University of Cyprus
MAI613 - Research Methodologies and Professional Practices in AI
Lecturer: Stelios Timotheou

Exercises on Vector Calculus

Exercises

Solve Exercises 5.1, 5.3, 5.4, 5.7, 5.8, and 5.9 from the book:

- M.P. Deisenroth, A. A. Faisal, and C. S. Ong, "Mathematics for Machine Learning," Cambridge University Press, 2020.
  https://mml-book.github.io/book/mml-book.pdf

Note: These exercises are intended for self-assessment purposes. Their solutions will be uploaded in one week from now.

5

Vector Calculus

Exercises

5.1 Compute the derivative $f'(x)$ for

$$f(x) = \log(x^4)\sin(x^3)\,.$$

$$f'(x) = \frac{4}{x}\sin(x^3) + 12x^2\log(x)\cos(x^3)$$

5.3 Compute the derivative $f'(x)$ of the function

$$f(x) = \exp(-\tfrac{1}{2\sigma^2}(x-\mu)^2)\,,$$

where $\mu,\ \sigma \in \mathbb{R}$ are constants.

$$f'(x) = -\tfrac{1}{\sigma^2} f(x)(x-\mu)$$

5.4 Compute the Taylor polynomials T_n, $n = 0, \ldots, 5$ of $f(x) = \sin(x) + \cos(x)$ at $x_0 = 0$.

$$
\begin{aligned}
T_0(x) &= 1\\
T_1(x) &= T_0(x) + x\\
T_2(x) &= T_1(x) - \frac{x^2}{2}\\
T_3(x) &= T_2(x) - \frac{x^3}{6}\\
T_4(x) &= T_3(x) + \frac{x^4}{24}\\
T_5(x) &= T_4(x) + \frac{x^5}{120}
\end{aligned}
$$

5.7 Compute the derivatives $\mathrm{d}f/\mathrm{d}\boldsymbol{x}$ of the following functions by using the chain rule. Provide the dimensions of every single partial derivative. Describe your steps in detail.

a.

$$f(z) = \log(1+z)\,, \quad z = \boldsymbol{x}^\top \boldsymbol{x}\,, \quad \boldsymbol{x} \in \mathbb{R}^D$$

b.

$$f(\boldsymbol{z}) = \sin(\boldsymbol{z})\,, \quad \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\,, \quad \boldsymbol{A} \in \mathbb{R}^{E\times D}, \boldsymbol{x} \in \mathbb{R}^D, \boldsymbol{b} \in \mathbb{R}^E$$

where $\sin(\cdot)$ is applied to every element of $\boldsymbol{z}$.

a.

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \underbrace{\frac{\partial f}{\partial z}}_{\in\mathbb{R}} \underbrace{\frac{\partial z}{\partial \boldsymbol{x}}}_{\in\mathbb{R}^{1\times D}} \in \mathbb{R}^{1\times D}$$

$$\frac{\partial f}{\partial z} = \frac{1}{1+z} = \frac{1}{1+\boldsymbol{x}^\top\boldsymbol{x}}$$

$$\frac{\partial z}{\partial \boldsymbol{x}} = 2\boldsymbol{x}^\top$$

$$\implies \frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \frac{2\boldsymbol{x}^\top}{1+\boldsymbol{x}^\top\boldsymbol{x}}$$

b.

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \underbrace{\frac{\partial f}{\partial \boldsymbol{z}}}_{\in\mathbb{R}^{E\times E}} \underbrace{\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}}_{\in\mathbb{R}^{E\times D}} \in \mathbb{R}^{E\times D}$$

$$\sin(\boldsymbol{z}) = \begin{bmatrix} \sin z_1 \\ \vdots \\ \sin z_E \end{bmatrix}$$

$$\frac{\partial \sin \boldsymbol{z}}{\partial z_i} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \cos(z_i) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^E$$

$$\implies \frac{\partial f}{\partial \boldsymbol{z}} = \mathrm{diag}(\cos(\boldsymbol{z})) \in \mathbb{R}^{E\times E}$$

$$\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \boldsymbol{A} \in \mathbb{R}^{E\times D}:$$

$$c_i = \sum_{j=1}^{D} A_{ij}x_j \implies \frac{\partial c_i}{\partial x_j} = A_{ij}\,, \quad i = 1,\dots,E, j = 1,\dots,D$$

Here, we defined c_i to be the ith component of $\boldsymbol{Ax}$. The offset $\boldsymbol{b}$ is constant and vanishes when taking the gradient with respect to $\boldsymbol{x}$. Overall, we obtain

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \mathrm{diag}(\cos(\boldsymbol{Ax}+\boldsymbol{b}))\boldsymbol{A}$$

5.8 Compute the derivatives $\mathrm{d}f/\mathrm{d}\boldsymbol{x}$ of the following functions. Describe your steps in detail.

a. Use the chain rule. Provide the dimensions of every single partial derivative.

$$f(z) = \exp(-\tfrac{1}{2}z)$$
$$z = g(\boldsymbol{y}) = \boldsymbol{y}^\top\boldsymbol{S}^{-1}\boldsymbol{y}$$
$$\boldsymbol{y} = h(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\mu}$$

where $\boldsymbol{x}, \boldsymbol{\mu} \in \mathbb{R}^D$, $\boldsymbol{S} \in \mathbb{R}^{D \times D}$.

