Dr. Stefan Richter
Statistische Analyse von maschinellen Lernalgorithmen
Wintersemester 2020/2021

4. Übungsblatt

Aufgabe 1 (Berechnung Erwartungswert in Satz 3.14).
Seien ε_i, $i = 1, ..., n$ i.i.d. Rademacher-verteilt ($\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = \frac{1}{2}$) und $X_i \sim N(0, \Sigma)$, $i = 1, ..., n$ i.i.d (unabhängig von den ε_i). Wir zeigen hier, dass

$$\mathbb{E}\Big[\max_{j=1,...,d}\Big|\sum_{i=1}^{n}\varepsilon_i X_{ij}\Big|\Big] \le 2\sqrt{n}\sqrt{\log(2d)} \cdot \max_{j=1,...,d}\Sigma_{jj}^{1/2}.$$

(a) Definiere $W_j := \frac{1}{\sqrt{n}\Sigma_{jj}^{1/2}}\sum_{i=1}^{n}\varepsilon_i X_{ij}$. Zeigen Sie, dass gegeben ε_i, $i = 1, ..., n$ gilt:

$$W_j \sim N(0, 1)$$

sowie $\mathbb{E}\Big[\max_{j=1,...,d}\Big|\sum_{i=1}^{n}\varepsilon_i X_{ij}\Big|\Big] \le \mathbb{E}[\max_{j=1,...,d}|W_j|] \cdot \sqrt{n}\max_{j=1,...,d}\Sigma_{jj}^{1/2}$.

(b) Die Funktion $\psi(x) = \exp(x^2/4)$ ist konvex. Zeigen Sie mit Jensen's Ungleichung, dass

$$\psi\big(\mathbb{E}[\max_{j=1,...,d}|W_j|]\big) \le \sum_{j=1}^{d}\mathbb{E}\psi(|W_j|).$$

(c) Folgern Sie aus (b): $\mathbb{E}[\max_{j=1,...,d}|W_j|] \le 2\log(2d)$.
Hinweis: Es gilt $\mathbb{E}[\exp(W_j^2/4)] = \sqrt{2}$.

Aufgabe 2 (Modellannahme der logistischen Regression).
Sei $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Y} = \{+1, -1\}$ (Klassifikationsproblem mit 2 Klassen). Es sei $(X, Y) \ll \lambda \times \mu_Z$, wobei λ das Lebesguemaß auf $\mathcal{X}$ und μ_Z das Zählmaß auf $\mathcal{Y}$ bezeichnet. Es sei g_k die bedingte Dichte von X gegeben $Y = k$, und $\pi_1 := \mathbb{P}(Y = 1)$. Erinnerung: $\eta(x) := \mathbb{P}(Y = 1|X = x)$.

(a) Ermitteln Sie die Dichte $g(x)$ von X.
Hinweis: Satz von der totalen Wahrscheinlichkeit.

(b) Zeigen Sie, dass

$$\log\Big(\frac{\eta(x)}{1-\eta(x)}\Big) = \log\Big(\frac{\pi_1}{1-\pi_1}\Big) + \log\Big(\frac{g_1(x)}{g_{-1}(x)}\Big).$$

Hinweis: Satz von Bayes.

Es sei nun für $k \in \mathcal{Y}$

$$g_k(x) = \frac{1}{((2\pi)^d\det(\Sigma_k))^{1/2}}\exp\Big(-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)\Big)$$

die Dichte einer $N(\mu_k, \Sigma_k)$-Verteilung, wobei $\mu_k \in \mathbb{R}^d$, $\Sigma_k \in \mathbb{R}^{d\times d}$ positiv definit.

(c) Ermitteln Sie $\log(\frac{g_1(x)}{g_{-1}(x)})$.

