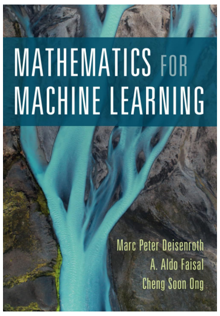
Mathematics for Machine Learning (Deisenroth, Faisal, Ong)

Chapter 5: Vector Calculus

Full solutions to exercises 5.1 - 5.9



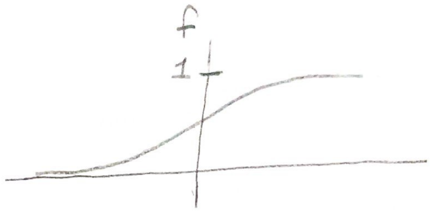
CONTENTS (clickable links)

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

$$f'(x) = \frac{d}{dx}\left(\log(x^4)\right)\sin(x^3) + \log(x^4)\frac{d}{dx}\left(\sin(x^3)\right)$$

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

5.2) $f(x) = \dfrac{1}{1+e^{-x}}$



$$f'(x) = \frac{d}{dx}\left(1+e^{-x}\right)^{-1}$$

$$= \frac{-1}{(1+e^{-x})^2}\cdot -e^{-x} = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{(e^{x/2}+e^{-x/2})^2}$$

5.3) $f(x) = e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

$$f'(x) = e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\cdot\frac{d}{dx}\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

$$= -\frac{x-\mu}{\sigma^2}e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = -\frac{x-\mu}{\sigma^2}f(x).$$

5.4) T_n @ $x_0 = 0$: $\quad T_n(x) = \sum_{i=0}^{n}\frac{f^{(i)}(0)}{i!}x^i$

$f(x) = \sin(x) + \cos(x) \Rightarrow f(0) = 1$

$f^{(1)}(x) = \cos(x) - \sin(x) \Rightarrow f^{(1)}(0) = 1$

$f^{(2)}(x) = -\sin(x) - \cos(x) \Rightarrow f^{(2)}(0) = -1$

$f^{(3)}(x) = -\cos(x) + \sin(x) \Rightarrow f^{(3)}(0) = -1$

$f^{(4)}(x) = \sin(x) + \cos(x) \Rightarrow f^{(4)}(0) = 1$

$f^{(5)}(x) = \cos(x) - \sin(x) \Rightarrow f^{(5)}(0) = 1$

Thus, $T_5(x) = 1 + x - \frac{1}{2!}x^2 - \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \frac{1}{5!}x^5$

$$= \underbrace{\underbrace{\underbrace{\underbrace{\underbrace{1}_{T_0} + x}_{T_1} - \frac{x^2}{2}}_{T_2} - \frac{x^3}{6}}_{T_3} + \frac{x^4}{24}}_{T_4} + \frac{x^5}{120}$$

5.5 a+b)

- $f_1 : \mathbb{R}^2 \longrightarrow \mathbb{R}$

$$\underline{x} \longmapsto \sin(x_1)\cos(x_2).$$

So $\frac{\partial f_1}{\partial \underline{x}} \in \mathbb{R}^{1\times 2}$

$$\frac{\partial f_1}{\partial x_1} = \cos(x_1)\cos(x_2) \quad , \quad \frac{\partial f_1}{\partial x_2} = -\sin(x_1)\sin(x_2)$$

Thus, $J = \frac{\partial f_1}{\partial \underline{x}} = \left(\frac{\partial f_1}{\partial x_1}, \frac{\partial f_1}{\partial x_2}\right) = \left(\cos(x_1)\cos(x_2), -\sin(x_1)\sin(x_2)\right).$

- $f_2 : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$

$$(\underline{x}, \underline{y}) \longmapsto x^T \underline{y} = \sum_{i=1}^{n} x_i y_i$$

$\frac{\partial f_2}{\partial \underline{x}} \in \mathbb{R}^{1\times n}$ ("$\underline{y}$" is constant)

$$J = \frac{\partial f_2}{\partial(\underline{x}, \underline{y})} = \left(\frac{\partial f_2}{\partial x_1}, \ldots, \frac{\partial f_2}{\partial x_n}, \frac{\partial f_2}{\partial y_1}, \ldots, \frac{\partial f_2}{\partial y_n}\right)$$

$$= (y_1, y_2, \ldots, y_n, x_1, \ldots, x_n). \in \mathbb{R}^{1\times(n+n)} = \mathbb{R}^{1\times(2n)}.$$

