Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

1. Übungsblatt

1. In einem Krankenhaus gab es in den vergangenen n Tagen $N_1, \dots, N_n$ Geburten.

   (a) Begründen Sie, weshalb $N_1, \dots, N_n$ näherungsweise als unabhängig und Poiss(λ)-verteilt mit unbekanntem Parameter $\lambda > 0$ angesehen werden können. Geben Sie das entsprechende statistische Modell formal an.

   (b) Untersuchen Sie in dem Modell den Schätzer $\hat{\lambda}_n := (N_1 + \cdots + N_n)/n$ von λ auf Erwartungstreue und bestimmen Sie seinen MSE.

   (c) Schätzer $\hat{\rho}_n$ eines abgeleiteten Parameters $\rho(\vartheta)$ in statistischen Modellen $(\mathcal{X}_n, \mathscr{F}_n, (\mathbb{P}_{\vartheta,n})_{\vartheta \in \Theta})$ für $n \in \mathbb{N}$ heißen *konsistent*, falls

   $$\forall \vartheta \in \Theta,\, \varepsilon > 0 : \lim_{n \to \infty} \mathbb{P}_{\vartheta,n}(|\hat{\rho}_n - \rho(\vartheta)| > \varepsilon) = 0$$

   gilt. Sind die Schätzer $\hat{\lambda}_n$ für $n \to \infty$ konsistent?

2. Für eine $N(\mu, \sigma^2)$-verteilte mathematische Stichprobe $X_1, \dots, X_n$ mit Parametern $\mu \in \mathbb{R}, \sigma^2 > 0$ betrachte den Schätzer $\hat{\sigma}_\alpha^2 := \frac{\alpha}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ von σ^2, wobei $\alpha > 0$ frei gewählt wird.

   (a) Zeigen Sie für den mittleren quadratischen Fehler (MSE)

   $$\mathbb{E}_{\mu,\sigma^2}[(\hat{\sigma}_\alpha^2 - \sigma^2)^2] = (\alpha - 1)^2 \sigma^4 + \alpha^2 \operatorname{Var}_{\mu,\sigma^2}(\hat{\sigma}_1^2).$$

   (b) Begründen Sie, weshalb $\operatorname{Var}_{\mu,\sigma^2}(\hat{\sigma}_1^2) = \sigma^4 V_n$ mit einer Konstanten V_n unabhängig von μ und σ^2 gilt. Schließen Sie, dass der MSE minimal ist für $\alpha = (1 + V_n)^{-1}$.

   (c) Bestimmen Sie V_n. Sie dürfen $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = (\frac{1}{n} \sum_{i=1}^n X_i^2) - \bar{X}^2$ sowie $\operatorname{Var}(X^2) = 2$ für $X \sim N(0,1)$ verwenden.

   (d) *(freiwillig)* Simulieren Sie das Modell für $n = 20$, $\mu = 0$, $\sigma^2 = 1$ in $M = 100$ Monte-Carlo-Iterationen und geben Sie (empirisch) Bias und MSE sowie die Verteilung von $\hat{\sigma}_\alpha^2$ als Boxplot ($\rightsquigarrow$ Funktion in Mathematik/Statistik-Software wie R) für geeignete Werte von α (in der Nähe von 1) aus.

Abgabe **zu Beginn** der Vorlesung am Montag, 1.11.21.

Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

2. Übungsblatt

1. Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind. Geben Sie jeweils eine kurze Begründung:

   (a) Auf dem statistischen Modell $(\{0,1\}^n, \mathcal{P}(\{0,1\}^n), (\mathrm{Ber}(p)^{\otimes n})_{p\in[0,1]})$ induziert die Statistik $T(x_1,\dots,x_n) = x_1 + \cdots + x_n$ das Modell $(\{0,1,\dots,n\}, \mathcal{P}(\{0,1,\cdots,n\}), (\mathrm{Bin}(n,p))_{p\in[0,1]})$.

   (b) Für eine $N(\mu,\sigma^2)$-verteilte mathematische Stichprobe $X_1,\dots,X_n$ ist $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ erwartungstreuer Schätzer von σ^2.

   (c) Für eine $\mathrm{Exp}(\lambda)$-verteilte mathematische Stichprobe $X_1,\dots,X_n$ ist $\hat{\lambda} = (\frac{1}{n}\sum_{i=1}^n X_i)^{-1}$ Momentenschätzer für $\lambda > 0$.

   (d) Der Schätzer $\hat{\lambda}$ in (c) ist auch MLE.

   (e) Der Schätzer $\hat{\lambda}$ in (c) ist erwartungstreu.

2. Unter N Wahlberechtigten gibt es N_{AfD} AfD-Wähler sowie jeweils $N/2$ Männer und Frauen.

   (a) Für gerades $n \in \mathbb{N}$ ziehe jeweils zufällig (mit Zurücklegen) $n/2$ Männer und $n/2$ Frauen und bestimme jeweils die Anzahl X_m bzw. X_w von AfD-Wählern/innen in den Stichproben. Formalisieren Sie das statistische Modell und zeigen Sie, dass $\hat{p}_{m/w} := \frac{X_m + X_w}{n}$ ein erwartungstreuer Schätzer von $p = N_{AfD}/N$ ist, dessen MSE nicht größer ist als der von $\hat{p} = \frac{X}{n}$, wenn sich X AFD-Wähler unter n geschlechtsneutral gezogenen Wahlberechtigten befinden.

   (b) Angenommen die Anzahl der AfD-Wähler unter den Männern ist etwa doppelt so groß wie der unter den Frauen. Wie können Sie durch geeignete Stichprobenzahlen einen erwartungstreuen Schätzer von p konstruieren, der noch kleineren MSE als $\hat{p}_{m/w}$ besitzt? (Stichwort: *stratified sampling*).

3. Es sei Y eine reellwertige Zufallsvariable mit Dichte f^Y bezüglich einem σ-endlichen Maß ν auf $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$. Zeigen Sie:

   (a) $m := \inf\{q \in \mathbb{R} \mid \int_{-\infty}^{q} f^Y(y)\nu(dy) \geqslant 1/2\}$ minimiert $\mathbb{E}[|Y - r|]$ über $r \in \mathbb{R}$.

