MASD 2020, Assignment 2

Hand-in in groups of 2 or 3 before 15.9.2020 at 10:00
One submission per group
Remember to include the names of all group memebers

Exercise 1 (Writing proofs: Peer feedback). In this exercise, you will read proofs for Assignment 1, Exercise 1 written by your fellow students and give feedback on these by answering the following 4 questions:

i) If the proof rests on assumptions, are they clearly stated?
ii) Does the proof contain a clearly stated claim?
iii) Does the proof contain a clearly explained logical argument for how the assumptions lead to the claim?
iv) Are the assumptions, claims and arguments correct? All feedback on the correctness goes here.

Your feedback should clearly answer all four questions. In answering questions i)-iii), you should focus on *clarity*: Are all details clearly presented.

Remember to be respectful when giving feedback to your fellow students – if you believe the proof has problems, say so, but keep it factual and respectful.

Deliverables. Feedback uploaded in Peer feedback.
Solution:
No standard solution here...

Exercise 2 (Concept of derivative).

a) Assume that $f\colon \mathbb{R} \to \mathbb{R}$ is a differentiable function (the notation means that f is a function that takes real numbers as input and returns real numbers as output).
Do the following expressions (in this exact form) give $\frac{df}{dt}$ at time t? Please give an argument for your answers.
   i) $\frac{f(t+\Delta t)-f(t)}{\Delta t}$
   ii) $\lim_{h\to 0} \frac{f(t+5h)-f(t)}{5h}$
   iii) $\lim_{\Delta t\to 0} \frac{f(t-\Delta t)-f(t)}{-\Delta t}$

b) Draw the "shape" of the derivative of the functions shown in Fig. 1. The important points are to get the signs and the zeros of the derivative roughly right.

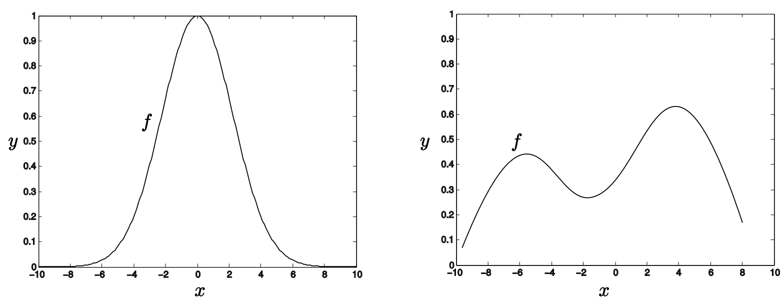

Figure 1: Draw the shapes of the derivative of these functions.

c) For each of the cases below, draw the graph of a differentiable function $f\colon \mathbb{R} \to \mathbb{R}$ for which the following holds:

i) $f'(-1) = f'(1) = 0$, $f''(-1) > 0$, $f''(1) < 0$
ii) $f'(-1) = 2$, $f'(3) = -2$, $f'(0) = 0$, $f''(0) > 0$

Deliverables. a) An answer and a one-liner explanation per item; b) and c) the drawings (photos of hand-drawn sketches are OK, and millimeter precision is not necessary).
Solution:

a) i) The fraction $\frac{f(t+\Delta t)-f(t)}{\Delta t}$ is a finite difference *approximation* of a derivative, but it is not the derivative. This is because $\Delta t > 0$ is a fixed number.

ii) Use the change of variable $u = 5h$ and note that

$$\lim_{u\to 0} \frac{f(t+u)-f(t)}{u} = \lim_{5h\to 0} \frac{f(t+5h)-f(t)}{5h} = \lim_{h\to 0} \frac{f(t+5h)-f(t)}{5h},$$

where the last equation holds because $h \to 0 \;\Leftrightarrow\; 5h \to 0$. But the left hand side is the definition of the derivative. Hence, this is a derivative.

iii) Use the change of variable $h = -\Delta t$; then we see that

$$\lim_{h\to 0} \frac{f(t+h)-f(t)}{h} = \lim_{-\Delta t\to 0} \frac{f(t-\Delta t)-f(t)}{-\Delta t} = \lim_{\Delta t\to 0} \frac{f(t-\Delta t)-f(t)}{-\Delta t}$$

because $\Delta t \to 0 \;\Leftrightarrow\; -\Delta t \to 0$. Hence, this is a derivative.

b) See Figure ??.

