

Regresja - ćwiczenia

Marcin K. Dyderski

18 kwietnia 2018

Regresja - zadania

1. Wczytaj plik 'prunus.csv' dostępny na githubie. Zawiera on dane wykorzystane w pracy Dyderski i Jagodziński 2015 https://www.forestry.actapol.net/pub/2_2_2015.pdf. Opis zmiennych: a - zwarcie warstwy drzew (%) b - zwarcie warstwy krzewów (%) c - pokrycie runa zielnego (%) d - pokrycie warstwy mszystej (%) prunusc - liczba sztuk czeremchy w warstwie zielnej prunusc - liczba sztuk czeremchy w warstwie krzewów richness - bogactwo gatunkowe runa shannon - wskaźnik różnorodności Shannona dla runa L - wskaźnik świetlny Ellenberga (1-9, 1-cień, 9-pełne słońce) M - wskaźnik wilgotności Ellenberga (1-12, 1- pustynia, 12 - rośliny zanurzone) SR - wskaźnik odczynu gleby (1-9, 1-kwaśne, 9-lekko zasadowe, 7- obojętne) N - wskaźnik żyzności (1-9, 1-ubogie, 9-bardzo żyzne)
2. Wygeneruj macierz korelacji dla danych. Pamiętaj, że do funkcji `cor()` można wrzucić tylko zmienne liczbowe.
3. Przygotuj model liniowy zależności zwarcia drzew (a) od wskaźnika wilgotności (M). Podaj współczynnik determinacji, istotność zmiennych oraz sprawdź za pomocą wykresu diagnostycznego potencjalne problemy modelu.
4. Opracuj i zwizualizuj model nieliniowy zależności pomiędzy wskaźnikiem Shannona a liczbą czeremch w warstwie b. Podaj R^2 i sprawdź czy model ten jest lepszy niż model liniowy za pomocą AIC.
5. Za pomocą GLM z rozkładem Poissona oceń które parametry najlepiej wyjaśniają liczebność czeremchy na polkach badawczych.
6. Na stronie <https://github.com/mkdyderski/BSS> znajduje się plik 'hotspots.csv'. Zawiera on dane z Appendixu do pracy Dawson et al. 2017. <https://www.nature.com/articles/s41559-017-0186> Opis danych: area - nazwa regionu continent - kontynent country - kraj plants, ants... - liczba obcych gatunków z grup taksonomicznych area - powierzchnia [km²] mainl - czy wyspa czy mainland popdens - zagęszczenie ludności GDP - PKB na głowę mieszkańca MAT - średnia roczna temp. MAP - średnie roczne opady N, E - współrzędne geograficzne
7. Dla wybranej grupy taksonomicznej sprawdź które czynniki najlepiej wyjaśniają bogactwo gatunków obcych przy zastosowaniu wyłącznie modelu liniowego
8. Sprawdź zależność pomiędzy bogactwem gatunkowym roślin a GDP za pomocą modeli GAM.
9. Co stanie się jeśli do modelu bogactwa gatunkowego mrówek wrzucimy MAT, MAP, N i E? Jak zmienia się VIFy po wyrzuceniu poszczególnych elementów i jak będzie zmieniać się AIC? Jaka będzie ostateczna postać modelu? Możesz wykorzystać dowolny typ modelu (liniowy, nieliniowy, glm, gam itp.)
10. Porównaj jakość GLM oraz GLMM (modelu mieszanego) bogactwa gatunkowego wybranej grupy taksonomicznej. W GLM jako predyktorów użyj GDP, MAT, MAP i popdens, w GLMM jako zmienne losowe wykorzystaj kraj i kontynent. Określ conditional i marginal R^2 i zinterpretuj wynik.