Mathematik des maschinellen Lernens (WiSe 2021)

1. Übung

02. November 2021

1. Aufgabe$^{(*)}$ (Lineare Regression)

Wir betrachten die klassische lineare Ausgleichsrechnung: Für Messdatenpaare $(\mathbf{x}_i, y_i) \in \mathbb{R}^{d+1}$, $i = 1, \ldots, m$, und Ansatzfunktionen $f_j \colon \mathbb{R}^d \to \mathbb{R}$, $j = 1, \ldots, n$, $n \leq m$, versuchen wir eine Funktion

$$f_{\mathbf{w}}(\mathbf{x}) = \sum_{j=1}^{n} w_j f_j(\mathbf{x}), \qquad \mathbf{w} = (w_1, \ldots, w_n)^\top,$$

an die Messdaten anzupassen durch

$$\frac{1}{m} \sum_{i=1}^{m} \|y_i - f_{\mathbf{w}}(\mathbf{x}_i)\|^2 \to \min_{\mathbf{w} \in \mathbb{R}^n},$$

wobei $\|\cdot\|$ die euklidische Norm bezeichne.

a) Zeigen Sie, dass diese Ausgleichsrechnung als (verzerrte) ERM-Lernregel verstanden werden kann. Geben Sie dazu explizit den Merkmals- und Markierungsraum $\mathcal{X}$ bzw. $\mathcal{Y}$, die Stichprobe S sowie die entsprechende Verlustfunktion ℓ und Hypothesenklasse $\mathcal{H}$ an.

b) Wir führen nun

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m, \qquad F = \begin{pmatrix} f_1(\mathbf{x}_1) & \ldots & f_n(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ f_1(\mathbf{x}_m) & \ldots & f_n(\mathbf{x}_m) \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

ein. Zeigen Sie, dass die die obige Minimierung dann äquivalent ist zu:

$$\|\mathbf{y} - F\mathbf{w}\|^2 \to \min_{\mathbf{w} \in \mathbb{R}^n}.$$

c) Für die lineare Regression können wir die Hypothese h_S, welche das empirische Risiko minimiert exakt berechnen. Nutzen Sie die Formulierung aus Aufgabe b) um zu zeigen, dass h_S gerade denn folgenden Parametervektor nutzt:

$$\mathbf{w}_S := (F^\top F)^{-1} F^\top \mathbf{y}.$$

(*Hinweis:* Mehrdimensionale Kettenregel zum Ableiten nutzen.) Welche Terme auf der rechten Seite hängen alle von der Stichprobe ab?

d) Wir betrachten nun den Spezialfall von Ausgleichsgeraden, d.h.,

$$f_{\mathbf{w}}(\mathbf{x}) = w_1 + w_2 \mathbf{x}.$$

In der Statistik nimmt man oft an, es gibt eine wahre Hypothese, z. B.

$$h^\dagger(\mathbf{x}) = 1 - 2\mathbf{x},$$

und die Daten $(\mathbf{X}, Y)$ folgen der wahren Hypothese plus einem normalverteilten „Messfehler“:

$$Y = h^\dagger(\mathbf{X}) + \epsilon, \qquad \epsilon \sim \mathrm{N}(0, \sigma^2), \quad \sigma > 0,$$

wobei $\mathbf{X}$ und ϵ stochastisch unabhängig seien. Wie groß ist dann das erwartete Risiko $\mathcal{R}_{\mathcal{D}}(h^\dagger)$? Hängt es von der Verteilung von $\mathbf{X}$ ab?

e) Wir nehmen nun an, uns sind folgende 10 Trainingsdaten S gegeben:

x	0.01	0.13	0.37	0.4	0.7	0.77	0.84	0.85	0.87	0.9
y	0.58	1.01	0.15	0.43	−0.23	−0.69	0.23	0.27	−0.96	−0.08

Berechnen Sie die daraus erlernte Hypothese h_S sowie deren empirisches Risiko $\mathcal{R}_S(h_S)$ und das empirische Risiko der „wahren Hypothese" $h^\dagger$. Nutzen Sie dazu das Matlab-Skript `Regress.m`, bei dem Sie noch die Zeilen 45 und 48 ergänzen müssen.

f) Wir wollen nun das erwartete Risiko der erlernten Hypothese berechnen. Zeigen Sie dazu, dass sich das erwartete Risiko wie folgt darstellen lässt:

$$\mathcal{R}_{\mathcal{D}}(h) = \mathbb{E}\left[|h^\dagger(\mathbf{X}) - h(\mathbf{X})|^2\right] + \sigma^2.$$

Nutzen Sie dabei die Unabhängigkeit von $\mathbf{X}$ und ϵ aus sowie, dass $\epsilon \sim \mathrm{N}(0, \sigma^2)$.

g) Wir nehmen nun an, dass $\mathbf{X} \sim \mathrm{U}[0,1]$ und $\sigma^2 = 1$. Berechnen Sie nun das erwartete Risiko $\mathcal{R}_{\mathcal{D}}(h_S)$ der in e) erlernten Hypothese. Runden Sie dazu das Ergebnis für w_S auf zwei Nachkommastellen. Wie hoch ist der Schätzfehler $\varepsilon_{\mathrm{est}}(S)$ in diesem Fall?

h) Angenommen, Sie werten die erlernte Hypothese h_S an 1000 neuen Datenpunkten $(\tilde{x}_1, \tilde{y}_1)$, …, $(\tilde{x}_{1000}, \tilde{y}_{1000})$ aus, die jeweils wieder aus der Verteilung von $(\mathbf{X}, Y) \sim \mathcal{D}$ gezogen seien, und berechnen den resultierenden mittleren quadratischen Fehler

$$\frac{1}{1000}\sum_{j=1}^{1000} |\tilde{y}_j - h_S(\tilde{x}_j)|^2.$$

Wie groß wird dieser in etwa sein? Überprüfen Sie dies numerisch, z. B. in Matlab.

i) Wir wollen an die Daten aus Aufgabe e) nun ein Polynom vom Grad 8 anpassen:

$$f_{\mathbf{w}}(x) = w_1 + w_2 x + w_3 x^2 + w_4 x^3 + w_5 x^4 + w_6 x^5 + w_7 x^6 + w_8 x^7 + w_9 x^8.$$

Erweitern bzw. ergänzen Sie das Matlab-Skript `Regress.m` entsprechend um hier die neun Koeffizienten $w_1, \ldots, w_9$ zu schätzen. Stellen Sie auch die neue erlernte Hypothese im Vergleich zu den Datenpunkten grafisch dar.

Berechnen Sie für das Ausgleichspolynom erneut das empirische Risiko und schätzen Sie das zu erwartende Risiko wie in Aufgabe h). Was stellen Sie im Vergleich zu den Werten der Ausgleichsgerade in Aufgabe e) und h) fest?