   (b) $M := \operatorname{argmax}_{y\in\mathbb{R}} f^Y(y)\nu(\{y\})$ (dies möge existieren) minimiert $\mathbb{E}[\mathbf{1}(Y \neq r)]$ über $r \in \mathbb{R}$.

   Folgern Sie das Korollar zu Bayesschätzern unter absolutem und 0-1-Verlust aus der Vorlesung.

4. Es sei $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\{0,1\}})$ ein binäres statistisches Modell sowie π eine a-priori-Verteilung auf $\mathcal{P}(\{0, 1\})$. $L(\vartheta, x)$ bezeichne die Likelihoodfunktion bezüglich einem dominierenden Maß μ.

   (a) Bestimmen Sie die a-posteriori-Verteilung von ϑ.

   (b) Leiten Sie den Bayes-optimalen Schätzer bezüglich 0-1-Verlust $\ell(\vartheta, r) = \mathbf{1}(\vartheta \neq r)$ her und berechnen Sie sein Bayesrisiko. Was ergibt sich im Fall $\pi(\{0\}) = \pi(\{1\})$?

Abgabe **zu Beginn** der Vorlesung am Montag, 8.11.21.

Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

3. Übungsblatt

1. Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind. Geben Sie jeweils eine kurze Begründung:

   (a) Jeder Bayesschätzer ist erwartungstreu.

   (b) Die Menge der (randomisierten) Tests vom Niveau α ist konvex.

   (c) Es gibt stets einen randomisierten Test vom Niveau α, dessen Wahrscheinlichkeit für einen Fehler 2. Art $1-\alpha$ beträgt; nämlich den trivialen Test $\varphi(x)=\alpha$.

   (d) Der zweiseitige Binomialtest ist stets UMP.

   (e) Der Beweis des Neyman-Pearson-Lemmas zeigt, dass ein Test $\tilde{\varphi}$ mit $\mathbb{E}_0[\tilde{\varphi}] \leqslant \mathbb{E}_0[\varphi]$ und $\mathbb{E}_1[\tilde{\varphi}] = \mathbb{E}_1[\varphi]$ für einen Neyman-Pearson-Test φ sogar $\tilde{\varphi}=\varphi$ fast sicher unter $\mathbb{P}_0$ und $\mathbb{P}_1$ erfüllt.

2. Die Beta-Verteilung $B(a,b)$ auf $[0,1]$ ist gegeben durch die Dichte

$$f_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad x \in (0,1),$$

   wobei $a,b>0$ und Γ die Gamma-Funktion bezeichnet. $B(a,b)$ hat Erwartungswert $\mu_{a,b} = \frac{a}{a+b}$ und Varianz $\sigma^2_{a,b} = \frac{ab}{(a+b)^2(a+b+1)}$.

   (a) Die Beobachtung X sei $\mathrm{Bin}(n,p)$-verteilt, wobei $n \geqslant 1$ bekannt ist sowie der unbekannte Parameter p gemäß $B(a,b)$ a priori verteilt ist. Zeigen Sie, dass der Bayes-optimale Schätzer (bzgl. MSE und $B(a,b)$) gegeben ist durch

$$\hat{p}_{a,b} = \frac{a+X}{a+b+n}.$$

   (b) Bestimmen Sie $\mathrm{MSE}_p(\hat{p}_{a,b})$ für $p \in [0,1]$. Finden Sie $a^*, b^* > 0$ so, dass $p \mapsto \mathrm{MSE}_p(\hat{p}_{a^*,b^*})$ konstant ist.

   (c) Zeigen Sie mittels Bayes-Optimalität, dass aus $p \mapsto \mathrm{MSE}_p(\hat{p}_{a^*,b^*})$ konstant folgt, dass $\hat{p}_{a^*,b^*}$ minimax ist. Folgern Sie, dass der MLE $\hat{p} = \frac{X}{n}$ nicht minimax ist.

3. Es werde eine $\mathbb{P}_\vartheta$-verteilte mathematische Stichprobe $X_1, \dots, X_n$ beobachtet mit $\vartheta \in \{\vartheta_0, \vartheta_1\}$, $0 < \vartheta_0 < \vartheta_1$, unbekannt. Bestimmen Sie die Form der Neyman-Pearson-Tests für $H_0 : \vartheta = \vartheta_0$ gegen $H_1 : \vartheta = \vartheta_1$ in folgenden Modellen:

   (a) $\mathbb{P}_\vartheta = N(\vartheta, 1)$,

   (b) $\mathbb{P}_\vartheta = \mathrm{Exp}(\vartheta)$,

   (c) $\mathbb{P}_\vartheta = U([0, \vartheta])$.

4. Recherchieren Sie den Begriff des *p-Wertes* eines Tests. Bestimmen Sie für die folgenden Datensätze den p-Wert für einen zweiseitigen Test unter der Nullhypothese, dass die Wahrscheinlichkeit für *männlich* und *weiblich* gleich groß ist. Die Binomialverteilung darf durch die Normalverteilung approximiert werden. Klären Sie jeweils, ob die Nullhypothese zu den Niveaus $\alpha = 0,05$ bzw. $\alpha = 0,01$ akzeptiert oder abgelehnt wird.

   (a) 2015 wurden in Berlin 19 614 Jungen und 18 416 Mädchen geboren.

   (b) In der 4. Coronawelle gab es in Berlin bis zum 8.11.21 51 949 Neuinfektionen, darunter $50,2\%$ weibliche und $48,4\%$ männliche Personen (beim Rest ist das Geschlecht unbekannt oder divers).

Abgabe **zu Beginn** der Vorlesung am Montag, 15.11.21.

Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

4. Übungsblatt

1. Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind. Geben Sie jeweils eine kurze Begründung:

   (a) Eine Konfidenzmenge C vom Niveau $1-\alpha$ gibt an, dass der wahre Parameter mit Wahrscheinlichkeit $1-\alpha$ in C liegt.

   (b) Für $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ i.i.d. und $\alpha \in (0,1)$ ist $C = [\bar{X} + q_\alpha \sigma n^{-1/2}, \infty)$ ein einseitiges Konfidenzintervall für μ zum Niveau $1-\alpha$ (mit α-Quantil q_α von $N(0,1)$).

   (c) Im linearen Modell existiert stets ein Kleinste-Quadrate-Schätzer.

   (d) Im linearen Modell ist der Kleinste-Quadrate-Schätzer stets eindeutig.

