

Miniprojekt 2

Julia Biały

26 kwietnia 2022

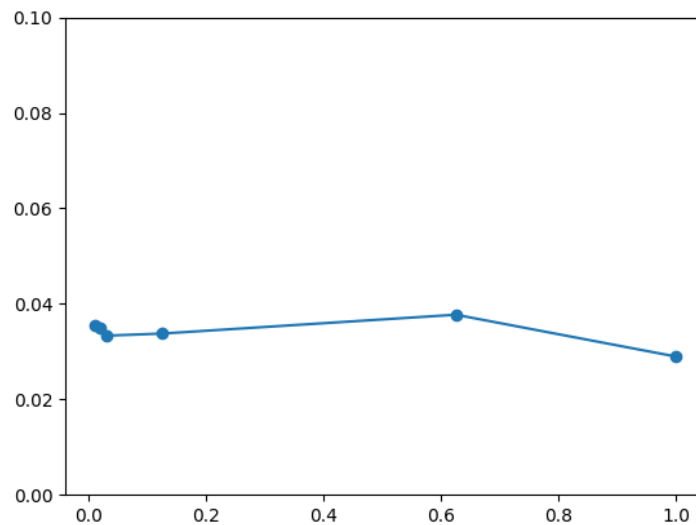
1 Model regresji logistycznej

Najpierw daną opisującą zamieniłam odpowiednio z 2 i 4 na 0 i 1. Dane będące parametrami skaluję tak by uzyskać na każdym parametrze wartość oczekiwaną 0 i standardowe odchylenie równe 1. Zbiór danych podzieliłam wybierając do zbioru treningowego losowo $\frac{2}{3}$ danych z każdej klasy. Podczas trenowania zbioru użyłam metody stochastycznego gradientu. Wyniki zostały uśrednione dla 10 przebiegów algorytmu.

Błędy klasyfikacyjne dla kolejnych frakcji zbioru treningowego wynoszą:

0.03552, 0.03508, 0.03333, 0.03377, 0.03771, 0.02894.

Poniższy wykres przedstawia krzywą uczenia.

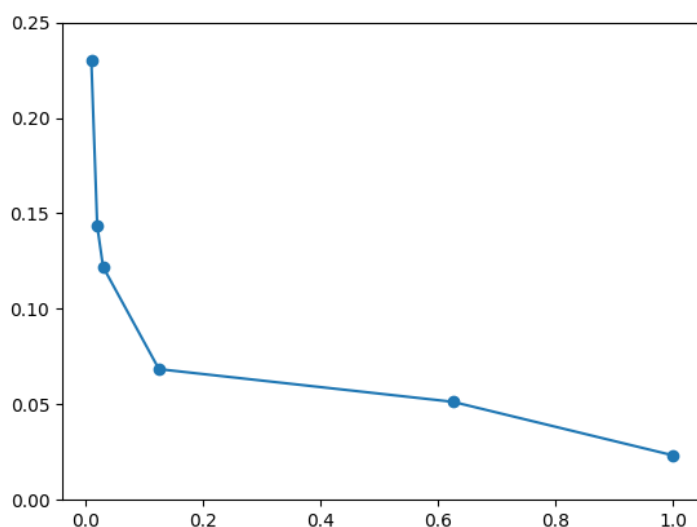


2 Model naiwnego klasyfikatora bayesowskiego

Daną opisującą zamieniłam tak samo jak w modelu regresji logistycznej z 2 i 4 na 0 i 1. Zbiór danych podzieliłam wybierając dane do zbioru treningowego w ten sam sposób co w powyższym modelu. Przyjęłam założenie, że dziewięć parametrów jest od siebie niezależnych oraz że zmienne warunkowe mają rozkład wielomianowy. Zastosowałam wygładzenie Laplace'a. Wyniki uśredniłam dla 10 przebiegów algorytmu. Błędy klasyfikacyjne wynoszą odpowiednio:

0.22982, 0.14342, 0.12192, 0.06842, 0.051315, 0.02324

Poniższy wykres przedstawia krzywą uczenia dla tego modelu.



3 Wnioski

W artykule "On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes" autorzy twierdzą, że błąd asymptotyczny naiwnego klasyfikatora bayesowskiego jest większy, ale osiągany szybciej, a błąd asymptotyczny regresji logistycznej niższy, ale osiągany wolniej. Rezultaty otrzymane na zaimplementowanych przeze mnie modelach jednak tego nie potwierdzają, bo to klasyfikator dyskryminatywny - model regresji logistycznej nauczył się szybciej, a klasyfikator generatywny - naiwny klasyfikator bayesowski osiągał gorsze wyniki na mniejszych frakcjach zbioru treningowego, ale lepszy wynik na pełnym zbiorze.