Machine Learning

Assignment 02: Matrix Calculus Review

Übung 01

$$x = (x_1 \quad x_2 \quad \cdots \quad x_n)^T \qquad\qquad A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

$$f(x) = \begin{pmatrix} x_1 a_{11} + x_2 a_{21} + \cdots + x_n a_{m1} \\ x_1 a_{12} + \cdots + x_n a_{m2} \\ \vdots \\ x_1 a_{1n} + \cdots + x_n a_{mn} \end{pmatrix}$$

$$Df(x) = \begin{pmatrix} \frac{\vartheta f(x_1)}{\vartheta x_1} = a_{11} & \frac{\vartheta f(x_1)}{\vartheta x_2} = a_{21} & \cdots & a_{m1} \\ \vdots & \cdots & \cdots & \vdots \\ a_{1n} & \cdots & \cdots & a_{mn} \end{pmatrix} = A^T$$

Übung 02

$$x = (x_1 \quad x_2 \quad \cdots \quad x_n) \in R^n \qquad g(x) = (x^T x)^{m+1} \qquad mit\ m \in N, m > 0$$

Aus $x^T x = \sum_i x_i^2$ *folgt:* $(x^T x)^{m+1} = (\sum_i x_i^2)^{m+1}$

1. *Schritt: eine Variable ableiten:*
$$\frac{d}{dx_1}(\,x^T x)^{m+1} = \frac{d}{dx_1}\left(\left(\sum_i x_i^2\right)^{m+1}\right) = \frac{d}{dx_1}(x_1^{2m+2}) = (2m+2)x_1^{2m+1}$$
2. *Schritt: auf den ganzen Vektor anwenden:*
$$\frac{d}{dx}(\,x^T x)^{m+1} = \left(\frac{d}{dx_1}(x_1^{2m+2}) \quad \cdots \quad \frac{d}{dx_n}(x_n^{2m+2})\right)$$
$$= ((2m+2)x_1^{2m+1} \quad \cdots \quad (2m+2)x_n^{2m+1}) = (2m+2)(x^T)^{2m+1}$$

Übung 03

<u>Zu zeigen:</u> *Wann ist* $tr(AA) = tr(A^T A)$*?*

$$tr(A^2) = \sum_{i=1}^{n}\sum_{j=2}^{n} a_{ij}a_{ji} \stackrel{?}{=} \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 = tr(A^T A)$$

Dies ist genau dann der Fall, wenn $a_{ij} = a_{ji}\ \forall i,j$ *ist, also bei symmetrischen Matrizen.*

<u>Kann es auch eine andere Lösung geben?</u>

Da $\sum_{i=1}^{n}\sum_{j=2}^{n} a_{ij}a_{ji} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2$ sein muss, müssen wir nun eine Möglichkeit finden, die Zahlen so anzuordnen, sodass wiederum das gleiche Ergebnis herauskommt, dabei dürfen die Zahlen aber nur so oft verwendet werden, wie sie auch vorkommen; Beispielsweise müssen wir die Zahlen 1, 2, 3 in die Gleichung $1a + 1b + 1c = 1^2 + 2^2 + 3^2$ so einordnen, dass es auf beiden Seiten dasselbe ergibt. Da aber die Summe der Quadrate immer das Maximum, welches auch nur auf diese Art und Weise erreicht werden kann, bildet, ist dies nicht möglich.

Somit besteht die Lösung für $tr(AA) = tr(A^T A)$ nur aus symmetrischen Matrizen.

Machine Learning

Assignment 02: Matrix Calculus Review

Übung 04

(i) $\nabla_x tr(AX^T) = A?$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \qquad X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix}$$

$$AX^T = \begin{pmatrix} \sum_{i=1}^{n} a_{1i}x_{1i} & \cdots & . \\ \vdots & \ddots & \vdots \\ . & \cdots & \sum_{i=1}^{n} a_{mi}x_{mi} \end{pmatrix} => tr(AX^T) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_{ij}$$

1. Schritt: eine Variable ableiten:

$\frac{d}{dx_{ij}}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_{ij} = a_{ij}$ *, da alle anderen Einträge bezogen auf x_{ij} Konstanten darstellen*

2.Schritt: auf die ganze Matrix anwenden:

$$\frac{d}{dx}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_{ij} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = A$$ *qed.*

(ii) $\nabla_x tr(X^T X) = 2X$

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix}$$

$$tr(X^T X) = \sum_{i=1}^{m}\sum_{j=1}^{n} {x_{ij}}^2$$

1. Schritt: eine Variable ableiten:

$\frac{d}{dx_{ij}}\sum_{i=1}^{m}\sum_{j=1}^{n} {x_{ij}}^2 = 2x_{ij}$ *, da alle anderen Einträge bezogen auf x_{ij} Konstanten darstellen*

2.Schritt: auf die ganze Matrix anwenden:

$$\frac{d}{dx}\sum_{i=1}^{m}\sum_{j=1}^{n} {x_{ij}}^2 = \begin{pmatrix} 2x_{11} & \cdots & 2x_{1n} \\ \vdots & \ddots & \vdots \\ 2x_{m1} & \cdots & 2x_{mn} \end{pmatrix} = 2\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix} = 2X$$ *qed.*

Übung 05

$A \in R^{m*n}$ voller Rang (alle Spalten linear unabhängig)
$b \in R^m$ $b \notin R(A)$
$x \in R^n$

<u>*Wir wissen damit:*</u> $(a_1 \quad \cdots \quad a_n) * \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = a_1x_1 + a_2x_2 + \cdots + a_nx_n \neq b$

$\exists x^*, mit\ Ax^* = v\ so\ nah\ wie\ möglich\ an\ b$ *=> minimiere* $||b - v|| = ||b - Ax^*||$

$$\left\|\begin{pmatrix} b_1 - v_1 \\ \vdots \\ b_n - v_n \end{pmatrix}\right\| = (b_1 - v_1)^2 + (b_2 - v_2)^2 + \cdots + (b_n - v_n)^2$$

Da $Ax^* - b \in R(A)^\perp$ *=>* $A^T(Ax^* - b) = A^TAx^* - A^Tb = 0$ *was genau eine Lösung hat qed.*

Machine Learning

Assignment 02: Matrix Calculus Review

Übung 06

$$f(x) = x^T Ax \rightarrow Maximum, mit\ x^T x = 1$$

Da A symmetrisch ist, ist sie diagonalisierbar in eine Orthonormalbasis von R^n, also:

$$x = \sum_i x_i e_i \quad und\ A * e_i = \lambda_i e_i$$

Daraus folgt:
$x^T Ax = x^T A \sum_i x_i e_i = x^T \sum_i x_i \lambda_i e_i$ *(Aufgrund der Linearität und der Definition der Eigenvektoren)*

Und da die Basis orthonormal ist:

$$x^T \sum_i x_i \lambda_i e_i = \sum_i \lambda_i {x_i}^2 = f(x)$$

Nun definieren wir: $max(\lambda_i) = \lambda_M$.
Da wir durch einsetzen von e_M in f λ_M erhalten haben wir damit bewiesen, dass f das Maximum erreicht.