

1. Załóżmy, że w Twojej skrzynce mailingowej pojawiła się nowa wiadomość, która ma temat „Cześć”. Masz statystyki wszystkich swoich maili i wiesz, że: (a) 10% wszystkich wiadomości jest spamem, (b) 50% spamu ma temat „Cześć”, (c) tylko 2% nie-spamu ma temat „Cześć”. Jakie jest prawdopodobieństwo, że ta wiadomość jest spamem?
2. Szacuje się, że 0,05% populacji USA ma HIV. Istnieje test na HIV: (a) jeśli badany ma HIV, test ma 98% szansy na pozytywny wynik; (b) jeżeli osoba nie ma HIV, test ma 3% szansy na pozytywny wynik. Tomek ma wynik pozytywny. Jakie jest prawdopodobieństwo, że ma HIV?
3. Rozważmy prosty zbiór treningowy, w którym każdy przykład ma cztery binarne atrybuty i przydzieloną jedną z dwóch klas (+/-):

Przykład	Atrybut_1	Atrybut_2	Atrybut_3	Atrybut_4	Klasa
x1	1	1	1	1	+
x2	1	1	0	1	+
x3	0	1	1	0	+
x4	1	0	0	1	+
x5	1	0	0	0	+
x6	1	0	1	0	-
x7	0	1	0	0	-
x8	0	0	1	0	-

W jaki sposób naiwny klasyfikator Bayesowski, wyuczony na powyższym zbiorze treningowym, zaklasyfikuje poniższy przykład? Policz ręcznie :)

Przykład	Atrybut_1	Atrybut_2	Atrybut_3	Atrybut_4	Klasa
x9	1	1	0	0	?

4. Sprawdźmy jak to działa w Wece wykonując następujące zadania:
 - Włącz Wekę i wczytaj plik `weather.numeric.arff`
 - Przejdź na zakładkę **Classify**. Weka udostępnia dwie wersje Naive Bayes. Zapoznaj się z ich opisami:
 - NaiveBayes
 - NaiveBayesUpdateable
 - Przetestuj obydwie wersje algorytmu? Jakie są różnice?
 - Następnie przetestuj te same algorytmy na pliku `weather.nominal.arff`. Jakie różnice występują teraz? Co jest przyczyną występowania różnic - odpowiedz korzystając z opisów algorytmów.,
5. Dokonaj weryfikacji działania klasyfikatora SVM i wpływu jego parametrów na jakość uzyskiwanych wyników. Zadanie wykonaj na danych z pliku: `irys2D`.
 1. Wykorzystaj funkcję: `LibLINEAR` oraz `LibSVM`.
Dla `LibLINEAR`: Zmieniaj parametr `C`. (`C` - Stała kontroluje wpływ każdego indywidualnego wektora nośnego na funkcję celu. Większe `C` zmniejsza efekt $|\mathbf{w}|^2/2$ i skutkuje węższym marginesem.)
 2. Dla `LibSVM`: Zmieniaj parametry `C` oraz `gamma`. Wykonaj obliczenia dla wszystkich dostępnych kernel's.