

# Egzamin SAD 2023

Dorota Celińska-Kopczyńska, Krzysztof Gogolewski, Błażej Miasojedow, Szymon Nowakowski,  
Kazimierz Oksza-Orzechowski, Piotr Pokarowski, Ewa Szczurek

20 czerwca 2023

Za każdy podpunkt każdego zadania jest przyznane tyle samo punktów. Punkty sumują się do 40.

**Zadanie 1 [Autor: ES, gr 1 – POPRAWKA PP]** Rozważmy metodę składowych głównych dla macierzy danych  $\mathbf{X}$  o  $n$  wierszach (obserwacje) i  $p$  kolumnach (cechy, atrybuty). Załóżmy, że  $\mathbf{X}$  jest scentrowana ( $\mathbf{X}^T \mathbf{1} = \mathbf{0}$ ). Niech  $\mathbf{S} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X}$  oznacza empiryczną macierz kowariancji danych oraz  $\mathbf{X} = \mathbf{U} \mathbf{D} \mathbf{V}^T$  – rozkład SVD macierzy  $\mathbf{X}$ , w którym  $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_p)$  oraz  $\lambda_1 \geq \lambda_2, \dots, \geq \lambda_p$ . Przypomnijmy, że kierunkami głównymi  $v_1, \dots, v_p$  nazywamy kolumny macierzy  $\mathbf{V}$ , natomiast składowymi głównymi  $Y_1, \dots, Y_p$  – kolumny macierzy  $\mathbf{Y} = \mathbf{X} \mathbf{V}$ . Oceń prawdziwość stwierdzeń:

- $\mathbf{Y}$  jest scentrowana
- $\sum_{i=1}^n Y_{ij}^2 = \lambda_j^2$  dla każdego  $j = 1, \dots, p$
- $\mathbf{S} v_j = \frac{\lambda_j^2}{n-1} v_j$ , dla każdego  $j = 1, \dots, p$
- $\max_{a: a^T a = 1} a^T \mathbf{S} a = \frac{\lambda_1^2}{n-1}$

**Rozwiązanie:** T, T, T, T

**Zadanie 1 [Autor: ES, gr 2 – POPRAWKA PP]** Treść zadania jak w gr 1.

- $\sum_{i=1}^n Y_{ij} = 0$  dla każdego  $j = 1, \dots, p$
- $\sum_{i=1}^n (Y_{ij} - \bar{Y}_j)^2 = \lambda_j^2$  dla każdego  $j = 1, \dots, p$
- $\mathbf{S} v_j = \frac{\lambda_j^2}{n-1} v_j$ , dla każdego  $j = 1, \dots, p$
- $\min_{a: a^T a = 1} a^T \mathbf{S} a = \frac{\lambda_p^2}{n-1}$

**Rozwiązanie:** T, T, T, T

**Zadanie 2 [Autor: KO, gr 1]** Z 16 elementowej próby prostej  $X_1, \dots, X_{16}$  zmiennej  $X$  dostaliśmy:

$$\bar{X} = \frac{1}{16} \sum_{i=1}^{16} X_i = 1.1$$

$$\sum_{i=1}^{16} (X_i - \bar{X})^2 = 4$$

Wiemy, że  $X$  ma rozkład normalny o wariancji  $1/4$  i nieznaney średniej  $\mu$ . Dane są hipotezy: zerowa  $H_0 : \mu = \mu_0 = 1$  i alternatywna  $H_1 : \mu = \mu_1 = 1.1$ . (Przypomnienie: kwantyl rozkładu standardowego normalnego rzędu  $0.05$  wynosi w przybliżeniu  $-1.64$ ).

- nieobciążony estymator wariancji z tej próbki to  $1/4$
- przy poziomie istotności  $\alpha = 0.05$  i najmocniejszym teście nie odrzucamy  $H_0$
- moc testu najmocniejszego przy poziomie istotności  $\alpha = 0.05$  jest większa od  $\alpha$
- gdyby badać hipotezy: zerową  $H_0 : \mu = 1$  i alternatywną  $H_1 : \mu \neq 1$ , to dwustronna  $p$ -wartość tej próby byłaby mniejsza od  $0.1$

**Rozwiązanie:** F, T, T, F

**Zadanie 2 [Autor: KO, gr 2]** Z 25 elementowej próby prostej  $X_1, \dots, X_{25}$  zmiennej  $X$  dostaliśmy:

$$\bar{X} = \frac{1}{25} \sum_{i=1}^{25} X_i = -0.1$$

$$\sum_{i=1}^{25} (X_i - \bar{X})^2 = \frac{25}{16}$$

Wiemy, że  $X$  ma rozkład normalny o wariancji  $1/16$  i nieznannej średniej  $\mu$ . Dane są hipotezy: zerowa  $H_0 : \mu = \mu_0 = 0$ , alternatywna  $H_1 : \mu = \mu_1 = -0.1$ . (Kwantyl rozkładu standardowego normalnego rzędu 0.05 wynosi w przybliżeniu  $-1.64$ ).