   (e) Für eine symmetrische positiv semi-definite Matrix $\Sigma \in \mathbb{R}^{n\times n}$ gilt $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$.

2. Es werde eine $N(\mu, \sigma^2)$-verteilte mathematische Stichprobe beobachtet mit $\sigma > 0$ bekannt und $\mu \in \mathbb{R}$ unbekannt. Für welche $a, b > 0$ ist $C = [\bar{X} - a, \bar{X} + b]$ ein Konfidenzintervall für μ vom Niveau $1-\alpha$, $\alpha \in (0,1)$? Für welche dieser Werte von a, b besitzt C minimale Länge?

3. Betrachten Sie das gewöhnliche lineare Modell $Y = X\beta + \varepsilon$ mit $\varepsilon \sim N(0, \sigma^2 E_n)$. Bestimmen Sie den Likelihood-Quotiententest für die Nullhypothese $H_0 : \beta = \beta_0$ gegen $H_1 : \beta \neq \beta_0$ zum Niveau $\alpha \in (0,1)$. Hierbei seien $X \in \mathbb{R}^{n\times p}, \beta_0 \in \mathbb{R}^p, \sigma > 0$ bekannt.

4. (*Polynomregression*) Wir beobachten

$$Y_i = \beta_0 + \beta_1 x_i + \cdots + \beta_{p-1} x_i^{p-1} + \varepsilon_i, \quad i = 1, \ldots, n,$$

wobei $x_1, \ldots, x_n \in \mathbb{R}$ bekannte Designpunkte sind und $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d. gilt.

(a) Schreiben Sie die Beobachtungen als ein lineares Modell $Y = X\beta + \varepsilon$ mit einer geeigneten Designmatrix $X \in \mathbb{R}^{n \times p}$ und weisen Sie nach, dass für $p \leqslant n$ und paarweise verschiedene x_i die Designmatrix vollen Rang p besitzt (Tipp: *Vandermonde-Determinante*).

(b) Für $n = 100$, $x_i = (i-1)/(n-1)$, $\varepsilon_i \sim N(0; 0,1)$ simuliere Beobachtungen $Y_i = \sin(\pi x_i) + \varepsilon_i$, $i = 1, \ldots, n$. Für Dimensionen $p = 1, 2, 3, 5, 10, 50$ bestimme jeweils das Regressionspolynom $\hat{f}_p(x) = \sum_{i=0}^{p-1} \hat{\beta}_i x^i$ mit Kleinste-Quadrate-Schätzer $\hat{\beta}$. Zeichnen Sie f, alle $\hat{f}_p$ und die Beobachtungen Y_i in ein gemeinsames Koordinatensystem. Diskutieren Sie kurz das Ergebnis.

Abgabe **zu Beginn** der Vorlesung am Montag, 22.11.21.

Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

5. Übungsblatt

1. Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind. Geben Sie jeweils eine kurze Begründung:

   (a) Die Spur einer Orthogonalprojektion auf einen p-dimensionalen Unterraum beträgt p, so dass $\operatorname{tr}(\Pi_{X_\Sigma}) = p$ gilt.

   (b) Im gewöhnlichen linearen Modell $Y = X\beta + \varepsilon$ mit $p = n = 1$ ist $\hat{\sigma}^2 = Y^2$ stets erwartungstreuer Schätzer von σ^2.

   (c) Im linearen Modell $Y_i = \beta + \varepsilon_i$ mit $\varepsilon_i \sim U([-1, 1])$ i.i.d., $i = 1, \dots, n$, ist der *mid-range* $\hat{\beta}_{mr} = \frac{1}{2}(\max_i Y_i + \min_i Y_i)$ ein MLE von $\beta \in \mathbb{R}$.

   (d) $\hat{\beta}_{mr}$ ist erwartungstreuer Schätzer von β.

   (e) $\hat{\beta}_{mr}$ besitzt einen größeren MSE als der Kleinste-Quadrate-Schätzer $\hat{\beta}_{KQ} = \frac{1}{n}\sum_{i=1}^n Y_i$.

2. Betrachten Sie das lineare Modell $Y = X\beta + \varepsilon$, wobei die ε_i unabhängig und gemäß der Laplaceverteilungsdichte $f(x) = \frac{\lambda}{2}e^{-\lambda|x|}$, $x \in \mathbb{R}$, für ein $\lambda > 0$ verteilt sei.

   (a) Weisen Sie nach, dass der Maximum-Likelihood-Schätzer für β Lösung des folgenden Minimierungsproblems ist:

   $$\hat{\beta}^{ML} = \min_{b \in \mathbb{R}^p} \sum_{i=1}^n |Y_i - (Xb)_i|.$$

   (b) Schließen Sie, dass im Fall $p = 1$ und $X = (1, \dots, 1)^\top \in \mathbb{R}^{n \times p}$ jeder Stichprobenmedian von $Y_1, \dots, Y_n$ Maximum-Likelihood-Schätzer ist.

   (c) Betrachten Sie den Fall in (b). Zeigen Sie $\operatorname{Var}(\hat{\beta}_{KQ}) = \frac{2}{\lambda^2 n}$ für den Kleinste-Quadrate-Schätzer $\hat{\beta}_{KQ}$. Vergleichen Sie den MSE von $\hat{\beta}_{KQ}$ und $\hat{\beta}_{ML}$ in Simulationen für $n = 99$ und $\beta = 0$, $\lambda = 1$.

   (d**) Beweisen Sie $\sqrt{n}(\hat{\beta}_{ML} - \beta) \xrightarrow{d} N(0, \lambda^{-2})$ für $n \to \infty$ in der Situation von (b).

3. Eine physikalische Größe $\mu \in \mathbb{R}$ wird in n verschiedenen Apparaturen gemessen. Für die Messwerte $Y_1, \ldots, Y_n$ nehmen wir $\mathbb{E}[Y_i] = \mu$, $\mathrm{Var}(Y_i) = \sigma_i^2$ mit $\sigma_i > 0$ bekannt und $\mathrm{Cov}(Y_i, Y_j) = 0$ für $i \neq j$ an. Schreiben Sie dies als lineares Modell und weisen Sie nach, dass das gewichtete Mittel
$$\hat{\mu} = \frac{\sum_{i=1}^n \sigma_i^{-2} Y_i}{\sum_{i=1}^n \sigma_i^{-2}}$$
Kleinste-Quadrate-Schätzer ist. Vergleichen Sie die Varianz von $\hat{\mu}$ mit der des Stichprobenmittels $\bar{Y}$ und beweisen Sie in diesem Fall direkt, dass $\hat{\mu}$ kleinste Varianz unter allen linearen erwartungstreuen Schätzern besitzt.

