



Metoda WRF

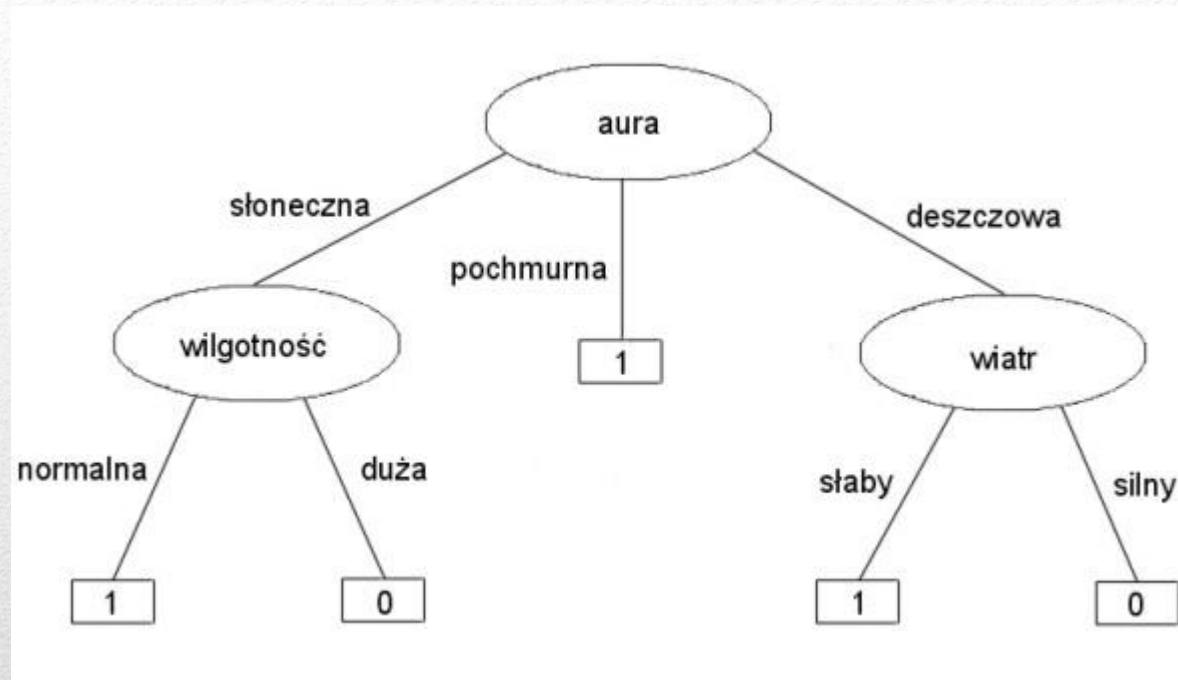
Weighted Random Forest

- Czy RF (random forest) sprawdzi się dla naszych danych?
- Dlaczego potrzebujemy BRF / WRF ?

Lasowy losowe

- Jak działa drzewo?
- Czy pojedyncze drzewo nadaje się do eksperymentu?

Jak działają lasy?



Jak działa drzewo?

"Drzewo klasyfikacyjne jest drzewem, które posiada dodatkową interpretację dla węzłów, gałęzi i liści:

- węzły odpowiadają testom przeprowadzonym na wartościach atrybutów przykładów,
- gałęzie odpowiadają możliwym wynikom tych testów,
- liście odpowiadają etykietom klas rozważanego problemu dyskryminacji,
- drzewo „rośnie” od góry do dołu (od korzenia do liści)."

Jak działa drzewo?

- Czy najpierw pytać o aurę, a dopiero potem o wilgotność?
- Czy użyć 2/3 atrybutów a nie 100?
- Jakie testy wykonać, by dojść do najlepszego rozróżnienia otrzymanych klas?

Dlaczego lasy, a nie drzewo?

1. Ze zbioru danych, losujemy z zwracaniem elementy i powtarzamy tę czynność kilka, kilkaset, kilka tysięcy razy.

Lasy losowe- metodologia

2. Otrzymujemy tzw. pseudo-próby (metoda bootstrap).

Lasy losowe- metodologia

3. Korzystając z każdej pseudo-próby, budujemy drzewo klasyfikacyjne, losując atrybuty, na których dokonujemy najlepszego testu.

Lasy losowe- metodologia

4. Zliczamy do jakiej klasy najwięcej drzew przyporządkowało nam dany przykład, tam go klasyfikujemy, a atrybuty tych drzew mają podniesioną wagę w stosunku do atrybutów drzew, które w tym głosowaniu przegrały. w ten sposób widzimy, jakie atrybuty są ważne w klasyfikacji, a jakie mają najmniejszy wpływ.

Lasy losowe- metodologia

- Cost sensitive learning.
- „Bias” w stronę klas większościowych.
- Przypisanie wagi – dla małych klas duża waga.

METODA WRF

- We shall place a heavier penalty on misclassifying the minority class.
- Wagi w 2 miejscach:
 - Proces indukcji drzewa – do znalezienie podziałów w kryterium Gini.
 - W każdym z końcowych węzłów drzewa.

Metoda WRF

- Implementacja
- Jak jest zdefiniowane wagowe kryterium Gini?
- Wsparcie

Problemy ?