- nieobciążony estymator wariancji z tej próbki to  $\frac{1}{16}$
- przy poziomie istotności  $\alpha = 0.05$  i najmocniejszym teście nie odrzucamy  $H_0$
- moc testu najmocniejszego przy poziomie istotności  $\alpha = 0.05$  jest większa od  $\alpha$
- gdyby badać hipotezy: zerową  $H_0 : \mu = 0$  i alternatywną  $H_1 : \mu \neq 0$ , to dwustronna  $p$ -wartość tej próby byłaby mniejsza od 0.1

**Rozwiązanie:** F,F,T,T

**Zadanie 3 [Autor: ES, gr 1]** Rozważmy metodę składowych głównych dla danych o  $n > 3$  obserwacjach i  $p$  atrybutach,  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , gdzie  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ ,  $1 \leq i \leq n$ . Oznaczmy ładunek dla  $j$ -tej składowej i  $i$ -tego punktu przez  $a_{j,i}$ . Rozważmy punkty  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$  takie, że ich zapisy (wyniki)  $y_{j,1}, y_{j,2}, y_{j,3}$  na  $j$ -tej składowej spełniają  $y_{j,1} < y_{j,2} < y_{j,3}$ . Rozstrzygnij, czy poniższe stwierdzenia zawsze zachodzą:

- $\|\mathbf{x}_2 - \mathbf{x}_1\| < \|\mathbf{x}_3 - \mathbf{x}_1\|$
- $Var[\{y_{j,1}, y_{j,2}, y_{j,3}\}] > Var[\{y_{(j+1),1}, y_{(j+1),2}, y_{(j+1),3}\}]$ , dla  $1 \leq j \leq p-1$
- Wektory  $(a_{j,1}, a_{j,2}, a_{j,3})^T$  oraz  $(a_{(j+1),1}, a_{(j+1),2}, a_{(j+1),3})$ , dla  $1 \leq j \leq p-1$ , są ortonormalne
- Kierunek dla składowej  $j$  będzie taki sam, nie zależnie od tego, czy dane były standardyzowane, czy nie

**Rozwiązanie:** F, F, F, F

**Zadanie 3 [Autor: ES, gr 2]** Treść zadania jak w gr 1.

- $\|\mathbf{x}_3 - \mathbf{x}_2\| < \|\mathbf{x}_3 - \mathbf{x}_1\|$
- $Var[\{y_{j,1}, y_{j,2}, y_{j,3}\}] < Var[\{y_{(j-1),1}, y_{(j-1),2}, y_{(j-1),3}\}]$ , dla  $2 \leq j \leq p$
- Wektory  $(a_{(j-1),1}, a_{(j-1),2}, a_{(j-1),3})$  oraz  $(a_{j,1}, a_{j,2}, a_{j,3})^T$ , dla  $2 \leq j \leq p$ , są ortonormalne
- Kierunek dla pierwszej składowej będzie taki sam, nie zależnie od tego, czy dane były standardyzowane, czy nie

**Rozwiązanie:** F, F, F, F

**Zadanie 4 [Autor: KO, gr 1]** Detektor rejestruje przepuszczoną przez niego cząsteczkę z nieznanym prawdopodobieństwem  $p$ . Przez ten detektor przepuszczono 100 cząsteczek, z których zarejestrował on 50. Wiemy, że detektor referencyjny rejestruje cząsteczki z prawdopodobieństwem 60%.