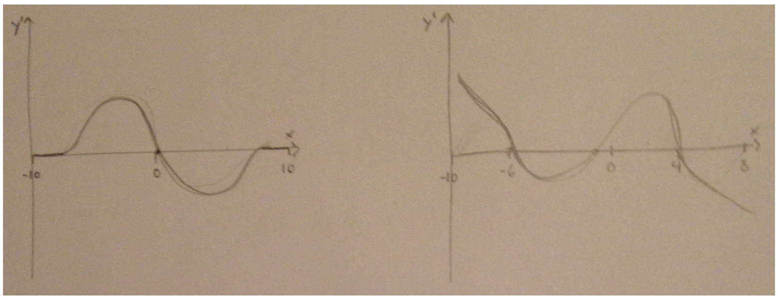

Figure 2: Drawings of the shapes of the derivatives.

c) See Figure ??.

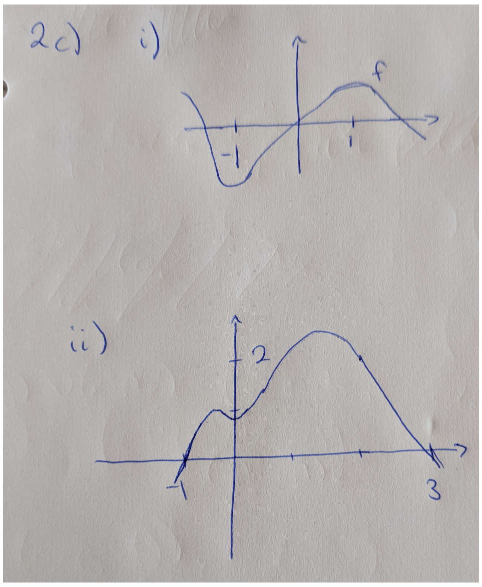

Figure 3: Drawings of i) and ii).

Exercise 3 (Techniques for differentiation). Derive the following derivatives. That is, write out the steps in finding the derivatives, and explain them. While careful typesetting is nice, it is not necessary – it is OK to write explanations of the form "at the first equality we used the sum rule, and in the second we used the quotient rule". You do not need to write proofs with "assumptions" and "claims", but clarity and correctness will be evaluated as with proofs.

a) $\frac{d}{dx}\left(x^3 + e^{2x}\right)$
b) $\frac{d}{dx}\left(e^{x^2+3x^3}\right)$
c) $\frac{d}{dx}\left(\frac{\ln x}{x^2}\right)$
d) $\frac{\partial}{\partial x}\left(e^{x^2+3xy+2y^3}\right)$
e) $\frac{\partial}{\partial y}\left(e^{xy}\ln(x^2+y^3)\right)$
f) $\frac{\partial}{\partial x_i}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)$, where $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$. Hint: Express $\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)$ as a double summation. If still in doubt, try to write it out for $n = 2$.
g) $\nabla_{\mathbf{x}}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)$, where $\mathbf{x}$ and $\mathbf{A}$ are as in f) – ideally expressed in matrix form (that is, in terms of matrices and vectors, not coordinates). Hint: Let $a_{(i,.)}$ denote the i-th row of $\mathbf{A}$, $i = 1, 2 \dots n$,and let $a_{(.,j)}$ denote the j-th column of $\mathbf{A}$. Start by expressingthe result from 3f using this column and row notation.

Deliverables. Derivations.

Solution:

Throughout the exercise, I will be using the following derivative rules:

$$
\begin{aligned}
f(x) = c \Rightarrow f'(x) &= 0 && \text{Constant function} \\
(f(x) + g(x))' &= f'(x) + g'(x) && \text{Sum rule} \\
f(g(x))' &= f'(g(x))g'(x) && \text{Chain rule} \\
(f(x)g(x))' &= f'(x)g(x) + f(x)g'(x) && \text{Product rule} \\
\left(\frac{f(x)}{g(x)}\right)' &= \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2} && \text{Quotient rule}
\end{aligned}
$$

a)