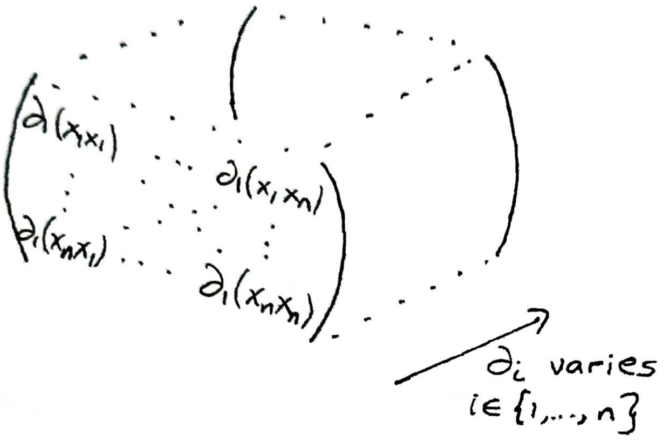
- $f_3 : \mathbb{R}^n \longrightarrow \mathbb{R}^{n \times n}$

$$\underline{x} \longmapsto \underline{x}\,\underline{x}^T = \begin{pmatrix} x_1x_1 & x_1x_2 & \cdots & x_1x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 & x_nx_2 & \cdots & x_nx_n \end{pmatrix}$$

$$\frac{\partial f_3}{\partial \underline{x}} \in \mathbb{R}^{(n\times n)\times n}$$

$J = \dfrac{\partial f_3}{\partial \underline{x}}$ is an $n \times n \times n$ cube.

$J =$



where $\partial_i(-) := \dfrac{\partial}{\partial x_i}(-)$

Let J_i be the "i^{th} sheet/layer" – corresponding to $\partial_i := \dfrac{\partial}{\partial x_i}$

Then $J_i = \begin{pmatrix} 0 & x_1 & 0 \\ & \vdots & \\ x_1 \cdots & 2x_i & \cdots x_n \\ & \vdots & \\ 0 & x_n & 0 \end{pmatrix}$

e.g. $J_1 = \begin{pmatrix} 2x_1 & x_2 & \cdots & x_n \\ x_2 & & & \\ \vdots & & 0 & \\ x_n & & & \end{pmatrix}$, $J_n = \begin{pmatrix} & & & x_1 \\ & 0 & & x_2 \\ & & & \vdots \\ x_1 & x_2 & \cdots & 2x_n \end{pmatrix}$

5.6) • $f(\underline{t}) = \sin(\log(\underline{t}^T\underline{t}))$

$$\frac{df}{d\underline{t}} = \cos(\log(\underline{t}^T\underline{t})) \cdot \frac{1}{\underline{t}^T\underline{t}} \cdot \frac{d}{d\underline{t}}(\underline{t}^T\underline{t})$$

$$= \cos(\log(\underline{t}^T\underline{t})) \cdot \frac{1}{\underline{t}^T\underline{t}} \cdot \frac{d}{d\underline{t}}(\underline{t} \longmapsto t_1^2 + \dots + t_n^2)$$

$$= \frac{\cos(\log(\underline{t}^T\underline{t}))}{\underline{t}^T\underline{t}} \cdot (2t_1, \dots, 2t_n)$$

$$= \frac{2\cos(\log(\underline{t}^T\underline{t}))}{\underline{t}^T\underline{t}} \underline{t}^T.$$

• $g(X) = tr(AXB)$

This is a map $\mathbb{R}^{E\times F} \longrightarrow \mathbb{R}$, so $\frac{\partial g}{\partial X} \in \mathbb{R}^{1\times(E\times F)} \cong \mathbb{R}^{E\times F}$.