	0.1	0.05	0.025	0.02	0.01	0.005
1	2.7055	3.8415	5.0239	5.4119	6.6349	7.8794
2	4.6052	5.9915	7.3778	7.8241	9.2103	10.5966
3	6.2514	7.8147	9.3484	9.8374	11.3449	12.8382

Tabela 1: Tablica wartości krytycznych rozkładu  $\chi^2$  dla zadanej liczby stopni swobody  $n$  i poziomu istotności

- estymator największej wiarygodności i estymator nieobciążony prawdopodobieństwa  $p$  z tej próby są takie same

- na poziomie istotności  $\alpha = 0.05$  nie odrzucamy hipotezy, że w tej próbie dostaliśmy rozkład zgodny z rozkładem detektora referencyjnego
- obniżono temperaturę w detektorze, po czym jeszcze raz przepuszczono przez niego 100 cząsteczek. Tym razem wykrył on 60 z nich. Na poziomie istotności  $\alpha = 0.05$ , przy użyciu testu niezależności  $\chi^2$ , nie odrzucamy hipotezy, że wystąpienie obserwacji cząsteczki jest niezależne od obniżenia temperatury
- w podpunkcie o obniżeniu temperatury moglibyśmy skorzystać również z dokładnego testu Fischera

**Rozwiązanie:** T,F,T,T

**Zadanie 4 [Autor: KO, gr 2]** Linia produkcyjna może wypuścić wadliwy komputer z nieznanym prawdopodobieństwem  $p$ . Ta linia wypuściła 100 komputerów, w tym 10 wadliwych. Wiemy, że linia referencyjna wypuszcza wadliwe komputery z prawdopodobieństwem 5%.

	0.1	0.05	0.025	0.02	0.01	0.005
1	2.7055	3.8415	5.0239	5.4119	6.6349	7.8794
2	4.6052	5.9915	7.3778	7.8241	9.2103	10.5966
3	6.2514	7.8147	9.3484	9.8374	11.3449	12.8382

Tabela 2: Tablica wartości krytycznych rozkładu  $\chi^2$  dla zadanej liczby stopni swobody  $n$  i poziomu istotności

- estymator największej wiarygodności i estymator nieobciążony prawdopodobieństwa  $p$  z tej próby są różne.
- na poziomie istotności  $\alpha = 0.025$  nie odrzucamy hipotezy, że w tej próbie dostaliśmy rozkład zgodny z rozkładem linii referencyjnej.
- na linii wprowadzono nowe parametry maszyn, po czym jeszcze raz wyprodukowano 100 komputerów. Tym razem wadliwych było 5 z nich. Na poziomie istotności  $\alpha = 0.1$ , przy użyciu testu niezależności  $\chi^2$ , nie odrzucamy hipotezy, że wystąpienie wady komputera jest niezależne od tej zmiany parametrów
- w podpunkcie o nowych parametrach moglibyśmy skorzystać również z dokładnego testu Fischera

**Rozwiązanie:** F,F,T,T

**Zadanie 5 [Autor: SN, gr 1]** Oceń prawdziwość podanych zdań.

- dla metody LOOCV (leave-one-out cross-validation) w zadaniu klasyfikacji: jedyne możliwe wartości estymatora błędu testowego na zbiorach walidacyjnych w poszczególnych iteracjach to 0.0 lub 1.0
- dla metody LOOCV (leave-one-out cross-validation) w zadaniu regresji: jedyne możliwe wartości estymatora błędu testowego na zbiorach walidacyjnych w poszczególnych iteracjach to 0.0 lub 1.0
- jeśli wynikiem  $k$ -krotnej walidacji krzyżowej w zadaniu klasyfikacji jest estymator błędu testowego  $\hat{b} = 0.47$ , a w zbiorach walidacyjnych poszczególnych iteracji jest  $n_i = 10$  ( $i = 1, \dots, k$ ) obserwacji, to był przynajmniej jeden fold, w którym poprawnie sklasyfikowano nie więcej niż 4 obserwacje
- jeśli wynikiem  $k$ -krotnej walidacji krzyżowej w zadaniu regresji jest estymator błędu testowego  $\hat{b} = 0.41$ , a w zbiorach walidacyjnych poszczególnych iteracji jest  $n_i = 10$  ( $i = 1, \dots, k$ ) obserwacji, to był przynajmniej jeden fold, w którym uzyskano błąd testowy nie przekraczający 0.4

**Rozwiązanie:** T, F, T, F

**Zadanie 5 [Autor: SN, gr 2]** Oceń prawdziwość podanych zdań.

- dla metody LOOCV (leave-one-out cross-validation) w zadaniu klasyfikacji: możliwe jest uzyskanie estymatora błędu testowego na zbiorze walidacyjnym w pojedynczej iteracji dokładnie równego  $\frac{1}{2}$
- dla metody LOOCV (leave-one-out cross-validation) w zadaniu regresji: możliwe jest uzyskanie estymatora błędu testowego na zbiorze walidacyjnym w pojedynczej iteracji dokładnie równego  $\frac{1}{2}$