$$
\begin{aligned}
\frac{d}{dx}\left(x^3 + \exp(2x)\right) &= \frac{d}{dx}x^3 + \frac{d}{dx}\exp(2x) && \text{Sum rule} \\
&= 3x^2 + \frac{d}{dx}(2x)\exp(2x) && \text{Chain rule} \\
&= 3x^2 + 2\exp(2x)
\end{aligned}
$$

b)

$$
\begin{aligned}
\frac{d}{dx}\left(\exp(x^2 + 3x^3)\right) &= \exp(x^2 + 3x^3)\frac{d}{dx}(x^2 + 3x^3) && \text{Chain rule} \\
&= \exp(x^2 + 3x^3)\left(\frac{d}{dx}x^2 + \frac{d}{dx}3x^3\right) && \text{Sum rule} \\
&= \exp(x^2 + 3x^3)(2x + 9x^2)
\end{aligned}
$$

c)

$$
\begin{aligned}
\frac{d}{dx}\frac{\ln x}{x^2} &= \frac{x^2\frac{1}{x} - \ln x(2x)}{x^4} && \text{Quotient rule} \\
&= \frac{1}{x^3} - 2\frac{\ln x}{x^3}
\end{aligned}
$$

d)

$$\begin{aligned}
\frac{\partial}{\partial x}\exp(x^2+3xy+2y^3) &= \exp(x^2+3xy+2y^3)\frac{\partial}{\partial x}\left(x^2+3xy+2y^3\right) && \text{Chain rule}\\
&= \exp(x^2+3xy+2y^3)(2x+3y)
\end{aligned}$$

e)

$$\begin{aligned}
\frac{\partial}{\partial y}\left(e^{xy}\ln(x^2+y^3)\right) &= e^{xy}\frac{\partial}{\partial y}\ln(x^2+y^3)+\ln(x^2+y^3)\frac{\partial}{\partial y}e^{xy} && \text{Product rule}\\
&= e^{xy}\frac{1}{x^2+y^3}\frac{\partial}{\partial y}(x^2+y^3)+\ln(x^2+y^3)e^{xy}\frac{\partial}{\partial y}(xy) && \text{2 x Chain rule}\\
&= \frac{3y^2e^{xy}}{x^2+y^3}+\ln(x^2+y^3)e^{xy}x.
\end{aligned}$$

f) First, we note that

$$\begin{aligned}
\mathbf{x}^T\mathbf{A}\mathbf{x} &= \mathbf{x}^T\left(\begin{array}{c} \sum_{k=1}^n a_{1k}x_k \dots \sum_{k=1}^n a_{nk}x_k \end{array}\right)^T\\
&= \sum_{l=1}^n x_l \sum_{k=1}^n a_{lk}x_k\\
&= \sum_{l=1}^n\sum_{k=1}^n x_l a_{lk} x_k,
\end{aligned}$$

from which we compute

$$\frac{\partial}{\partial x_i}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right) = \sum_l\sum_k \frac{\partial}{\partial x_i}\left(x_l a_{lk} x_k\right) = \sum_{l\neq i} x_l a_{li} + \sum_{k\neq i} a_{ik}x_k + \underbrace{2a_{ii}x_i}_{=a_{ii}x_i+x_ia_{ii}} = \sum_{l=1}^n x_l a_{li} + \sum_{k=1}^n a_{ik}x_k.$$

It would be fine to stop here, but for the sake of the following exercise, let's try to move this into something that resembles matrix form. Note that

$$\sum_{l=1}^n x_l a_{li} = (x_1,\dots,x_n)\left(\begin{array}{c} a_{1i}\\ \vdots \\ a_{ni}\end{array}\right) = \mathbf{x}^T\mathbf{a}_{(\cdot,i)},$$

where $\mathbf{a}_{(\cdot,i)} = \left(\begin{array}{c} a_{1i}\\ \vdots \\ a_{ni}\end{array}\right)$ is the i^{th} column of $\mathbf{A}$. Similarly,

$$\sum_{k=1}^n a_{ik}x_k = (a_{i1},\dots,a_{in})\left(\begin{array}{c} x_1\\ \vdots \\ x_n\end{array}\right) = \mathbf{a}_{(i,\cdot)}^T\mathbf{x},$$

