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

2. Übungsblatt

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Aufgabe 1 (Diskussion: Aussage von Satz 2.10).
Die Konvergenzrate des Lasso-Schätzers aus Satz 2.10. lautet

$$\frac{1}{\Lambda_{min}(\Sigma)} \cdot \frac{\sigma^2 s}{n} \cdot \log(d).$$

Dafür muss laut Vorlesung erfüllt sein: $n \geq c_1 \frac{\|\Sigma\|}{\Lambda_{min}(\Sigma)^2} s \cdot \log(ed/s)$.

(a) Vergleichen Sie die Annahme an n mit der Annahme $n \geq c_1' \cdot \frac{\|\Sigma\|^2}{\lambda_{min}(\Sigma)^2} \cdot d$ des KQ-Schätzers.

(b) Rekapitulieren Sie: Warum ergibt der Beweis den zusätzlichen Faktor $\frac{1}{\Lambda_{min}(\Sigma)}$, der beim KQ-Schätzer nicht vorhanden ist?

Zur Restricted Eigenvalue Property: Es gilt

$$\Lambda_{min}(\Sigma) = \inf_{v \in C} \frac{v^T \Sigma v}{\|v_S\|_2^2}, \qquad C := \{v \in \mathbb{R}^d : \|v_{S^c}\|_1 \leq 3\|v_S\|_1\}.$$

(c) Zeigen Sie: Gilt $\Sigma = I_{d \times d}$, so ist $\Lambda_{min}(\Sigma) \geq 1$.

(d) Diskussion: Warum ist das Szenario aus (a) unrealistisch, insbesondere für große d?

Betrachte nun für $\rho \in (0, 1)$

$$\Sigma = \begin{pmatrix} 1 & 0 & \dots & 0 & \rho \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \rho \\ 0 & \dots & 0 & 1 & \rho \\ \rho & \dots & \rho & \rho & 1 \end{pmatrix}$$

Man kann zeigen: Die Eigenwerte von Σ sind $\{1, ..., 1, 1 - (d-1)^{1/2}\rho, 1 + (d-1)^{1/2}\rho\}$.

(e) Ermitteln Sie eine Bedingung für ρ, so dass noch $\lambda_{min}(\Sigma) > 0$ gilt.

(f) Es sei $S = \{1, ..., s\}$. Ermitteln Sie eine untere Schranke für $\Lambda_{min}(\Sigma)$, die nur von s, nicht von d abhängt. Was ist nun die Bedingung an ρ für $\Lambda_{min}(\Sigma) > 0$?

Aufgabe 2 (Beweis von Satz 2.10.: Untersuchung des Ereignisses B_1).
In Satz 2.10 wurde das Ereignis

$$B_1 := \{\forall j \in \{1,...,d\} : \hat{\Sigma}_{jj} \leq \frac{3}{2}\Sigma_{jj}\}$$

definiert, wobei $\Sigma = \mathbb{E}[XX^T]$, $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n X_i X_i^T$ und $\Sigma_{jj} = 1$ vorausgesetzt wurde.

(a) Zeigen Sie, dass für $t < \frac{1}{2}$ gilt:

$$\mathbb{P}(B_1^c) \leq d \cdot ((1-2t)e^{3t})^{-n/2}.$$

Hinweis: Für $Z \sim N(0,1)$ gilt: $\mathbb{E}e^{tZ^2} = (1-2t)^{-1/2}$.

(b) Es ist $c := \frac{3}{4} + \log(\frac{1}{2}) > 0$. Zeigen Sie, dass für $n \geq \frac{2}{c}(\log(d) + x)$ gilt: $\mathbb{P}(B_1^c) \leq e^{-x}$.

Aufgabe 3 (Verallgemeinerungen von Satz 2.10).
In der Situation von Satz 2.10 untersuchen wir die Wahrscheinlichkeit

$$\mathbb{P}\Big(\frac{2}{n}\|\mathbb{e}^T\mathbb{X}\|_\infty \geq \frac{\lambda}{2}\Big|\mathbb{X}\Big) = \mathbb{P}\Big(\max_{j=1,...,d}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i X_{ij}\Big| > \frac{\lambda\sqrt{n}}{4}\Big|\mathbb{X}\Big)$$

und die damit verbundene Wahl von λ. Der Einfachheit halber nehmen wir hier an, dass $\mathbb{X}$ deterministisch ist und $\frac{1}{n}\sum_{i=1}^n X_{ij}^2 = 1$ erfüllt und ε_i, $i = 1,...,n$ i.i.d. mit $\mathbb{E}[\varepsilon_1^2] = \sigma^2$.

(a) Wiederholen Sie den Beweis der Vorlesung: Zeigen Sie unter der Annahme $\varepsilon \sim N(0, \sigma^2)$, dass

$$\mathbb{P}(\frac{2}{n}\|\mathbb{e}^T\mathbb{X}\|_\infty \geq \frac{\lambda}{2}) \leq 2d\exp\Big(-\frac{1}{2}(\frac{\lambda\sqrt{n}}{4\sigma})^2\Big).$$

(b) Zeigen Sie: Es gilt

$$\mathbb{P}\Big(\frac{2}{n}\|\mathbb{e}^T\mathbb{X}\|_\infty \geq \frac{\lambda}{2}\Big) \leq d \cdot \Big(\frac{4\sigma}{\sqrt{n}\lambda}\Big)^2$$

(c) Sei $p \geq 2$. Zeigen Sie: Gilt $\mathbb{E}[|\varepsilon|^p] = \mu_p^p$ mit einem $\mu_p \in \mathbb{R}$, so gilt

$$\mathbb{P}\Big(\frac{2}{n}\|\mathbb{e}^T\mathbb{X}\|_\infty \geq \frac{\lambda}{2}\Big) \leq d \cdot \Big(\frac{4p^{1/2}\mu_p}{\lambda\sqrt{n}}\Big)^p$$

Hinweis: Für unabhängige Zufallsvariablen A_i, $i = 1,...,n$ gilt die Ungleichung $\mathbb{E}[|\sum_{i=1}^n A_i|^p]^{1/p} \leq p^{1/2}(\sum_{i=1}^n \mathbb{E}[|A_i|^p]^{2/p})^{1/2}$.

Ab jetzt gelte $|X_{ij}| \leq C$ $(i = 1,...,n,\ j = 1,...,d)$ mit einer Konstante $C > 0$.

(d) Es gilt folgende Ungleichung (Nemirovski's Ungleichung, vgl. Lemma 14.24 in Bühlmann/Geer: 'Statistics for high-dimensional data'): Für unabhängige Zufallsvariablen $A_i \in \mathbb{R}^d$, $i = 1,...,n$:

$$\mathbb{E}\big[\max_{j=1,...,d}\big|\sum_{i=1}^n (A_{ij} - \mathbb{E}A_{ij})\big|\big] \quad \leq \quad (8\log(2d))^{1/2} \cdot \mathbb{E}\big[\max_{j=1,...,d}\sum_{i=1}^n A_{ij}^2\big]^{1/2}.$$

Zeigen Sie, dass

$$\mathbb{P}\Big(\frac{2}{n}\|\mathbb{e}^T\mathbb{X}\|_\infty \geq \frac{\lambda}{2}\Big) \leq \frac{4(8\log(2d))^{1/2}C\sigma}{\lambda\sqrt{n}}.$$

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