- jeśli wynikiem k-krotnej walidacji krzyżowej w zadaniu klasyfikacji jest estymator błędu testowego  $\hat{b} = 0.71$ , a w zbiorach walidacyjnych poszczególnych iteracji jest  $n_i = 10$  ( $i = 1, \dots, k$ ) obserwacji, to był przynajmniej jeden fold, w którym poprawnie sklasyfikowano co najmniej 8 obserwacji
- jeśli wynikiem k-krotnej walidacji krzyżowej w zadaniu regresji jest estymator błędu testowego  $\hat{b} = 0.79$ , a w zbiorach walidacyjnych poszczególnych iteracji jest  $n_i = 10$  ( $i = 1, \dots, k$ ) obserwacji, to był przynajmniej jeden fold, w którym uzyskano błąd testowy wynoszący co najmniej 0.8

**Rozwiązanie:** F, T, T, F

**Zadanie 6 [Autor: SN, gr 1]** Rozważmy metodę wyboru modelu poprzez minimalizację kryteriów informacyjnych AIC lub BIC. Załóżmy, że rodzina modeli liniowych jest budowana inkrementacyjnie (algorytm zachłanny przeszukiwania w przód) poprzez dodawanie predyktorów spośród  $P_1, P_2, P_3, P_4$  do początkowo pustego modelu dla  $n \geq 2$  obserwacji. W procesie konstrukcji rodziny modeli, otrzymano wartości podwojonego logarytmu wiarygodności  $2 \cdot \log(L)$  dla modeli zawierających narastającą liczbę predyktorów przedstawione w Tabeli 3. Poniżej oznaczenie "Px+...+Py" oznacza model, w którym użyto Px, ..., Py jako osobnych predyktorów.

Model	$\emptyset$	$P_2$	$P_2 + P_3$	$P_2 + P_3 + P_4$	$P_2 + P_3 + P_4 + P_1$
$2 \cdot \log(L)$	-840.33	-270.35	-148.00	-145.66	-144.00

Tabela 3: Tablica przedstawiająca zależność podwojonego logarytmu wiarygodności od wielkości modelu

- Modelem minimalizującym kryterium informacyjne Akaike AIC w opisanym procesie jest  $P_2 + P_3 + P_4$
- Nie ma gwarancji, że znaleziony model minimalizuje AIC wśród wszystkich modeli będących podzbiorami zbioru predyktorów, gdyż zastosowano algorytm zachłanny przeszukiwania w przód
- Gdyby zastosowano algorytm zachłanny przeszukiwania w obie strony (mogący zarówno dodać, jak i usunąć predyktor z modelu), znaleziony model minimalizowałby AIC wśród wszystkich modeli będących podzbiorami zbioru predyktorów
- Model  $P_2 + P_3 + P_4 + P_1$  nie minimalizuje kryterium informacyjnego Schwarza BIC w opisanym procesie, dla żadnego  $n$

**Rozwiązanie:** T, T, F, F

**Zadanie 6 [Autor: SN, gr 2]** (2 pkt) Treść zadania jak w gr 1.

- Modelem minimalizującym kryterium informacyjne Schwarza BIC dla  $n = 2$  w opisanym procesie jest  $P_2 + P_3 + P_4 + P_1$
- Nie ma gwarancji, że znaleziony model minimalizuje BIC wśród wszystkich modeli będących podzbiorami zbioru predyktorów, gdyż zastosowano algorytm zachłanny przeszukiwania w przód
- Gdyby zastosowano algorytm zachłanny przeszukiwania w obie strony (mogący zarówno dodać, jak i usunąć predyktor z modelu), znaleziony model minimalizowałby BIC wśród wszystkich modeli będących podzbiorami zbioru predyktorów
- Model  $P_2 + P_3 + P_4 + P_1$  nie minimalizuje kryterium informacyjnego Akaike AIC w opisanym procesie, dla żadnego  $n$

**Rozwiązanie:** T, T, F, T

**Zadanie 7 [Autor: KG, gr 1, ]** Określ prawdziwość poniższych stwierdzeń. W poniższych odpowiedziach jako przeuczenie (ang. overfitting) rozumiemy sytuację, w której model osiąga wysoką dokładność (ang. accuracy) na zbiorze danych treningowych, ale poziom tej dokładności maleje dla nowych, niebiorących udziału w procesie trenowania modelu, danych testowych.

