University of Oslo
Department of Mathematics

Fall, 2023
Weekly Exercises: Week 35

Author: Jonas Semprini Næss

FYS-STK4155: Applied Data Analysis and Machine Learning

Exercise 1: Analytical Exercises

a.) *Show that*

$$\frac{\partial(\mathbf{b}^T\mathbf{a})}{\partial\mathbf{a}} = \mathbf{b}.$$

Solution:

By convention we assume that both $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{n\times 1}$. Hence can we re-write the given expression as

$$\frac{\partial\left(\sum_i b_i a_i\right)}{\partial a_k}$$

[1] $\forall i, k = 0, 1, 2, \ldots, n-1$, which yields

$$\frac{\partial\left(\sum_i b_i a_i\right)}{\partial \mathbf{a_k}} = b_k$$

$$\Downarrow$$

$$\frac{\partial(\mathbf{b}^T\mathbf{a})}{\partial\mathbf{a}} = \begin{bmatrix} b_0 \\ b_2 \\ \vdots \\ b_{n-1} \end{bmatrix} = \mathbf{b}$$

$\blacksquare$.

b.) *Show that*

$$\frac{\partial(\mathbf{a}^T\mathbf{A}\mathbf{a})}{\partial\mathbf{a}} = \mathbf{a}^T(\mathbf{A} + \mathbf{A}).$$

Let $\mathbf{W} = \mathbf{A}\mathbf{a}$ such that the original quadratic form now becomes $\mathbf{a}^T\mathbf{W}$. By this we can initiate the product rule for derivatives. Thus

$$\frac{\partial\mathbf{a}^T}{\partial\mathbf{a}}\mathbf{W} + \mathbf{a}^T\frac{\partial\mathbf{W}}{\partial\mathbf{a}}$$

[1] I have decided to adapt standard typing and not boldface for specifying when we are working with scalar elements of respective vectors or matricies

which by applying (a.) gives that

$$\frac{\partial \mathbf{a}^T}{\partial \mathbf{a}}\mathbf{W} + \mathbf{a}^T\frac{\partial \mathbf{W}}{\partial \mathbf{a}} = \mathbf{W}^T + \mathbf{a}^T\mathbf{A}.$$

Finally exploit the mathematical fact of $\boxed{(AB)^T = B^T A^T}$ yielding

$$\begin{aligned}\mathbf{W}^T + \mathbf{a}^T\mathbf{A} &= \mathbf{a}^T\mathbf{A}^T + \mathbf{a}^T\mathbf{A}\\ &= \mathbf{a}^T(\mathbf{A} + \mathbf{A}^T)\end{aligned}$$

which is what we wanted to show. ■

c.) *Show that*

$$\frac{\partial\left(\mathbf{x} - \mathbf{A}\mathbf{s}\right)^T\left(\mathbf{x} - \mathbf{A}\mathbf{s}\right)}{\partial \mathbf{s}} = -2\left(\mathbf{x} - \mathbf{A}\mathbf{s}\right)^T\mathbf{A}.$$

We follow the same reasoning as in (b.) which gives us the following

$$\frac{\partial\left(\mathbf{x} - \mathbf{A}\mathbf{s}\right)^T}{\partial \mathbf{s}}\left(\mathbf{x} - \mathbf{A}\mathbf{s}\right) + \left(\mathbf{x} - \mathbf{A}\mathbf{s}\right)^T\frac{\partial\left(\mathbf{x} - \mathbf{A}\mathbf{s}\right)}{\partial \mathbf{s}}$$

. Split up the parentheses such that

$$\frac{\partial \mathbf{x}^T}{\partial \mathbf{s}}\left(\mathbf{x} - \mathbf{A}\mathbf{s}\right) - \left(\frac{\partial\left(\mathbf{A}\mathbf{s}\right)^T}{\partial \mathbf{s}}\right)\left(\mathbf{x} - \mathbf{A}\mathbf{s}\right) + \left(\mathbf{x} - \mathbf{A}\mathbf{s}\right)^T\frac{\partial \mathbf{x}}{\partial \mathbf{s}} - \left(\mathbf{x} - \mathbf{A}\mathbf{s}\right)^T\left(\frac{\partial\left(\mathbf{A}\mathbf{s}\right)}{\partial \mathbf{s}}\right).$$

Then differentiate with respect to $\mathbf{s}$, yielding

$$-\mathbf{A}(\mathbf{x} - \mathbf{A}\mathbf{s})^T - (\mathbf{x} - \mathbf{A}\mathbf{s})^T\mathbf{A} = -2(\mathbf{x} - \mathbf{A}\mathbf{s})^T\mathbf{A}.$$

Which is what we wanted to show. ■

d.)

To analyse the second derivate of the above mentioned result, we first differentiate $-2(\mathbf{x} - \mathbf{A}\mathbf{s})^T\mathbf{A}$ with respect to $\mathbf{s}$ again. This gives:

$$\begin{aligned}-2\frac{\partial(\mathbf{x} - \mathbf{A}\mathbf{s})^T\mathbf{A}}{\partial \mathbf{s}} &= 2\frac{(\mathbf{A}\mathbf{s})^T}{\partial \mathbf{s}}\mathbf{A}\\ &= 2\mathbf{A}^T\mathbf{A}\end{aligned}$$

Which describes the Hessian Matrix. If $\mathbf{A}$ happens to be symmetric then obviously we would have that $-2\frac{\partial(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{A}}{\partial \mathbf{s}} = 2\mathbf{A}^2$, and if orthogonal then $-2\frac{\partial(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{A}}{\partial \mathbf{s}} = 2$. ■