4. Betrachten Sie das statistische Modell
$$Y_i = f(x_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$
mit n verschiedenen Punkten $x_i \in \mathbb{R}$, $f : \mathbb{R} \to \mathbb{R}$ beliebig und unbekannt sowie $\mathbb{E}[\varepsilon] = 0$, $\mathrm{Cov}(\varepsilon) = \sigma^2 E_n$. Unter der polynomiellen Regressionsannahme $f = q_a^d$ mit
$$q_a^d(x) = a_0 + a_1 x + \cdots + a_d x^d$$
sei $\hat{a} = (\hat{a}_0, \ldots, \hat{a}_d)$ der Kleinste-Quadrate-Schätzer sowie $q_{\hat{a}}^d(x)$ die Vorhersage des Funktionswerts an der Stelle $x \in \mathbb{R}$. Zeigen Sie unter der Verwendung der empirischen Norm $\|g\|_n^2 := \frac{1}{n} \sum_{i=1}^n g(x_i)^2$:

(a) Unter der Annahme $f = q_a^d$ gilt $\mathbb{E}_{q_a^d}[\|q_{\hat{a}}^d - q_a^d\|_n^2] = \frac{\sigma^2}{n}(d+1)$.
Tipp: Betrachten Sie $\|X\hat{\beta} - X\beta\|^2$ allgemein im linearen Modell.

(b) Für f beliebig im Modell ergibt sich die Bias-Varianz-Zerlegung
$$\mathbb{E}_f[\|q_{\hat{a}}^d - f\|_n^2] = \min_{q_a^d} \|f - q_a^d\|_n^2 + \frac{\sigma^2}{n}(d+1),$$
wobei sich der Bias durch den kleinsten Abstand von f zum Raum der Polynome vom maximalen Grad d in empirischer Norm ergibt.

(c) Gilt für jedes quadratische Polynom f, dass der MSE $\mathbb{E}_f[\|q_{\hat{a}}^d - f\|_n^2]$ für $d = 1$ größer als für $d = 2$ ist?

Abgabe **zu Beginn** der Vorlesung am Montag, 29.11.21.

Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

6. Übungsblatt

1. Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind. Geben Sie jeweils eine kurze Begründung:

   (a) Für unabhängige rellwertige Zufallsvariablen X, Y mit Lebesguedichten f^X, f^Y gilt $f^{XY}(z) = \int_{-\infty}^{\infty} f^Y(z/x) f^X(x) dx$.

   (b) Für $X \sim t(n)$ gilt $\mathbb{E}[|X|^p] < \infty$, $p > 0$, genau dann, wenn $p < n$.

   (c) Für $X, Y \sim N(0,1)$ unabhängig ist X/Y $t(1)$-verteilt, also Cauchy-verteilt.

   (d) Die Beobachtungen $Y_i = \mu_1 + \varepsilon_i$, $i = 1, \ldots, n_1$, und $Y_i = \mu_2 + \varepsilon_i$, $i = n_1 + 1, \ldots, n_1 + n_2$, mit $\varepsilon \sim N(0, \sigma^2 E_n)$, $n = n_1 + n_2$, lassen sich als lineares Modell mit $p = 2$, $\beta = (\mu_1, \mu_2)^\top$ und Designmatrixeinträgen $X_{ij} = \mathbf{1}(i \leqslant n_1, j = 1) + \mathbf{1}(i > n_1, j = 2)$ schreiben.

   (e) Das $(1 - \alpha)$-Konfidenzellipsoid für β in (d) ist stets ein Kreis.

2. Im linearen Modell $Y = X\beta + \varepsilon$ mit $\varepsilon \sim N(0, \sigma^2 E_n)$ sei $\rho = \langle v, \beta \rangle$ für ein $v \in \mathbb{R}^p \setminus \{0\}$ ein abgeleiteter Parameter. Zeigen Sie für das Testproblem $H_0 : \rho = \rho_0, \sigma > 0$ gegen $H_1 : \rho \neq \rho_0, \sigma > 0$, dass

$$\varphi_\alpha := \mathbf{1}\Big(|\hat{\rho} - \rho_0| > \hat{\sigma} \sqrt{\langle (X^\top X)^{-1} v, v \rangle} q_{t(n-p), 1-\alpha/2} \Big)$$

   mit $\hat{\rho} = \langle v, \hat{\beta} \rangle$ Likelihood-Quotienten-Test zum Niveau $\alpha \in (0,1)$ ist.

3. Untersuchen Sie im linearen Modell $Y = X\beta + \varepsilon$ mit $\varepsilon \sim N(0, \sigma^2 E_n)$ folgende lineare Testprobleme:

   (a) Test auf signifikanten Einfluss einer Kovariablen auf die Responsevariable: $H_0 : \beta_j = 0, \sigma > 0$ gegen $H_1 : \beta_j \neq 0, \sigma > 0$ für ein fest vorgegebenes $j \in \{1, \ldots, p\}$.

   (b) Test eines Untervektors $\beta^* = (\beta_1^*, \ldots, \beta_r^*)^\top \in \mathbb{R}^r$ mit $r \leqslant p$: $H_0 : \forall j = 1, \ldots, r : \beta_j = \beta_j^*, \sigma > 0$ gegen $H_1 : \exists j = 1, \ldots, r : \beta_j \neq \beta_j^*, \sigma > 0$.

   Stellen Sie diese Hypothesen in der Form $H_0 : K\beta = c$ mit geeigneter Kontrastmatrix K und mit einem Vektor c dar. Konstruieren Sie die zugehörigen F-Statistiken und bestimmen Sie deren Verteilung.

4. Betrachten Sie die lineare Hypothese $H_0 : K\beta = 0$ und den auf H_0 eingeschränkten Kleinste-Quadrate-Schätzer $\hat{\beta}_{H_0}$. Zeigen Sie im Fall von orthogonalem Design $X^\top X = nE_p$ mit der expliziten Formel für $\hat{\beta}_{H_0}$:

   (a) $\hat{\beta}_{H_0} = (E_p - \Pi_{K^\top})\hat{\beta}$;

   (b) $\hat{\beta}_{H_0}$ ist die Orthogonalprojektion von $\hat{\beta}$ auf den Kern $\ker(K) = \{\beta \in \mathbb{R}^p \mid K\beta = 0\}$ von K.