Given matrix M, write M_i for the i^{th} column, and M_{ij} for the (i,j)-entry.

$$(AX)_{ij} = \sum_{k=1}^{E} A_{ik} X_{kj}$$

$$\Rightarrow (AXB)_{ij} = \sum_{l=1}^{F} (AX)_{il} B_{lj} = \sum_{l=1}^{F}\sum_{k=1}^{E} A_{ik} X_{kl} B_{lj}$$

$$\Rightarrow tr(AXB) = \sum_{i=1}^{D} (AXB)_{ii} = \sum_{i=1}^{D}\sum_{k=1}^{E}\sum_{l=1}^{F} A_{ik} X_{kl} B_{li}$$

so $$\frac{\partial\, tr(AXB)}{\partial X_{kl}} = \sum_{i=1}^{D} A_{ik} B_{li} = (A_k)^T \cdot (B^T)_l$$

i.e. the (k,l) entry of $\frac{\partial g}{\partial X}$ is the dot product of k^{th} column of A & l^{th} column of B^T

$$\Rightarrow \frac{\partial\, tr(AXB)}{\partial X} = A^T B^T$$

5.7 a) $f(z) = \log(1+z)$, $z = \underline{x}^T\underline{x}$, $\underline{x} \in \mathbb{R}^D$.

$$\begin{array}{ccccc} \mathbb{R}^D & \xrightarrow{z} & \mathbb{R} & \xrightarrow{f} & \mathbb{R} \\ \underline{x} & \longmapsto & \underline{x}^T\underline{x} & \longmapsto & \log(1+z). \end{array}$$

$$\underbrace{\frac{\partial f}{\partial \underline{x}}}_{\in \mathbb{R}^{1\times D}} = \underbrace{\frac{\partial f}{\partial z}}_{\in \mathbb{R}} \cdot \underbrace{\frac{\partial z}{\partial \underline{x}}}_{\in \mathbb{R}^{1\times D}}$$

$$\frac{\partial f}{\partial z} = \frac{\partial}{\partial z}\left(\log(1+z)\right) = \frac{1}{1+z} = \frac{1}{1+\underline{x}^T\underline{x}}$$

$$\frac{\partial z}{\partial \underline{x}} = \frac{\partial}{\partial \underline{x}}\left(\underline{x}^T\underline{x}\right) = 2\underline{x}^T \quad \left(\text{to see this, write } \underline{x}^T\underline{x} = \sum_{i=1}^{D} x_i^2\right)$$

So $$\frac{\partial f}{\partial \underline{x}} = \frac{2\underline{x}^T}{1+\underline{x}^T\underline{x}}$$

b) $f(\underline{z}) = \sin(\underline{z})$, $\underline{z} = A\underline{x} + \underline{b}$

$$\begin{array}{ccccc} \mathbb{R}^D & \xrightarrow{z} & \mathbb{R}^E & \xrightarrow{f} & \mathbb{R}^E \\ \underline{x} & \longmapsto & A\underline{x}+\underline{b} & \longmapsto & \underline{v} \text{ s.t } \underline{v}_i = \sin\left((A\underline{x}+\underline{b})_i\right). \end{array}$$

$$\underbrace{\frac{\partial f}{\partial \underline{x}}}_{\in \mathbb{R}^{E\times D}} = \underbrace{\frac{\partial f}{\partial \underline{z}}}_{\in \mathbb{R}^{E\times E}} \cdot \underbrace{\frac{\partial \underline{z}}{\partial \underline{x}}}_{\in \mathbb{R}^{E\times D}}$$

Calculating $\frac{\partial f}{\partial \underline{z}}$, where $f(\underline{z}) = \sin(\underline{z}) = \begin{pmatrix} \sin(z_1) \\ \vdots \\ \sin(z_E) \end{pmatrix}$

$f_i(\underline{z}) = \sin(z_i) \Rightarrow \frac{\partial f_i}{\partial z_i} = \cos(z_i)$, $\frac{\partial f_i}{\partial z_j} = 0$ if $i \neq j$.

so $\frac{\partial f}{\partial \underline{z}}$ is diagonal $E\times E$ matrix with $\cos(z_i)$ in (i,i) entry.

i.e. $$\frac{\partial f}{\partial \underline{z}} = \begin{pmatrix} \cos(z_1) & & \\ & \ddots & \\ & & \cos(z_E) \end{pmatrix}$$

Calculating $\frac{\partial \underline{z}}{\partial \underline{x}}$, where $\underline{z} = A\underline{x} + \underline{b}$.

