Universität Salzburg | Florian Graf

Machine Learning

Übungsblatt **5** | 24 Punkte

Aufgabe 1. *Lineare Regression – MLE für σ^2* 6 P.

Es seien Einflussgrößen $(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^d$ und Zielgrößen $(y_1, \dots, y_n) \in \mathbb{R}$ gegeben. Wir modellieren die Zielgrößen durch ein lineares Regressionsmodell der Form $p(y_i|\mathbf{x}_i, \theta) = \mathcal{N}(y_i|\mathbf{w}^\top \mathbf{x}_i, \sigma^2)$.
In der Vorlesung wurde gezeigt, dass der Maximum-Likelihood Schätzer für $\mathbf{w}$ durch $\hat{\mathbf{w}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$ gegeben ist. Wie schätzen nun den Varianzparameter σ^2.

(a) Geben Sie die NLL (negative log likelihood) Funktion unter den genannten Verteilungsannahmen an. Nutzen Sie dafür die Datenmatrix $X := \left(\mathbf{x}_1, \dots, \mathbf{x}_n\right)^\top$ und den Zielvektor $\mathbf{y} = \left(y_1, \dots, y_n\right)^\top$.

(b) Bestimmen Sie die Nullstelle $\hat{\sigma}^2$ der partiellen Ableitung von $\mathrm{NLL}(\mathbf{w}, \sigma^2)$ nach σ^2.

(c) Bestimmen Sie die zweite partielle Ableitung von $\mathrm{NLL}(\mathbf{w}, \sigma^2)$ nach σ^2 und zeigen Sie, dass diese an der Stelle $\hat{\sigma}^2$ positiv ist.

(d) Genügen (b) und (c) um zu zeigen, dass $(\hat{\mathbf{w}}, \hat{\sigma}^2)$ ein lokales Maximum der Likelihoodfunktion ist? Begründen Sie Ihre Antwort.

Aufgabe 2. *Lineare Regression – Geometrische Interpretation* 6 P.

In der Vorlesung wurde gezeigt, dass unter den Annahmen von Aufgabe 1 der Maximum-Likelihood Schätzer für $\mathbf{w}$ durch $\hat{\mathbf{w}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$ gegeben ist.
Die Vorhersage der Zielvariable ist demnach $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{w}} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \mathbf{P}\mathbf{y}$. Zeigen Sie, dass $\mathbf{P} := \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ eine orthogonale Projektion auf den Spaltenraum von $\mathbf{X}$ definiert, indem Sie zeigen, dass

(a) $\mathbf{PP} = \mathbf{P}$

(b) Für jeden Vektor $\mathbf{v} \in \mathbb{R}^n$ gilt $\mathbf{Pv} \in \mathrm{span}(\{\mathbf{x}_{:,i} : i = 1, \dots, d\})$. D.h., zeigen Sie, dass es Faktoren $a_1, \dots, a_d \in \mathbb{R}$ gibt, sodass $\mathbf{Pv} = \sum_{i=1}^d a_i \mathbf{x}_{:,i}$.

(c) Für jeden Spaltenvektor $\mathbf{x}_{:,i}$ von X und für alle $\mathbf{v} \in \mathbb{R}^n$ ist $\mathbf{v} - \mathbf{Pv}$ orthogonal zu $\mathbf{x}_{:,i}$.

Aufgabe 3. *Lineare Regression – Beispiel 1* 6 P.

Ein Gleichstrom-Motor erbrachte bei $n = 8$ Messungen die folgenden Werte für die Leistung P [kW] in Abhängigkeit von der Drehzahl f [1000 U/min]:

f_k	0.8	1.5	2.5	3.5	4.2	4.7	5.0	5.5
P_k	8.8	14.7	22.8	29.4	38.2	44.1	47.8	51.5

(a) Zeichnen Sie die Messwerte (sauber!) in ein Koordinatensystem ein.

(b) In dem für den vorliegenden Fall von skalaren Einflussgrößen $x_i \in \mathbb{R}$ reduziert sich das Regressionsmodell zu $p(y_i|x_i, \theta) = \mathcal{N}(y_i|ax + b)$. Leiten Sie die Maximum-Likelihood Schätzer für a und b aus der Formel $\hat{\mathbf{w}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$ für den Fall von $d = 1$ her.

(c) Berechnen Sie mit diesen Formeln (und einem Taschenrechner) die Regressionsgerade und zeichnen Sie diese (mit Lineal!) ein.

(d) Bestimmen Sie die Residuenquadratsumme (RSS) des Regressionsmodells.

(e) Bestimmen Sie aus der Regressionsgerade einen Schätzwert für die Motorleistung bei 4000 U/min und zeichnen Sie ihn ein.

Aufgabe 4. *Lineare Regression – Beispiel 2* 6 P.

Betrachten Sie das folgende lineare Regressionsmodell mit einer 2-dimensionalen Zielvariable $y_i \in \mathbb{R}^2$ und einer binären Einflussgröße $x_i \in \{0, 1\}$.

Die Trainingsdaten sind wie in der Tabelle rechts angegeben. Zusätzlich, betten wir die Einflussgrößen folgendermaßen in $\mathbb{R}^2$ ein:

$$\phi(0) = (1, 0)^\top, \qquad \phi(1) = (0, 1)^\top \ ,$$

sodass $\mathbf{y} = \mathbf{W}^\top \phi(\mathbf{x})$ wobei $\mathbf{W}$ eine 2×2 Matrix ist.

Bestimmen Sie den Maximum-Likelihood Schätzer für $\mathbf{W}$ mithilfe der Formel

$$\hat{\mathbf{W}} = \left(\boldsymbol{\phi}(\mathbf{X})^\top \boldsymbol{\phi}(\mathbf{X})\right)^{-1} \boldsymbol{\phi}(\mathbf{X})^\top \mathbf{Y} \ .$$

x	y
0	$(-1, -1)^\top$
0	$(-1, -2)^\top$
0	$(-2, 1)^\top$
1	$(1, 1)^\top$
1	$(1, 2)^\top$
1	$(2, 1)^\top$