- Klastrowanie hierarchiczne  $n$ -elementowego zbioru obserwacji  $D$  wymaga określenia liczby poszukiwanych klastrów ( $k$ ) przed przystąpieniem do obliczenia macierzy odległości między obserwacjami

- Algorytm  $k$ -średnich (ang.  $k$ -means) zawsze zbiega do tego samego rozwiązania, niezależnie od tego, jak klastry są inicjowane. Dla ustalenia uwagi rozważamy heurystykę przedstawioną na wykładzie
- Klasyfikator 1-NN ( $k=1$  najbliższych sąsiadów) najczęściej wiąże się z większym ryzykiem przeuczenia (ang. overfitting) niż 10-NN, choć w szczególnych przypadkach może być skuteczniejszy
- Ryzyko przeuczenia (ang. overfitting) w drzewach decyzyjnych wzrasta wraz ze wzrostem głębokości konstruowanego drzewa

**Rozwiązanie:** F, F, T, T

**Zadanie 7 [Autor: KG, gr 2, ]** Treść zadania jak w gr 1.

- Dla zadanego  $n$ -elementowego zbioru obserwacji  $D$  i macierzy odległości między obserwacjami  $M$  wymiaru  $n \times n$  przyporządkowanie obserwacji do  $k$  klastrów przy użyciu klastrowania hierarchicznego jest jednoznacznie wyznaczone za pomocą  $M$
- Dla ustalonego klastrowania początkowego algorytm  $k$ -średnich (ang.  $k$ -means) zawsze zbiega do tego samego rozwiązania
- Klasyfikator 1-NN ( $k=1$  najbliższych sąsiadów) najczęściej wiąże się z mniejszym ryzykiem przeuczenia (ang. overfitting) niż 10-NN, choć w szczególnych przypadkach może być mniej skuteczny
- Jednym z celów redukcji złożoności drzewa decyzyjnego (ang. pruning) jest zmniejszenie ryzyka przeuczenia (ang. overfitting)

**Rozwiązanie:** F, T, F, T

**Zadanie 8 [Autor: KG, gr 1, ]** Mamy próbę prostą  $X_1, X_2, \dots, X_n$  pochodzącą z rozkładu o skończonej wartości oczekiwanej  $\mu$  i skończonej wariancji  $\sigma^2 > 0$ . Rozważamy estymator wartości oczekiwanej dla tego rozkładu:

$$\hat{\mu}_n = \frac{X_1 + \dots + X_n}{n + 1}$$

Wskaż poprawne stwierdzenia spośród poniższych.

- estymator  $\hat{\mu}_n$  jest asymptotycznie nieobciążony
- błąd średniokwadratowy estymatora  $\hat{\mu}_n$  wynosi:  $\frac{\mu^2 + n \cdot \sigma^2}{(n+1)^2}$
- wariancja estymatora  $\hat{\mu}_n$  wynosi:  $\frac{n}{n+1} \cdot \sigma^2$
- obciążenie estymatora  $\hat{\mu}_n$  wynosi:  $\frac{n}{(n+1)^2} \cdot \mu$

**Rozwiązanie:** T, T, F, F

**Zadanie 8 [Autor: KG, gr 2, ]** Treść zadania jak w gr 1.

- dla  $n \rightarrow \infty$  zachodzi  $Var(\hat{\mu}_n) \rightarrow 0$
- estymator  $\hat{\mu}_n$  jest asymptotycznie obciążony
- błąd średniokwadratowy estymatora  $\hat{\mu}_n$  wynosi:  $\frac{\mu + n \cdot \sigma^2}{(n+1)^2}$
- obciążenie estymatora  $\hat{\mu}_n$  wynosi:  $\frac{n}{n+1} \cdot \mu$

