Regresja logistyczne - ćwiczenia

Piotr J. Sobczyk

12 January 2017

1. Ściągnij dane z adresu "http://www.biecek.pl/R/dane/dane0.csv" (można to zrobić za pomocą funkcji **read.table**)
2. Zbuduj model, który tłumaczy wystąpienie wznowienia choroby nowotworowej w zależności od zmiennych **Nowotwor** i logarytmu zmiennej **VEGF**
3. Sprawdź, że na modelu można zastosować kryterium wyboru modelu AIC. Jak ono działa? Jak wygląda maksymalizowana wartość? Jaki problem niosą ze sobą braki danych?
4. Sprawdź jak przewidywać wartość p-stwa dla nowych obserwacji
5. Przeanalizuj się jak zmiana wartości zmiennych objaśniających wpływa na przewidywane prawdopodobieństwo wznowienia choroby. Czy potrafisz wyjaśnić tę zależność matematycznie?
6. Sterując punktem odcięcia na prawdopodobieństwo $y = 1$, sprawdź jak zmienia się sensitivity i specificity. Ze względu potencjalne zastosowanie tego modelu, który błąd jest dla nas bardziej kosztowny?
7. Zbuduj pełny model. Dokonaj wyboru zmiennych za pomocą kryterium informacyjnego Akaike
8. Zbuduj model regresji lasso. Jak wybierzesz parametr λ?