

1. Wyestymuj łączny rozkład prawdopodobieństwa tych danych, zgodnie z zasadą maksymalnej wiarygodności, a następnie oblicz prawdopodobieństwo, że  $P(y = \star | \text{słonecznie, false})$ . Do jakiej klasy zostałby przydzielony przykład (słonecznie, false) wg. klasyfikatora skonstruowanego na tym prawdopodobieństwie?

Outlook	Windy	Play?
słonecznie	false	o
słonecznie	true	o
pochmurnie	false	★
deszcz	false	★
deszcz	false	★
deszcz	true	o
pochmurnie	true	★
słonecznie	false	o
słonecznie	false	★
deszcz	false	★
słonecznie	true	★
pochmurnie	true	★
pochmurnie	false	★
deszcz	true	o

2. Analizując klasyfikator z poprzedniego zadania, odpowiedz na pytania:

- Jakich rozkładów prawdopodobieństwa nie możemy się na uczyć (tj. zamodelować)?
- Czy dostrzegasz jakieś wady zaproponowanego podejścia?
- Ile parametrów ma ten klasyfikator?
- Zakładając klasyfikację binarną i  $d$  cech binarnych, podaj wzór na liczbę parametrów tego klasyfikatora.

3. Zakładając rozkład normalny cech pod warunkiem klasy:

$$P(\mathbf{x}, y) = N(\mathbf{x} | \mu_y, \Sigma_y) P(y)$$

dokonaj estymacji tego klasyfikatora zgodnie z zasadą maksymalnej wiarygodności, a następnie podaj wyrażenie na  $P(+ | x_1 = 3, x_2 = 0)$ .

Uwaga: przy estymowaniu wariancji i kowariancji możesz użyć nieobciążonych estymatorów pomimo innego wyniku dyktowanego przez MLE.

$x_1$	$x_2$	$y$
1	-2	+
2	0	+
3	2	+
1	5	-
1	-5	-

4. Analizując klasyfikator z poprzedniego zadania, odpowiedz na pytania:

- Jakich rozkładów prawdopodobieństwa nie możemy się nauczyć (tj. zamodelować)?
- Czy dostrzegasz jakieś wady zaproponowanego podejścia?
- Ile parametrów ma ten klasyfikator?
- Zakładając klasyfikację binarną i  $d$  cech binarnych, podaj wzór na liczbę parametrów tego klasyfikatora.

5. Zakładając dane z pierwszego zadania wyznacz prawdopodobieństwo  $P(y = \star | \text{słonecznie, false})$  zgodnie z klasyfikatorem naiwnego Bayesa, a następnie odpowiedz na pytania:

- Jakich rozkładów prawdopodobieństwa nie możemy się nauczyć (tj. zamodelować)?
- Czy dostrzegasz jakieś wady zaproponowanego podejścia?

- Ile parametrów ma ten klasyfikator?
  - Zakładając klasyfikację binarną i  $d$  cech binarnych, podaj wzór na liczbę parametrów tego klasyfikatora.
6. Klasyfikator naiwnego Bayesa zakłada warunkową niezależność cech. Czy warunkowa niezależność jest implikowana przez niezależność cech? Podaj kontrprzykład.
7. Oblicz wartość szansy tj.  $\frac{p}{1-p}$  dla  $p = 0.5$ ,  $p = 0.2$  oraz  $p = 0.9$ .
8. Jeżeli wynik modelu liniowego  $g(x)$  będzie:
- bliski  $-\infty$  to funkcja logistyczna zwróci wartość .....
  - bliski  $\infty$  to funkcja logistyczna zwróci wartość .....
  - równy 0 to funkcja logistyczna zwróci wartość .....
  - równy 2 to funkcja logistyczna zwróci wartość ok. ....
9. Próbuje się zamodelować prawdopodobieństwo ataku cybernetycznego w danym dniu przy użyciu liczby ataków z dnia poprzedniego. Otrzymano następujący model regresji logistycznej o współczynnikach  $b = 0.5$  oraz  $w = 0.1$ . Ile wynosi prawdopodobieństwo ataku, jeżeli wczoraj było ich 5?
10. Przeanalizuj modele LDA, naiwnego Bayesa i regresji logistycznej w kontekście zasady minimalizacji ryzyka empirycznego. Jakie są klasy hipotez? Jak jest optymalizowana funkcja? Jaki algorytm optymalizacyjny może być stosowany?
11. Wykorzystując poniższe aspekty wskaż na wady i zalety podejść dyskryminacyjnych i generatywnych.
- łatwość treningu (w sensie kosztu obliczeniowego)
  - łatwość dodania nowej klasy do już wytrenowanego klasyfikatora
  - łatwość obsługi brakujących danych
  - łatwość wykorzystania niezaetykietowanych danych podczas uczenia
  - łatwość wykorzystania przetworzonych cechy (tworzonych np. poprzez wymnażanie cech ze sobą lub wykonując na nich transformacje nieliniowe)
  - zwracanie dobrze wykalibrowanych prawdopodobieństw