   (c) Interpretieren Sie im Fall orthogonalen Designs $RSS_{H_0} - RSS = |X(\hat{\beta} - \hat{\beta}_{H_0})|^2$ geometrisch.

Abgabe **zu Beginn** der Vorlesung am Montag, 6.12.21.

Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

7. Übungsblatt

1. Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind. Geben Sie jeweils eine kurze Begründung:

   (a) Das Quadrat einer $t(n)$-verteilten Zufallsvariable ist $F(1,n)$-verteilt.

   (b) Die Differenz zweier unabhängiger $F(m,n)$-verteilten Zufallsvariablen ist $F(2m,n)$-verteilt.

   (c) Eine Exponentialfamilie $(P_\vartheta)_{\vartheta\in\Theta}$ besitzt von ϑ unabhängigen Träger, also $\forall\vartheta,\vartheta'\in\Theta : \operatorname{supp}(P_\vartheta)=\operatorname{supp}(P_{\vartheta'})$.

   (d) Bilden $(P_\vartheta)_{\vartheta\in\Theta}$ und $(Q_\vartheta)_{\vartheta\in\Theta}$ Exponentialfamilien, so auch ihr Produkt $(P_\vartheta\otimes Q_\vartheta)_{\vartheta\in\Theta}$.

   (e) Der natürliche Parameterraum einer Exponentialfamilie ist stets konvex.

2. Beweisen oder widerlegen Sie die Aussage, dass folgende Verteilungen Exponentialfamilien bilden. Bestimmen Sie gegebenenfalls den natürlichen Parameterraum.

   (a) Poissonverteilung $(\mathrm{Poiss}(\lambda))_{\lambda>0}$

   (b) Multinomialverteilung $(M(p_0,\dots,p_s;n))_{0<p_i<1,\sum_{i=1}^s p_i=1}$;

   (c) p-dimensionale Normalverteilung $(N(\mu,\Sigma))_{\mu\in\mathbb{R}^p}$ mit bekannter Kovarianzmatrix $\Sigma\in\mathbb{R}^{p\times p}$;

   (d) Gleichmäßige Verteilung $(U([0,\vartheta]))_{\vartheta>0}$;

   (e) Gammaverteilung $(\Gamma(a,b))_{a,b>0}$.

3. Simulieren Sie selbständig die ANOVA-Tafeln für das Düngemittelbeispiel im Buch (Beispiel 2.53, basierend auf den vorherigen Beispielen). Belesen Sie sich zur zweifaktoriellen Varianzanalyse und erklären Sie kurz dieses Modell mit seinen Teststatistiken. Führen Sie auch dafür Simulationen durch analog zu Beispiel 2.59.

Abgabe **zu Beginn** der Vorlesung am Montag, 13.12.21.

Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

8. Übungsblatt

1. Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind. Geben Sie jeweils eine kurze Begründung:

   (a) Für $v, w \in \mathbb{R}^n$ gilt $|\langle v, w\rangle| \leqslant \frac{1}{2}|v|^2 + \frac{1}{2}|w|^2$.

   (b) Für die ℓ^p-Normen gilt im Fall $p \leqslant q$ stets $|v|_p \leqslant |v|_q$, $v \in \mathbb{R}^n$.

   (c) Ist $\zeta : \mathbb{R} \to \mathbb{R}$ konvex, so auch $\beta \mapsto \sum_{i=1}^n \zeta(\langle x_i, \beta\rangle)$ für $\beta \in \mathbb{R}^p$ und beliebige $x_1, \dots, x_n \in \mathbb{R}^p$.

   (d) Besitzt $X \in \mathbb{R}^{n\times p}$ nur Einträge $|X_{ij}| \leqslant 1$, so gilt $|X\beta| \leqslant |\beta|$ für alle $\beta \in \mathbb{R}^p$.

   (e) In einem korrekt spezifizierten GLM mit wahrem Parameter β_0 erfüllt das Exzessrisiko $\mathcal{E}(\hat\beta) \leqslant \frac{1}{n}\langle Y - \mathbb{E}[Y], X(\hat\beta - \beta_0)\rangle$.

2. Die Poissonverteilung mit Parameter $\lambda > 0$ bildet eine Exponentialfamilie in $T(k) = k$ mit natürlichem Parameter $\vartheta = \log\lambda \in \mathbb{R}$. Überprüfen Sie die Identitäten $\mathbb{E}_\vartheta[T] = \zeta'(\vartheta)$, $\mathrm{Var}_\vartheta(T) = \zeta''(\vartheta)$.

   Betrachten Sie nun das GLM-Modell der Poissonregression mit $\log\lambda_i = ax_i + b$ für gegebene Designpunkte $x_1, \dots, x_n \in \mathbb{R}$ und $a, b \in \mathbb{R}$ unbekannt. Stellen Sie eine Gleichung für den MLE auf und untersuchen Sie, ob der MLE existiert und eindeutig ist. Wie sieht die Iteration in Fishers Scoring-Methode aus?

   Lektüretipp: In *Report 490 des Imperial College* werden *Growth, population distribution and immune escape of Omicron in England* mittels logistischer und Poisson-Regression untersucht.

3. In der Vorlesung wurde $M_X := \sqrt{n/p}\max_{j=1,\dots,n}|\Pi_X e_j|$ definiert für die Orthogonalprojektion Π_X auf das Bild der Designmatrix $X \in \mathbb{R}^{n\times p}$ vom Rang p. Zeigen Sie:

   (a) $M_X \leqslant \sqrt{n/p}$.

   (b) Es gilt $\sum_{j=1}^n |\Pi_X e_j|^2 = p$, so dass $M_X \geqslant 1$ folgt.