(d) Zeigen Sie: ist $\Sigma_1 = \Sigma_2$, $\pi_1 = \frac{1}{2}$ und $\mu_1^T \Sigma^{-1} \mu_1 = \mu_{-1}^T \Sigma^{-1} \mu_{-1}$, so liefert die Bayes-Regel lineare optimale Entscheidungsränder, und die Modellannahme der logistischen Regression ist erfüllt für (X, Y).

(e) Geben Sie eine einfache Bedingung an μ_1, μ_{-1} an, so dass $\mu_1^T \Sigma^{-1} \mu_1 = \mu_{-1}^T \Sigma^{-1} \mu_{-1}$.

(f) Zeigen Sie: ist $\Sigma_1 = \Sigma_2$, so liefert die Bayes-Regel affin lineare optimale Entscheidungsränder, und die Modellannahme der logistischen Regression ist erfüllt für $(\tilde{X}, Y)$, wobei $\tilde{X} = (1, X)$.

(g) Zeigen Sie: Die Bayes-Regel liefert quadratische optimale Entscheidungsränder und die Modellannahme der logistischen Regression ist erfüllt für $(\tilde{X}, Y)$, wobei $\tilde{X} = h(X)$ mit $h(x) = (1, (x_j)_{j=1,\dots,d}, (x_j x_l)_{1 \le j \le l \le d})$.

(h) Welche Konvergenzraten erwartet man für den Klassifizierer der logistischen Regression aus Satz 3.12 angewandt auf $(\tilde{X}, Y)$ in den Fällen (d),(f),(g)?

Aufgabe 3 (Berechnung eines Bayes-Risikos bei Klassifikation).

Sei $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Y} = \{+1, -1\}$ (Klassifikationsproblem mit 2 Klassen). Es gelten dieselben Voraussetzungen wie in Aufgabe 2, insbesondere sei X gegeben $X = k$ normalverteilt $N(\mu_k, \Sigma)$ mit Dichte

$$g_k(x) = \frac{1}{((2\pi)^d \det(\Sigma))^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1} (x - \mu_k)\Big),$$

wobei $\mu_k \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$ positiv definit. In dieser Aufgabe ermitteln wir das Bayes-Risiko $R(f^*)$.

(a) Definiere $\delta^*(x) := \log(\frac{\eta(x)}{1-\eta(x)})$. Zeigen Sie, dass

$$f^* = \operatorname{sign}(\delta^*),$$

und ermitteln Sie $\delta^*(x)$.

(b) Zeigen Sie, dass

$$R(f^*) = \pi_1 \mathbb{P}(\delta^*(X) < 0 | Y = 1) + (1 - \pi_1)\mathbb{P}(\delta^*(X) > 0 | Y = -1).$$

(c) Zeigen Sie, dass gegeben $Y = 1$ gilt: $\delta^*(X) \sim N(T + \frac{1}{2}\Delta, \Delta)$ und gegeben $Y = -1$ gilt: $\delta^*(X) \sim N(T - \frac{1}{2}\Delta, \Delta)$, wobei

$$\Delta := (\mu_1 - \mu_{-1})^T \Sigma^{-1} (\mu_1 - \mu_{-1}) \qquad \text{sog. Mahalanobis-Distanz,} \qquad T := \log\Big(\frac{\pi_1}{1-\pi_1}\Big).$$

(d) Zeigen Sie, dass

$$R(f^*) = \pi_1 \Phi\Big(\frac{-T - \frac{1}{2}\Delta}{\sqrt{\Delta}}\Big) + (1 - \pi_1)\Big(1 - \Phi\Big(\frac{-T + \frac{1}{2}\Delta}{\sqrt{\Delta}}\Big)\Big),$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung bezeichnet.

(e) Zeigen Sie: Im Spezialfall $\pi_1 = \frac{1}{2}$, $\Sigma = I_{d \times d}$, $\mu_{-1} = -\mu_1$ ist $R(f^*) = \Phi(-\|\mu_1\|_2)$.

Homepage der Vorlesung:
`https://ssp.math.uni-heidelberg.de/sam-ws2020/`