**Rozwiązanie:** T, F, F, F

**Zadanie 9 [Autor: DCK, gr 1]** Jak wiadomo z kolokwium, malutki Jaś od jakiegoś czasu każdego ranka w swoich dwóch butach odnajduje po nocy 0, 1 albo 2 cukierki, w taki sposób umieszczone, że w każdym bucie jest co najwyżej jeden cukierek. Jaś jest uzdolniony matematycznie i wobec tego jest przekonany, że nie odkrył jeszcze wszystkich sekretów skrzatów, które według niego podrzucają mu w nocy cukierki. Od 184 nocy liczy, ile cukierków łącznie danej nocy dostał i jego obserwacje są następujące: było 70 nocy, kiedy otrzymał dwa cukierki, 44 noce, kiedy otrzymał jeden cukierek i 70 nocy, kiedy w ogóle nie otrzymał cukierków. Jaś chciałby się dowiedzieć, jakie czynniki mogą mieć wpływ na to, że po niektórych nocach nie otrzymuje ani jednego cukierka. W tym celu zapisuje również następujące dane: subiektywną ocenę, czy był poprzedniego dnia grzeczny (*grzeczny*: 3 – bardzo, 2 – przeciętnie, 1 – zupełnie nie), liczbę godzin przespanych przez jego kota w badanym dniu (*kot*), aktualną fazę księżyca (*ksiezyc*: 1 – nów, 0 – pozostałe fazy) oraz liczbę gości w domu poprzedniego dnia (*goscie*). W oparciu o sporządzoną bazę, estymuje model logitowy, gdzie 1 – w danym dniu nie otrzymał cukierków, 0 – w danym dniu otrzymał co najmniej jeden cukierek; przyjęty próg odcięcia to  $p^* = 0,50$ . Jaś szacuje model na całości bazy (jeszcze nie dowiedział się o walidacji krzyżowej). Korzystając z wydruków zamieszczonych poniżej (oszacowania współczynników oraz macierz konfuzji), oceń prawdziwość zdań.

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.059571	0.0807450	13.12	< 2e-16
kot	0.0169666	0.0047302	3.587	0.00033
ksiezyc1	-0.0011382	0.0019075	-0.597	0.55071
grzeczny3	-0.5777502	0.0445178	-12.79	< 2e-16
grzeczny2	-0.1816369	0.0332112	-5.469	4.5e-08
goscie	0.1313188	0.0246349	5.331	9.8e-08

	Y = 1	Y = 0
$\hat{Y} = 1$	42	14
$\hat{Y} = 0$	28	100

- Wyższa liczba godzin przespanych w ciągu dnia przez kota zwiększa prawdopodobieństwo braku cukierka, efekt ten jest istotny statystycznie na poziomie istotności  $\alpha = 0,01$
- Model naiwny, prognozujący, że Jaś zawsze znajdzie cukierka, osiągałby na podanym zbiorze danych ( $n = 184$ ) dokładność (ang. accuracy) równą 0,62 (z dokładnością do dwóch miejsc po przecinku)
- 95% przedział ufności dla współczynnika przy zmiennej *ksiezyc1* zawiera zero
- False Positive Rate dla modelu Jasia przy przyjętym progu odcięcia wynosi 0,39 (z dokładnością do dwóch miejsc po przecinku)

**Rozwiązanie:** T, T, T, F

**Zadanie 9 [Autor: DCK, gr 2]** Treść zadania jak w gr 1.

- Dla nocy, przed którymi Jaś subiektywnie ocenił, że był przeciętnie grzeczny prawdopodobieństwo braku cukierka jest o 18,16 punktów procentowych niższe (z dokładnością do dwóch miejsc po przecinku) w porównaniu do nocy, przed którymi Jaś subiektywnie ocenił, że zupełnie nie był grzeczny.
- Model naiwny, prognozujący, że Jaś zawsze znajdzie cukierka, osiągałby na podanym zbiorze danych ( $n = 184$ ) dokładność (accuracy), która byłaby niższa niż dokładność osiągnięta przez model Jasia przy przyjętym progu odcięcia.
- Jeśli Jaś chce jak najdokładniej prognozować brak cukierka, powinien optymalizować swój model pod kątem swoistości (ang. specificity)
- Precyzja (ang. precision) dla modelu Jasia przy przyjętym progu odcięcia wynosi 0,70 (z dokładnością do dwóch miejsc po przecinku)

**Rozwiązanie:** F, T, F, F

**Zadanie 10 [Autor: DCK, gr 1]** Badacz wyestymował model regresji liniowej  $y = \mathbb{X}\beta + \varepsilon$ , uzyskując wektor oszacowań parametrów  $\hat{\beta}$ . Dla podanego modelu wszystkie założenia Klasycznego Modelu Regresji Liniowej były spełnione, w szczególności  $\varepsilon \sim N(0, \sigma^2 I)$ ; w modelu uwzględniono wyraz wolny. Następnie badacz stworzył macierz  $\mathbb{X}^* = \mathbb{X}\mathbb{A}$ , gdzie  $\mathbb{A}$  jest pewną macierzą nieosobliwą. W ostatnim kroku oszacował regresję  $y$  na  $\mathbb{X}^*$ , uzyskując z tego modelu wektor oszacowań parametrów  $\hat{\gamma}$ .

