AIMS African Masters in Machine Intelligence 2024

Exercises: Gradients and Jacobians

In all these exercises, $\langle \cdot, \cdot \rangle$ is the standard euclidean inner product $\langle x, y \rangle = x^T y$

1. Gradients

Let $x \in \mathbb{R}^n$ with $n \in \mathbb{N}$.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function.

The *gradient* $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ can be defined either by its components

$$[\nabla f(x)]_i = \frac{\partial f}{\partial x_i}(x),$$

or as the unique vector $in \mathbb{R}^n$ such that for small $h \in \mathbb{R}^n$,

$$f(x+h) \underset{\|h\| \to 0}{=} f(x) + \langle \nabla f(x), h \rangle + o(h).$$

In each of the questions, find ∇f, using either one (or both) of the above definitions.

1. $f(x) = \langle c, x \rangle, \quad c \in \mathbb{R}^n$
2. $f(x) = x^T A x + c^T x + b \quad A \in \mathbb{R}^{n \times n} c \in \mathbb{R}^n, b \in \mathbb{R}$
3. $f(x) = \langle a(x), b(x) \rangle, \quad a, b : \mathbb{R}^n \to \mathbb{R}^m$ differentiable functions with Jacobians J_a and J_b
4. (a) $f(x) = \sigma(w^T x + b), \quad \sigma : \mathbb{R} \to \mathbb{R}$ a differentiable function $c \in \mathbb{R}^n$

   (b) Simplify (a) when σ is one of these well-known activation functions: $\sigma(x) = x^+$, $\sigma(x) = \frac{1}{1+e^{-x}}$
5. (a) Chain rule: $f = a \circ b$

   $f(x) = a(b(x)), \quad a : \mathbb{R} \to \mathbb{R}$, derivative a', $\quad b : \mathbb{R}^n \to \mathbb{R}$, gradient ∇b

   (b) what is expression (a) when $a(x) = \log(1 + e^{-x})$

2. Jacobians

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a differentiable function.
The *Jacobian* $J_f : \mathbb{R}^n \to \mathbb{R}^{m\times n}$ can be defined in two equivalent ways:

- The matrix of partial derivatives: $[J_f(x)]_{ij} = \frac{\partial f_i}{\partial x_j}(x)$
- The matrix such that, for $x, h \in \mathbb{R}^n$, $f(x+h) \underset{\|h\|\to 0}{=} f(x) + J_f(x)h + o(h)$

1. What is the relationship between ∇f_i – the gradient of the i-th component of f – and the Jacobian J_f ?

   In each of the following questions, find J_f, using either one (or both) of the above definitions.

2. Polar-Cartesian change of coordinates: $f(r,\theta) = [r\cos\theta, r\sin\theta]^T$
3. Spherical-Cartesian change of coordinates. $f(r,\theta,\varphi) = [r\cos\theta\sin\varphi, r\sin\theta\sin\varphi, r\cos\varphi]^T$
4. $f(x) = [g(x_1), \ldots, g(x_n)]^T, \quad g : \mathbb{R} \to \mathbb{R}$ differentiable function.
5. *Chain rule*: $f = a \circ b, \quad a, b : \mathbb{R}^n \to \mathbb{R}^m$ differentiable functions with Jacobians J_a and J_b

   Hint: Use the Taylor expansion of b*, then the Taylor expansion of* a *with* $h' := J_b h$
6. (a) $f(x) = \sigma(Wx + b), \quad W \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, \sigma : \mathbb{R} \to \mathbb{R}$ an elementwise function with derivative σ'

   (b) *higher dimensions:* What is the dimension of $\nabla_W f$?