(This is clearly A itself, but let's check.)

$$z_i = \left(\sum_{k=1}^{D} A_{ik} x_k\right) + b_i \Rightarrow \frac{\partial z_i}{\partial x_j} = A_{ij}. \Rightarrow \frac{\partial \underline{z}}{\partial \underline{x}} = A.$$

so $$\frac{\partial f}{\partial \underline{x}} = \begin{pmatrix} \cos(z_1) & & \\ & \ddots & \\ & & \cos(z_E) \end{pmatrix} A$$

$$= \begin{pmatrix} \cos(z_1)a_{11} & \cos(z_1)a_{12} & \cdots & \cos(z_1)a_{1D} \\ \vdots & & \ddots & \vdots \\ \cos(z_E)a_{E1} & \cdots & \cdots & \cos(z_E)a_{ED} \end{pmatrix}.$$

5.8 a)
$$\mathbb{R}^D \xrightarrow{h} \mathbb{R}^D \xrightarrow{g} \mathbb{R} \xrightarrow{f} \mathbb{R}$$
$$\underline{x} \longmapsto \underline{y} = \underline{x} - \underline{\mu} \longmapsto z = \underline{y}^T S^{-1} \underline{y} \longmapsto \exp\left(-\frac{z}{2}\right).$$

$$\underbrace{\frac{df}{d\underline{x}}}_{\in \mathbb{R}^{1\times D}} = \underbrace{\frac{df}{dz}}_{\in \mathbb{R}} \cdot \underbrace{\frac{dz}{d\underline{y}}}_{\in \mathbb{R}^{1\times D}} \cdot \underbrace{\frac{d\underline{y}}{d\underline{x}}}_{\in \mathbb{R}^{D\times D}}.$$

$$\frac{df}{dz} = \frac{d}{dz}\left(\exp\left(-\frac{z}{2}\right)\right) = -\frac{1}{2}\exp\left(-\frac{z}{2}\right).$$

$$\frac{dz}{d\underline{y}} = \frac{d}{d\underline{y}}\left(\underline{y}^T S^{-1} \underline{y}\right) = y^T\left(S^{-1} + S^{-T}\right) \qquad \left(\text{see } §5.5, (5.107)\right).$$

$$\frac{d\underline{y}}{d\underline{x}} = \frac{d}{d\underline{x}}\left(\underline{x} - \underline{\mu}\right) = \frac{d}{d\underline{x}}\left(Id\,\underline{x} - \underline{\mu}\right) = Id$$ ← $D \times D$ identity matrix.

So $$\frac{df}{d\underline{x}} = -\frac{1}{2}\exp\left(-\frac{z}{2}\right) \cdot y^T\left(S^{-1} + S^{-T}\right)$$

$$= \underbrace{-\frac{1}{2}\exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu})^T S^{-1}(\underline{x}-\underline{\mu})\right)}_{\text{scalar}} \cdot \underbrace{(\underline{x}-\underline{\mu})^T\left(S^{-1}+S^{-T}\right)}_{\in \mathbb{R}^{1\times D}}.$$

b) $f(\underline{x}) = tr\left(\underline{x}\underline{x}^T + \sigma^2 I\right)$, $\underline{x} \in \mathbb{R}^D$.

$f: \mathbb{R}^D \to \mathbb{R}$, so $\frac{df}{d\underline{x}} \in \mathbb{R}^{1\times D}$.

$$\underline{x}\underline{x}^T + \sigma^2 I = \begin{pmatrix} x_1 \\ \vdots \\ x_D \end{pmatrix}\left(x_1 \dots x_D\right) + \begin{pmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{pmatrix} = \begin{pmatrix} x_1^2+\sigma^2 & & * \\ & x_2^2+\sigma^2 & \\ * & & \ddots \\ & & & x_D^2+\sigma^2 \end{pmatrix}$$

so $f(\underline{x}) = tr\left(\underline{x}\underline{x}^T + \sigma^2 I\right) = \left(x_1^2 + x_2^2 + \dots + x_D^2\right) + D\sigma^2$

