$y = \text{sign} \left(w_0 + \sum_{i=1}^2 w_i x_i \right),$$

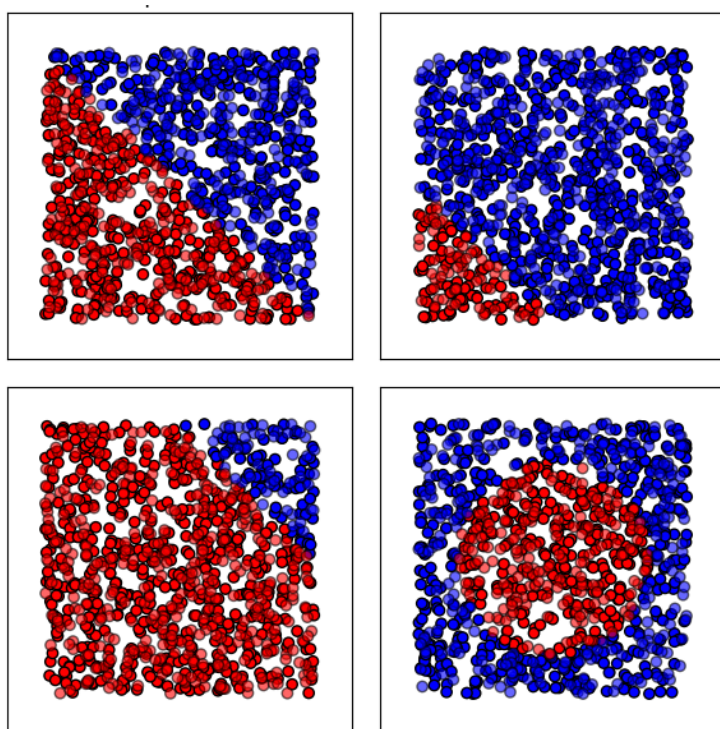
gdzie w_1, w_2 oraz w_0 są parametrami wejściowymi.

- Możesz też sprawdzić inny typ danych:

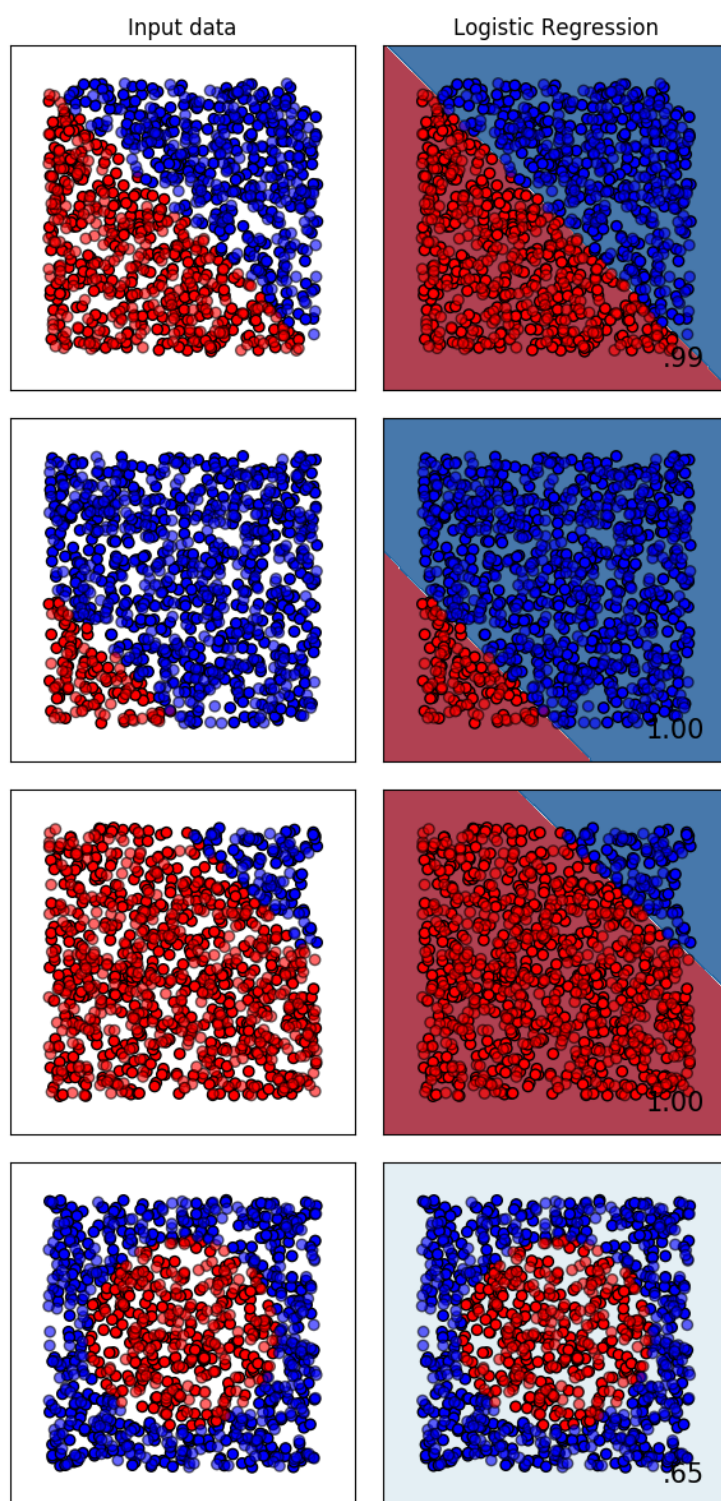
$$y = \text{sign} \left((x_1 - w_1)^2 + (x_2 - w_2)^2 - w_0 \right)$$

- Przedstaw dane na wykresie.
 - Czy dane są liniowo separowalne?
- (b) Przedstaw na wykresie granicę decyzyjną NKB, regresji logistycznej, KNN oraz drzewa decyzyjnego na otrzymanych danych. Czym są granice decyzyjne? Dlaczego mają one taką postać?
 - (c) Sprawdź wyniki dla różnych parametrów modelu. Co zauważyłeś?
 - (d) Czy potrafisz zmienić swoją implementację NKB, tak aby granica decyzyjna była hyperpłaszczyzną? Czy możesz to przedstawić formalnie? (**bonus 5p**)

Przykładowe zwizualizowane sztuczne zbiory danych przedstawione są na Rysunku 1. Granica decyzyjna może być zwizualizowana jak na Rysunku 2.



Rysunek 1: Przykładowe dane.



Rysunek 2: Granica decyzyjna regresji logistycznej.