   (c) Mit den Bezeichnungen $\Sigma_n = \frac{1}{n}\sum_{i=1}^n x_i x_i^\top$ für $x_i = X^\top e_i$ sowie $\|X\|_{max} = \max_{i,j}|X_{ij}|$ gilt

$$M_X \leqslant \frac{\|X\|_{max}}{\lambda_{min}(\Sigma_n)^{1/2}}.$$

4. Für jedes $\tau > 0$ gelte mit Wahrscheinlichkeit mindestens $1 - e^{-\tau}$ die Orakelungleichung

$$\mathcal{E}(\hat{\beta}) \leqslant C \inf_{\beta \in \mathbb{R}^p} \mathcal{E}(\beta) + F(\tau)$$

für eine Konstante $C \geqslant 1$ und eine deterministische, monoton wachsende Funktion $F : \mathbb{R}^+ \to \mathbb{R}^+$. Folgern Sie für den Erwartungswert die Orakelungleichung

$$\mathbb{E}\left[\mathcal{E}(\hat{\beta})\right] \leqslant C \inf_{\beta \in \mathbb{R}^p} \mathcal{E}(\beta) + \mathbb{E}\left[F(Z)\right]$$

mit einer Exp(1)-verteilten Zufallsvariablen Z.
Tipp: Für nicht-negative Zufallsvariablen X wurde in Stochastik I $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geqslant t)\, dt$ bewiesen.

Abgabe **zu Beginn** der Vorlesung am Montag, 10.1.22.

Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

9. Übungsblatt

1. Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind. Geben Sie jeweils eine kurze Begründung.

   Dabei seien (X, Y) eine $\mathcal{X} \times \{-1, +1\}$-wertige Zufallsvariable, $\pi_{\pm 1} = \mathbb{P}(Y = \pm 1) \in (0, 1)$, $\mathbb{P}_{\pm 1}(B) = P(X \in B \,|\, Y = \pm 1)$ für Ereignisse $B \subseteq \mathcal{X}$, $p_{\pm 1}$ Dichten von $\mathbb{P}_{\pm 1}$ bezüglich einem Maß μ auf $\mathcal{X}$ und $\eta(x) = \pi_{+1} p_{+1}(x)/(\pi_{+1} p_{+1}(x) + \pi_{-1} p_{-1}(x))$.

   (a) $\mathbb{P}(X \in B, Y = +1) = \pi_{+1}\, \mathbb{P}_{+1}(B)$ für alle Ereignisse B.

   (b) $\mathbb{P}(X \in B) = \pi_{+1}\, \mathbb{P}_{+1}(B) + \pi_{-1}\, \mathbb{P}_{-1}(B)$ für alle Ereignisse B.

   (c) $\mathbb{P}_{+1}(B) + \mathbb{P}_{-1}(B) = 1$ für alle Ereignisse B.

   (d) X besitzt die μ-Dichte $p(x) := \pi_{+1} p_{+1}(x) + \pi_{-1} p_{-1}(x)$, $x \in \mathcal{X}$.

   (e) $\mathbb{E}[\eta(X)\varphi(X)] = \mathbb{E}[\mathbf{1}(Y = +1)\varphi(X)]$ für alle beschränkten, messbaren Funktion $\varphi : \mathcal{X} \to \mathbb{R}$, insbesondere $\mathbb{E}[\eta(X)] = \mathbb{P}(Y = +1)$.

2. Es sei $\hat{\beta}$ der Maximum-Likelihood-Schätzer bei der logistischen Regression mit Absolutglied, das heißt, $X_{i,1} = 1$ gilt für $i = 1, \ldots, n$. Zeigen Sie, dass dann die beobachtete Anzahl der Einsen mit der erwarteten Anzahl übereinstimmt:

$$\sum_{i=1}^n Y_i = \sum_{i=1}^n p(\langle x_i, \hat{\beta} \rangle) \text{ mit } p(\eta) := e^\eta (1 + e^\eta)^{-1},\ x_i = X^\top e_i.$$

3. Beweisen Sie die *Höffding-Ungleichung*: Für positive Zahlen R_i sowie unabhängige und zentrierte Zufallvariablen X_i mit $|X_i| \leqslant R_i$ f.s. gilt

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n X_i\Big| \geqslant t\Big) \leqslant 2 \exp\Big(-\frac{t^2}{2\sum_{i=1}^n R_i^2}\Big), \quad t \geqslant 0.$$

   Zeigen Sie dazu $e^{\alpha x} \leqslant \frac{R-x}{2R} e^{-\alpha R} + \frac{R+x}{2R} e^{\alpha R}$ für $\alpha \geqslant 0$ und $|x| \leqslant R$ und schließen Sie $\mathbb{E}[e^{\alpha X_i}] \leqslant e^{\alpha^2 R_i^2/2}$. Verwenden Sie die Markovungleichung in geeigneter Weise, um zunächst $\mathbb{P}(\sum_{i=1}^n X_i \geqslant t)$ abzuschätzen.

   Nun sei $X_i = Y_i - \mathbb{E}[Y_i]$ mit $Y_i \sim \mathrm{Bern}(p)$ für $p \in (0, 1)$. Vergleichen Sie in diesem Fall die Schranke der Höffding-Ungleichung mit der der Tschebyschew-Ungleichung.

4. Wir betrachten die Klassifikation bei $K \in \mathbb{N}$ Klassen, es gilt also $Y \in \{1, \dots, K\}$. Zeigen Sie, dass bezüglich dem Klassifikationsfehler $R(C) = \mathbb{P}(C(X) \neq Y)$ eines Klassifizierers $C : \mathcal{X} \to \{1, \dots, K\}$ der Bayesklassifizierer

$$C^*(x) := \operatorname{argmax}_{k=1,\dots,K} \eta_k(x) \text{ mit } \eta_k(x) := \mathbb{P}(Y = k \mid X = x)$$

minimalen Fehler besitzt. Wie groß ist das entsprechende Bayes-Risiko?

Abgabe **zu Beginn** der Vorlesung am Montag, 17.1.22.

Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

10. Übungsblatt

1. Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind. Geben Sie jeweils eine kurze Begründung.

   (a) Jeder Zufallsvektor $X \in \mathbb{R}^p$ mit $|X| \leqslant R$ f.s. für eine Konstante $R \in (0, \infty)$ ist (Σ, C)-subgaußsch für geeignete Σ, C.

   (b) Sind X, Y (Σ, C)-subgaußsch, so ist $X + Y$ $(\Sigma, 2C)$-subgaußsch.

   (c) Aus $\sup_n \mathbb{E}[|X_n|^p] < \infty$ für ein $p > 0$ folgt $X_n = O_P(1)$.