- $\hat{\gamma} = \mathbb{A}^{-1}\hat{\beta}$
- $\text{Var } \hat{\gamma} = \sigma^2 \mathbb{A}^{-1}(\mathbb{X}^T \mathbb{X})^{-1}(\mathbb{A}^{-1})^T$
- RSS z regresji  $y$  na  $\mathbb{X}^*$  jest równe RSS z regresji  $y$  na  $\mathbb{X}$
- $R^2$  w obu regresjach jest sobie równe

**Rozwiązanie:** T, T, T, T

**Zadanie 10 [Autor: DCK, gr 2]** Treść jak w grupie 1

- $\text{Var } \hat{\beta} = \sigma^2(\mathbb{X}^T \mathbb{X})^{-1}$
- TSS z regresji  $y$  na  $\mathbb{X}^*$  zmalało w stosunku do TSS z regresji  $y$  na  $\mathbb{X}$
- $\hat{\gamma} = \mathbb{A}^{-1}(\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T y$
- $R^2$  w regresji  $y$  na  $\mathbb{X}^*$  jest niższe niż w regresji  $y$  na  $\mathbb{X}$

**Rozwiązanie:** T, F, T, F

**Zadanie 11 [Autor: BM, gr 1]** Rozważmy problem regresji postaci

$$y = f(x) + \varepsilon.$$

Na danych treningowych  $\mathcal{D}$  nauczony został model  $\hat{f}$ , jako  $\hat{f} = \arg \min_{g \in \mathcal{F}} \sum_{i=1}^n (y_i - g(x_i))^2$  dla pewnej dopuszczalnej klasy funkcji  $\mathcal{F}$ . Oceń prawdziwość podanych zdań.

- Dla nowej obserwacji  $(x^*, y^*)$  zachodzi: jeśli  $\mathbb{E}_{\mathcal{D}, y^*} [(y^* - \hat{f}(x^*))^2 | x^*] = \text{Var}(\varepsilon)$ , to  $\hat{f}$  jest nieobciążony
- Na danych treningowych dla prawdziwej funkcji  $f$  zachodzi

$$\sum_{i=1}^n (y_i - f(x_i))^2 \geq \sum_{i=1}^n (y_i - \hat{f}(x_i))^2$$

- Dla nowej obserwacji  $(x^*, y^*)$  zachodzi

$$(f(x^*) - y^*)^2 \leq (\hat{f}(x^*) - y^*)^2$$

- Jeśli  $\hat{f}$  jest obliczony przy użyciu większej rodziny  $\mathcal{F}$  niż inny model  $\tilde{f}$ , to dla nowej obserwacji  $(x^*, y^*)$  zachodzi

$$(\tilde{f}(x^*) - y^*)^2 \leq (\hat{f}(x^*) - y^*)^2$$

**Rozwiązanie:** F, T, F, F

**Zadanie 11 [Autor: BM, gr 2]** Treść jak w grupie 1

- Jeśli  $\mathbb{E}(y - \hat{f}(x))^2 = \text{Var}(\varepsilon)$  to  $\hat{f}$  osiąga najmniejszy możliwy MSE.
- Zawsze istnieje obserwacja  $x_i$  ze zbioru treningowego taka, że

$$(y_i - f(x_i))^2 \geq (y_i - \hat{f}(x_i))^2$$

- Dla nowej obserwacji  $x^*$  i odpowiadającej jej  $y^*$  zachodzi

$$(f(x^*) - y^*)^2 \geq (\hat{f}(x^*) - y^*)^2$$

- Jeśli  $\hat{f}$  jest uzyskane w modelu bardziej elastycznym niż inny estymator  $\tilde{f}$  to dla danych testowych zachodzi

$$\sum_{i=1}^n (\tilde{f}(x_i) - y_i)^2 \leq \sum_{i=1}^n (\hat{f}(x_i) - y_i)^2$$

**Rozwiązanie:** T, T, F, F

**Zadanie 12 [Autor: BM, gr 1]** Rozważmy problem wielokrotnego testowania. Niech  $p_1, \dots, p_m$  będą  $p$ -wartościami kolejnych testów.

- Dla każdego testu o  $p_i < \alpha$ , z prawdopodobieństwem  $\frac{1}{m}$  odrzucamy  $H_0$ . Taka procedura kontroluje FWER na poziomie  $\alpha$ .
- Jeśli procedura Benjaminiego – Hochberga odrzuca  $H_0$  pewnego testu, to tak samo będzie w procedurze Bonferoniego.
- Jeśli wszystkie hipotezy zerowe są fałszywe, to procedury wielokrotnego testowania pogarszają  $ACC$  (dokładność).
- Procedury wielokrotnego testowania poprawiają moc testów.