The desired derivative can be computed using the chain rule:

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \underbrace{\frac{\partial f}{\partial z}}_{1\times 1} \underbrace{\frac{\partial g}{\partial \boldsymbol{y}}}_{1\times D} \underbrace{\frac{\partial h}{\partial \boldsymbol{x}}}_{D\times D} \in \mathbb{R}^{1\times D}$$

Here

$$\begin{aligned}
\frac{\partial f}{\partial z} &= -\tfrac{1}{2}\exp(-\tfrac{1}{2}z)\\
\frac{\partial g}{\partial \boldsymbol{y}} &= 2\boldsymbol{y}^\top \boldsymbol{S}^{-1}\\
\frac{\partial h}{\partial \boldsymbol{x}} &= \boldsymbol{I}_D
\end{aligned}$$

so that

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = -\exp\big(-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big)(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{S}^{-1}$$

b.

$$f(\boldsymbol{x}) = \mathrm{tr}(\boldsymbol{x}\boldsymbol{x}^\top + \sigma^2 \boldsymbol{I}), \quad \boldsymbol{x} \in \mathbb{R}^D$$

Here $\mathrm{tr}(\boldsymbol{A})$ is the trace of $\boldsymbol{A}$, i.e., the sum of the diagonal elements A_{ii}.
Hint: Explicitly write out the outer product.

Let us have a look at the outer product. We define $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\top$ with

$$X_{ij} = x_i x_j$$

The trace sums up all the diagonal elements, such that

$$\frac{\partial}{\partial x_j}\mathrm{tr}(\boldsymbol{X} + \sigma^2 \boldsymbol{I}) = \sum_{i=1}^{D} \frac{\partial X_{ii} + \sigma^2}{\partial x_j} = 2x_j$$

for $j = 1, \ldots, D$. Overall, we get

$$\frac{\partial}{\partial \boldsymbol{x}}\mathrm{tr}(\boldsymbol{x}\boldsymbol{x}^\top + \sigma^2 \boldsymbol{I}) = 2\boldsymbol{x}^\top \in \mathbb{R}^{1\times D}$$

c. Use the chain rule. Provide the dimensions of every single partial derivative. You do not need to compute the product of the partial derivatives explicitly.

$$\begin{aligned}
\boldsymbol{f} &= \tanh(\boldsymbol{z}) \in \mathbb{R}^M\\
\boldsymbol{z} &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \quad \boldsymbol{x} \in \mathbb{R}^N, \boldsymbol{A} \in \mathbb{R}^{M\times N}, \boldsymbol{b} \in \mathbb{R}^M.
\end{aligned}$$

Here, $\tanh$ is applied to every component of $\boldsymbol{z}$.

$$\begin{aligned}
\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}} &= \mathrm{diag}(1 - \tanh^2(\boldsymbol{z})) \in \mathbb{R}^{M\times M}\\
\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} &= \frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A} \in \mathbb{R}^{M\times N}
\end{aligned}$$

We get the latter result by defining $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, such that

$$y_i = \sum_j A_{ij} x_j \implies \frac{\partial y_i}{\partial x_k} = A_{ik} \implies \frac{\partial y_i}{\partial \boldsymbol{x}} = [A_{i1}, ..., A_{iN}] \in \mathbb{R}^{1\times N}$$
$$\implies \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \boldsymbol{A}$$

The overall derivative is an $M \times N$ matrix.

5.9 We define

$$g(\boldsymbol{z}, \boldsymbol{\nu}) := \log p(\boldsymbol{x}, \boldsymbol{z}) - \log q(\boldsymbol{z}, \boldsymbol{\nu})$$
$$\boldsymbol{z} := t(\boldsymbol{\epsilon}, \boldsymbol{\nu})$$

for differentiable functions p, q, t. By using the chain rule, compute the gradient

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\nu}} g(\boldsymbol{z}, \boldsymbol{\nu}) .$$

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\nu}} g(\boldsymbol{z}, \boldsymbol{\nu}) &= \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\nu}} \left(\log p(\boldsymbol{x}, \boldsymbol{z}) - \log q(\boldsymbol{z}, \boldsymbol{\nu})\right) = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\nu}} \log p(\boldsymbol{x}, \boldsymbol{z}) - \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\nu}} \log q(\boldsymbol{z}, \boldsymbol{\nu}) \\
&= \frac{\partial}{\partial \boldsymbol{z}} \log p(\boldsymbol{x}, \boldsymbol{z}) \frac{\partial t(\boldsymbol{\epsilon}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} - \frac{\partial}{\partial \boldsymbol{z}} \log q(\boldsymbol{z}, \boldsymbol{\nu}) \frac{\partial t(\boldsymbol{\epsilon}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} - \frac{\partial}{\partial \boldsymbol{\nu}} \log q(\boldsymbol{z}, \boldsymbol{\nu}) \\
&= \left(\frac{\partial}{\partial \boldsymbol{z}} \log p(\boldsymbol{x}, \boldsymbol{z}) - \frac{\partial}{\partial \boldsymbol{z}} \log q(\boldsymbol{z}, \boldsymbol{\nu})\right) \frac{\partial t(\boldsymbol{\epsilon}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} - \frac{\partial}{\partial \boldsymbol{\nu}} \log q(\boldsymbol{z}, \boldsymbol{\nu}) \\
&= \left(\frac{1}{p(\boldsymbol{x}, \boldsymbol{z})} \frac{\partial}{\partial \boldsymbol{z}} p(\boldsymbol{x}, \boldsymbol{z}) - \frac{1}{q(\boldsymbol{z}, \boldsymbol{\nu})} \frac{\partial}{\partial \boldsymbol{z}} q(\boldsymbol{z}, \boldsymbol{\nu})\right) \frac{\partial t(\boldsymbol{\epsilon}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} - \frac{\partial}{\partial \boldsymbol{\nu}} \log q(\boldsymbol{z}, \boldsymbol{\nu})
\end{aligned}$$