   (d) $X_n = O_P(1)$ impliziert $X_n \xrightarrow{P} 0$ (stochastisch).

   (e) Aus $X_n = O_P(a_n)$, $Y_n = O_P(b_n)$ folgt $X_n + Y_n = O_P(a_n + b_n)$, $X_n Y_n = O_P(a_n b_n)$.

2. Betrachten Sie für Klassifizier $C : \mathcal{X} \to \{-1, +1\}$ und $\alpha \in (0, 1)$ den gewichteten Klassifikationsfehler

   $$R_\alpha(C) = \alpha P\big(C(X) = -1, Y = +1\big) + (1 - \alpha) P\big(C(X) = +1, Y = -1\big).$$

   Bestimmen Sie unter Benutzung von $\eta(x) = P(Y = +1 \,|\, X = x)$ den zugehörigen Bayes-Klassifizierer C_α^*, der $R_\alpha(C_\alpha^*) = \min_C R_\alpha(C)$ erfüllt. Wie groß ist $R_\alpha(C_\alpha^*)$?

3. Für einen Klassifizierer C und eine mathematische Stichprobe (*Trainingsdaten*) $(X_i, Y_i)_{1 \leqslant i \leqslant n}$ bezeichnet

   $$R_n(C) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(Y_i \neq C(X_i))$$

   das sogenannte *empirische Risiko*. $\hat{C}$ heißt *ERM-Klassifizierer* (*empirical risk minimizer*) in einer Klasse $\mathscr{C}$ von Klassifizierern, falls $R_n(\hat{C}) = \min_{C \in \mathscr{C}} R_n(C)$ gilt. Zeigen Sie für das *Risiko* (den Klassifizierungsfehler) R die Fundamentalungleichung

   $$R(\hat{C}) \leqslant \inf_{C \in \mathscr{C}} R(C) + 2 \sup_{C \in \mathscr{C}} |R_n(C) - R(C)|.$$

4. Betrachten Sie eine endliche Familie $\mathscr{C} = \{C_1, \ldots, C_M\}$ von Klassifizierern und den zugehörigen ERM-Klassifizierer $\hat{C}$. Verwenden Sie Aufgabe 3 und die Hoeffding-Ungleichung, um für alle $\tau > 0$ zu zeigen, dass mit Wahrscheinlichkeit mindestens $1 - e^{-\tau}$

$$R(\hat{C}) \leqslant \min_{1 \leqslant m \leqslant M} R(C_m) + \frac{\sqrt{8(\log(2M) + \tau)}}{\sqrt{n}}.$$

Abgabe **zu Beginn** der Vorlesung am Montag, 24.1.22.

Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

11. Übungsblatt

1. Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind. Geben Sie jeweils eine kurze Begründung.

   (a) Der Bayesklassifizierer hat die Form eines nicht-randomisierten Neyman-Pearson-Tests für $H_0 : P = P_{+1}$ gegen $H_1 : P = P_{-1}$.

   (b) Durch Randomisierung im Fall $\eta(x) = 1/2$ kann der Klassifikationsfehler des Bayesklassifizierers verringert werden.

   (c) Im Fall $X \sim N(0, \mathbf{E}_p)$ gilt für die empirische Kovarianzmatrix stets $\mathbb{E}[\|\Sigma_n - \mathbf{E}_p\|^2] \geqslant \sum_{j=1}^p \mathbb{E}[((\Sigma_n - \mathbf{E}_p)_{1,j})^2] \geqslant \frac{p}{n}$.

   (d) Im Fall $\pi_{+1} = 1$ ist $C^*(x) = -1$ der Bayesklassifizierer im LDA-Modell.

   (e) Im Modell von LDA und logistischer Regression gilt jeweils $\eta(x) = S(\langle x, \beta\rangle + c)$ mit $S(z) = e^z/(1+e^z)$ und geeigneten $\beta \in \mathbb{R}^p$, $c \in \mathbb{R}$.

2. Betrachten Sie für K Klassen mit Klassenwahrscheinlichkeiten $\pi_k \in (0, 1)$, $\sum_{k=1}^K \pi_k = 1$, Mittelwerten $\mu_k \in \mathbb{R}^p$ und invertierbaren Kovarianzmatrizen $\Sigma_k \in \mathbb{R}^{p\times p}$ die Normalverteilungsmischungsdichte

$$f(x) = \sum_{k=1}^K \pi_k \varphi_{\mu_k,\Sigma_k}(x), \quad x \in \mathbb{R}^p\,.$$

   Weisen Sie nach, dass der beste Klassifizierer (bei Standard-Klassifikationsfehler) gegeben ist durch $C^*(x) = \operatorname{argmax}_{k=1,\dots,K} \delta_k(x)$ mit quadratischen Diskrimanten

$$\delta_k(x) = -\tfrac{1}{2}\log(\det(\Sigma_k)) - \tfrac{1}{2}\langle \Sigma_k^{-1}(x - \mu_k), x - \mu_k\rangle + \log \pi_k.$$

   Welche geometrischen Formen für $p = 2$ können die Entscheidungsgrenzen $\{x \in \mathbb{R}^2 \mid \delta_k(x) = \delta_l(x)\}$ zwischen Klasse k und l besitzen?

3. Weisen Sie nach, dass das Bayesrisiko im LDA-Modell gegeben ist durch

$$R^* = \pi_{-1}\Phi\Big(\big(\log(\tfrac{\pi_{+1}}{\pi_{-1}})-\Delta^2/2\big)/\Delta\Big)+\pi_{+1}\Big(1-\Phi\Big(\big(\log(\tfrac{\pi_{+1}}{\pi_{-1}})+\Delta^2/2\big)/\Delta\Big)\Big)$$

mit der $N(0,1)$-Verteilungsfunktion Φ und dem *Mahalanobisabstand* Δ. Diskutieren Sie das Verhalten für $\pi_{+1} \to 1$, $\Delta \to \infty$ und $\Delta \to 0$.

4. Klassifikation in der Praxis: Laden Sie die Cleveland-Daten (`http://www.archive.ics.uci.edu/ml/datasets/heart+disease`, vgl. Beispiel 5.1 im Buch) herunter. Benutzen Sie die ersten $n = 151$ Patienten mit den Kovariablen Ruheblutdruck und maximale Herzfrequenz und ihren Labels (0=gesund, 1-4=krank) als Trainingsdaten. Die restlichen $m = 150$ Patienten dienen als *Testdaten*.

