

Klasyfikacja 2

6 grudnia 2016

1 Na następnych zajęciach

W następnym tygodniu pojawi się treść nowego raportu. Termin oddania: 10 stycznia 2017. Na wykładach 13 i 20 stycznia będą prezentacje raportów.

2 Minimalizacja straty z konkursu OLX ★

Treść

Zakładamy, że dla każdej etykiety-liścia mamy dostępne estymaty prawdopodobieństwa dla danych wartości cech. Pokaż jak obliczyć optymalną decyzję ze względu na funkcję straty użytą w tym konkursie.

3 Model predykcyjny dla danych z OLX ★

Treść

Użyj danych przygotowanych na potrzeby poprzednich zajęć – reprezentacji bag of words (lub innej jeśli uznasz to za stosowne), nie ogarniaczaj się tylko to najpopularniejszych słów. Wyodrębnij zbiór treningowy i testowy ze zbioru treningowego. Utwórz prosty model pozwalający przewidywać prawdopodobieństwa poszczególnych etykiet. Sprawdź wartość funkcji straty gdy przewidujesz etykietę z najwyższą estymatą prawdopodobieństwa.

Użyj opracowanej w poprzednim zadaniu metody wyboru etykiety minimalizującej stratę do przewidywania etykiet. Porównaj wyniki, tzn wartość straty, z użyciem i bez tej metody.