Thus, $$\frac{\partial f}{\partial x_i} = 2x_i$$

$$\Rightarrow \frac{df}{d\underline{x}} = \left(2x_1, 2x_2, \dots, 2x_D\right)$$

$$= 2\underline{x}^T.$$

c) $$\mathbb{R}^N \xrightarrow{\underline{z}} \mathbb{R}^M \xrightarrow{\underline{f}} \mathbb{R}^M$$

$$\underline{x} \longmapsto \underline{z} = A\underline{x} + \underline{b} \longmapsto \tanh(\underline{z}) := \begin{pmatrix} \tanh(z_1) \\ \vdots \\ \tanh(z_M) \end{pmatrix}.$$

$$\underbrace{\frac{d\underline{f}}{d\underline{x}}}_{\in \mathbb{R}^{M\times N}} = \underbrace{\frac{d\underline{f}}{d\underline{z}}}_{\in \mathbb{R}^{M\times M}} \cdot \underbrace{\frac{d\underline{z}}{d\underline{x}}}_{\in \mathbb{R}^{M\times N}}$$

$$\frac{d\underline{f}}{d\underline{z}} = \frac{d}{d\underline{z}}\left(\tanh(\underline{z}) := \begin{pmatrix} \tanh(z_1) \\ \vdots \\ \tanh(z_M) \end{pmatrix}\right)$$

$\left(\text{Note: } \frac{d}{dt}\left(\tanh(t)\right) = \operatorname{sech}^2 x\right)$

We have $$\frac{\partial \underline{f}}{\partial z_i} = \begin{pmatrix} 0 \\ \vdots \\ \frac{d}{dz_i}(\tanh(z_i)) \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \operatorname{sech}^2(z_i) \\ \vdots \\ 0 \end{pmatrix}.$$

so $$\frac{d\underline{f}}{d\underline{z}} = \begin{pmatrix} \operatorname{sech}^2(z_1) & & & 0 \\ & \operatorname{sech}^2(z_2) & & \\ & & \ddots & \\ 0 & & & \operatorname{sech}^2(z_M) \end{pmatrix}$$

$$\frac{d\underline{z}}{d\underline{x}} = \frac{d}{d\underline{x}}(A\underline{x} + b) = A.$$

So $$\frac{d\underline{f}}{d\underline{x}} = \begin{pmatrix} \operatorname{sech}^2(z_1) & & \\ & \ddots & \\ & & \operatorname{sech}^2(z_m) \end{pmatrix} A$$

$z_i = (A\underline{x} + \underline{b})_i = \sum_{j=1}^{N} A_{ij} x_j + b_i$

$$= \begin{pmatrix} \operatorname{sech}^2\left(\sum_{j=1}^{N} A_{1j} x_j + b_1\right) & & 0 \\ & \ddots & \\ 0 & & \operatorname{sech}^2\left(\sum_{j=1}^{N} A_{Mj} x_j + b_M\right) \end{pmatrix} A$$

5.9)

$$\begin{tikzcd}
\nu \arrow[rr] \arrow[dr] & & q=q(z,\nu) \arrow[dr] & \\
\varepsilon \arrow[r] & z:=t(\varepsilon,\nu) \arrow[ur] \arrow[dr] & & g=\log p-\log q. \\
x \arrow[rr] & & p=p(x,z) \arrow[ur] &
\end{tikzcd}$$

$$\frac{d}{d\nu} g(x,z,\nu) = \frac{\partial g}{\partial x}\cdot\cancelto{0}{\frac{dx}{d\nu}} + \frac{\partial g}{\partial z}\cdot\frac{dz}{d\nu} + \frac{\partial g}{\partial \nu}\cdot\overbrace{\frac{d\nu}{d\nu}}^{=1}.$$

$$= \frac{\partial g}{\partial z}\cdot\frac{dz}{d\nu} + \frac{\partial g}{\partial \nu}$$

$$= \frac{\partial}{\partial z}\Big(\log(p(x,z)) - \log(q(z,\nu))\Big)\frac{d}{d\nu}(t(\varepsilon,\nu))$$

$$- \frac{\partial}{\partial \nu}\Big(\log(q(z,\nu))\Big)$$

$$= \left(\frac{\frac{\partial}{\partial z}(p(x,z))}{p(x,z)} - \frac{\frac{\partial}{\partial z}(q(z,\nu))}{q(z,\nu)}\right)\frac{dz}{d\nu} - \frac{\frac{\partial}{\partial \nu}(q(z,\nu))}{q(z,\nu)}.$$

$$= \frac{1}{p(x,z)}\left[\frac{\partial}{\partial z}(p(x,z))\frac{dz}{d\nu}\right] - \frac{1}{q(z,\nu)}\left[\frac{\partial}{\partial z}(q(z,\nu))\frac{dz}{d\nu} + \frac{\partial}{\partial \nu}(q(z,\nu))\right]$$