**Rozwiązanie:** T, F, T, F

**Zadanie 12 [Autor: BM, gr 2]** Treść jak w grupie 1

- Jeśli wylosujemy do odrzucenia hipotezy zerowej testy z z prawdopodobieństwem  $\frac{\alpha}{m}$ , to będziemy kontrolować FWER na poziomie  $\alpha$ .
- Jeśli procedura Bonferoniego odrzuca  $H_0$  dla pewnego testu to tak samo będzie w procedurze Benjaminiego–Hochberga.
- Jeśli wszystkie hipotezy zerowe są prawdziwe to procedury wielokrotnego testowania poprawiają accuracy.
- Procedury wielokrotnego testowania pogarszają moc testów.

**Rozwiązanie:** T, T, T, T

**Zadanie 13 [Autor: PP, gr 1]** Rozważmy model logistyczny o zmiennej odpowiedzi  $y \in \{-1, 1\}$ , w którym  $\mathbf{P}\{y = 1\} = (1 + \exp(-f))^{-1}$ . Oceń poprawność wzorów na funkcję wiarygodności.

- $(1 + \exp(-2yf))^{-1}$
- $(1 + \exp(-yf/2))^{-1}$
- $\exp(yf)/(1 + \exp(f))$
- $(1 + \exp(-yf))^{-1}$

**Rozwiązanie:** F, F, F, T

**Zadanie 13 [Autor: PP, gr 2]** Rozważmy model logistyczny o zmiennej odpowiedzi  $y \in \{-1, 1\}$ , w którym  $\mathbf{P}\{y = 1\} = (1 + \exp(-f))^{-1}$ . Oceń poprawność wzorów na funkcję straty, czyli minus log-wiarygodność.

- $\log(1 + \exp(-yf/2))$
- $\log(1 + \exp(-2yf))$
- $\log(1 + \exp(-yf))$
- $yf - \log(1 + \exp(f))$

**Rozwiązanie:** F, F, T, F

**Zadanie 14 [Autor: PP, gr 1]** Rozważmy klasyfikację binarną o zmiennej odpowiedzi  $y \in \{-1, 1\}$  oraz funkcji straty  $\ell(y, f(x))$ . Załóżmy, że przy znanym  $x$  minimalizujemy  $\mathbf{E}_x \ell(y, f(x))$  ze względu na  $f(x)$ . Niech  $p_+ = \mathbf{P}_x\{y = 1\}$  oraz  $p_- = \mathbf{P}_x\{y = -1\}$ . Oceń prawdziwość poniższych zdań.

- Dla  $\ell(y, f(x)) = \exp(-yf(x))$  optymalna  $f(x) = (p_+ - p_-)/2$
- Dla  $\ell(y, f(x)) = (1 - yf(x))^2$  optymalna  $f(x) = p_+ - p_-$
- Dla  $\ell(y, f(x)) = \log(1 + \exp(-yf(x)))$  optymalna  $f(x) = \log p_+ - \log p_-$
- Dla  $\ell(y, f(x)) = (y - f(x))^2$  optymalna  $f(x) = (p_+ - p_-)/2$

**Rozwiązanie:** F, T, T, F



**Zadanie 14 [Autor: PP, gr 2 ]** Rozważmy klasyfikację binarną o zmiennej odpowiedzi  $y \in \{-1, 1\}$  oraz funkcji straty  $\ell(y, f(x))$ . Załóżmy, że przy znanym  $x$  minimalizujemy  $\mathbf{E}_x \ell(y, f(x))$  ze względu na  $f(x)$ . Niech  $p_+ = \mathbf{P}_x\{y = 1\}$  oraz  $p_- = \mathbf{P}_x\{y = -1\}$ . Oceń prawdziwość poniższych zdań.

- Dla  $\ell(y, f(x)) = (1 - yf(x))^2$  optymalna  $f(x) = \log p_+ - \log p_-$
- Dla  $\ell(y, f(x)) = \log(1 + \exp(-yf(x)))$  optymalna  $f(x) = (\log p_+ - \log p_-)/2$
- Dla  $\ell(y, f(x)) = \exp(-yf(x))$  optymalna  $f(x) = (\log p_+ - \log p_-)/2$
- Dla  $\ell(y, f(x)) = (y - f(x))^2$  optymalna  $f(x) = p_+ - p_-$

**Rozwiązanie:** F, F, T, T