   (a) Zeichnen Sie die Kovariablen aller Patienten in ein Koordinatensystem und markieren Sie die Fälle gesund/krank (mit Farben) sowie Training-/Test-Datum (mit dunkel/hell).

   (b) Führen Sie eine logistische Regression für die Trainingsdaten durch (mit $p = 3$, d.h. unter Verwendung eines konstanten Glieds), zeichnen Sie die erhaltene Entscheidungsgrenze in das Koordinatensystem und bestimmen Sie den Klassifikationsfehler jeweils auf den Trainings- und Testdaten (relative Häufigkeit von Fehlklassifikationen).

   (c) Führen Sie eine lineare Diskriminanzanalyse durch und verfahren Sie wie in (b).

Abgabe **zu Beginn** der Vorlesung am Montag, 31.1.22.

Vorlesung *Methoden der Statistik*
Wintersemester 2021/22
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dr. Martin Wahl

12. Übungsblatt

1. Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind. Geben Sie jeweils eine kurze Begründung.

   (a) Die Entscheidungsgrenze des kNN-Klassifizierers ist stets linear.

   (b) Je größer k ist, desto größer ist der stochastische Fehler beim kNN-Klassifizierer.

   (c) Der kNN-Klassifizierer mit $k = 1$ klassifiziert alle X_i korrekt, wenn $X_i \neq X_j$ für $i \neq j$ gilt.

   (d) Besitzt X eine positive Lebesguedichte f^X auf dem Hyper-Würfel $\mathcal{X} = [0,1]^p$ mit $p \geqslant 2$, so ist $\liminf_{t\downarrow 0} \mathbb{P}(|X-x| \leqslant t)/t > 0$ für alle $x \in \mathcal{X}$.

   (e) Ist $f : \mathbb{R} \to \mathbb{R}$ stetig und existieren die einseitigen Ableitungen $f'(x+)$, $f'(x-)$, so gilt $f'(x^*-) \leqslant 0 \leqslant f'(x^*+)$ an jeder Minimalstelle x^* von f.

2. Betrachten Sie die Funktion

   $$J : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^+ \text{ mit } J(\beta,\beta_0) = \frac{1}{n}\sum_{i=1}^n \big(1 - Y_i(\langle X_i,\beta\rangle + \beta_0)\big)_+ + \tfrac{\lambda}{2}|\beta|^2$$

   für gegebene $Y_i \in \{-1,+1\}$, $X_i \in \mathbb{R}^p$, $\lambda > 0$. Zeigen Sie:

   (a) Die Funktion $\beta \mapsto J(\beta,\beta_0)$ ist für jedes $\beta_0 \in \mathbb{R}$ strikt konvex und besitzt genau ein Minimum $\hat\beta = \hat\beta(\beta_0)$. J besitzt ein Minimum $(\hat\beta,\hat\beta_0)$ auf $\mathbb{R}^p \times \mathbb{R}$, das aber nicht notwendigerweise eindeutig ist.

   (b) Für beliebige Richtungsvektoren $v \in \mathbb{R}^p$ erhalten wir die einseitige Ableitung

   $$\lim_{h\downarrow 0} \frac{J(\beta + hv,\beta_0) - J(\beta,\beta_0)}{h} = \lambda\langle\beta,v\rangle +$$
   $$\frac{1}{n}\sum_{i=1}^n \Big(- Y_i\langle X_i,v\rangle \mathbf{1}\big(Y_i(\langle X_i,\beta\rangle + \beta_0) < 1\big) + (Y_i\langle X_i,v\rangle)_- \mathbf{1}\big(Y_i(\langle X_i,\beta\rangle + \beta_0) = 1\big)\Big).$$

(c) Ist $v \in \mathbb{R}^p$ orthogonal zu allen Punkten X_i mit $Y_i(\langle X_i, \hat\beta\rangle + \beta_0) = 1$ für den Minimierer $\hat\beta$, so gilt

$$\langle \hat\beta, v\rangle = (\lambda n)^{-1} \sum_{i=1}^n Y_i \langle X_i, v\rangle \mathbf{1}\big(Y_i(\langle X_i, \hat\beta\rangle + \beta_0) < 1\big).$$

(d) Es folgt die Existenz einer Darstellung

$$\hat\beta = \sum_{i=1}^n \alpha_i Y_i X_i$$

mit $\alpha_i = (\lambda n)^{-1}$ im Fall $Y_i(\langle X_i, \hat\beta\rangle + \beta_0) < 1$, $\alpha_i = 0$ im Fall $Y_i(\langle X_i, \hat\beta\rangle + \beta_0) > 1$ und geeigneten α_i im Fall $Y_i(\langle X_i, \hat\beta\rangle + \beta_0) = 1$.

3. Betrachten Sie $(\hat\beta, \hat\beta_0) \in \operatorname{argmin}_{\beta,\beta_0} J(\beta, \beta_0)$ von oben und den SV-Klassifizierer $\hat C^{SV}(x) = \operatorname{sgn}(Y_i(\langle x, \hat\beta\rangle + \hat\beta_0))$. Punkte X_i mit $\alpha_i \neq 0$ in der Darstellung 2(d) heißen *Stützvektoren* (*support vectors*). Punkte X_i mit $Y_i(\langle X_i, \hat\beta\rangle + \beta_0) = 1$ liegen *auf dem margin* und solche mit $|Y_i(\langle X_i, \hat\beta\rangle + \beta_0)| > 1$ (bzw. < 1) *außerhalb (bzw. innerhalb) des margin*. Weisen Sie nach:

   (a) Punkte X_i innerhalb des margin sowie von $\hat C^{SV}$ falsch klassifizierte Punkte X_i sind Stützvektoren.

   (b) Von $\hat C^{SV}$ korrekt klassifizierte Punkte X_i außerhalb des margin sind keine Stützvektoren.

   (c) Punkte X_i auf dem margin können Stützvektoren sein, müssen aber nicht. (*Tipp:* Denken Sie an Beispiele, wo alle Punkte auf dem margin liegen.)

4. Finden Sie mindestens fünf Fehler im Buch.

Abgabe **zu Beginn** der Vorlesung am Montag, 7.2.22.