where

$$\mathbf{a}_{(i,\cdot)} = \left(\begin{array}{c} a_{i1}\\ \dots \\ a_{in}\end{array}\right)$$

is the (column vector corresponding to) the i^{th} row of $\mathbf{A}$. As a result, we get

$$\frac{\partial}{\partial x_i}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right) = \mathbf{x}^T\mathbf{a}_{(\cdot,i)} + \mathbf{a}_{(i,\cdot)}^T\mathbf{x} = \mathbf{a}_{(\cdot,i)}^T\mathbf{x} + \mathbf{a}_{(i,\cdot)}^T\mathbf{x},$$

where the last equality holds because for any two vectors $\mathbf{u}$ and $\mathbf{v}$, we have

$$\mathbf{u}^T\mathbf{v} = u_1v_1 + u_2v_2 + \ldots + u_nv_n = v_1u_1 + v_2u_2 + \ldots + v_nu_n = \mathbf{v}^T\mathbf{u}.$$

g) Since

$$\nabla_{\mathbf{x}} \left(\mathbf{x}^T \mathbf{A}\mathbf{x}\right) = \left(\frac{\partial}{\partial x_1}\left(\mathbf{x}^T \mathbf{A}\mathbf{x}\right), \ldots, \frac{\partial}{\partial x_n}\left(\mathbf{x}^T \mathbf{A}\mathbf{x}\right)\right),$$

and

$$\frac{\partial}{\partial x_i}\left(\mathbf{x}^T \mathbf{A}\mathbf{x}\right) = \mathbf{x}^T \mathbf{a}_{(\cdot,i)} + \mathbf{a}_{(i,\cdot)}^T \mathbf{x} = \mathbf{a}_{(\cdot,i)}^T \mathbf{x} + \mathbf{a}_{(i,\cdot)}^T \mathbf{x},$$

Recall that $\mathbf{a}_{(\cdot,i)}$ is the i^{th} column of $\mathbf{A}$. Moreover, since $\mathbf{a}_{(i,\cdot)}$ is the (column vector corresponding to) the i^{th} row of $\mathbf{A}$, it corresponds to the i^{th} column of $\mathbf{A}^T$. Aggregating the terms, we obtain (on matrix form):

$$\nabla_{\mathbf{x}} \left(\mathbf{x}^T \mathbf{A}\mathbf{x}\right) = (\mathbf{A}^T + \mathbf{A})\mathbf{x}.$$

Exercise 4 (Gradients for data fitting). . The file `A2-exercise4.ipynb` contains data on chocolate consumption (in kg per person per year) and nobel prize laureates (per 100,000 inhabitants) for 20 countries. The data can be described as follows: $(C_1, N_1), ..., (C_{20}, N_{20})$, where C_i and N_i denote chocolate consumption and nobel prize laureates for country i. We now describe the data using the model

$$N_i = a * C_i + b, i = 1, 2, ..., 20$$

To do so, for each value of a and b, we define a badness function

$$badness : \begin{array}{lcc} \mathbb{R} \times \mathbb{R} & \to & \mathbb{R} \\ (a, b) & \to & \frac{1}{20}\sum_{i=1}^{20}(N_i - aC_i - b)^2 \end{array}$$

that indicates how badly our model fits the data. The task is now to find values for a and b that minimize badness.

a) Implement *badness* as the function `def badness(a, b, cho, nob)` in the python notebook.

b) Compute the gradient $\nabla_{(a,b)} badness$. Implement this as the function `def badnessgradient(a, b, cho, nob)` in the python notebook.

c) Start with the values $a = 1$ and $b = 0$, compute the value of *badness* and its gradient and plot the model fit (using `plotdatafit`). Use the information in the gradient to find values of a and b that yield a *badness* of less than 50. Report the three values of a, b and *badness*. (Hint: If you move a tiny step into the *negative* direction of the gradient, i.e.,

$$(a', b') = (a, b) - \lambda \nabla_{(a,b)} badness$$

for a small $\lambda > 0$, you should be able to decrease badness. From there you again compute the gradient, move a step, and so on.) Optional: what is the smallest badness you can achieve?

Deliverables. a) code (with suitable comments) in the ipynb- and pdf-file b) formula; code (both uploaded and in the report) c) values for a, b and *badness*.

Solution:

See Jupyter notebook.