# Exercise: Linear Models

Due: Optional

---

**Problem 1** (Gradient for quadratic functions)

In this problem we develop gradients for quadratic functions.

(1) We first consider the *two-dimensional case*. Let

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2.$$

and

$$f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} = a_{1,1}x_1^2 + (a_{1,2} + a_{2,1})x_1x_2 + a_{2,2}x_2^2.$$

Prove that

$$\nabla f(\mathbf{x}) = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} a_{1,1} & a_{2,1} \\ a_{1,2} & a_{2,2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

(2) We now turn to more *general cases*. Let (this is a challenging question:))

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,d} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,d} \\ \vdots & \vdots & \ldots & \vdots \\ a_{d,1} & a_{d,2} & \cdots & a_{d,d} \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d. \end{pmatrix} \tag{1}$$

Define

$$f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} = \sum_{i,j=1}^{d} a_{i,j}x_ix_j.$$

Prove that

$$\nabla\big(\mathbf{x}^\top A\mathbf{x}\big) = A\mathbf{x} + A^\top\mathbf{x}.$$

**Solution 1**

The case with $d = 2$ follows directly from the definition of gradients

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2a_{1,1}x_1 + (a_{1,2} + a_{2,1})x_2 \\ (a_{1,2} + a_{2,1})x_1 + 2a_{2,2} \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} a_{1,1} & a_{2,1} \\ a_{1,2} & a_{2,2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

We can group the summands in $f(\mathbf{w})$ as follows

$$\sum_{i,j=1}^{d} a_{i,j}x_ix_j = \underbrace{\sum_{i,j\neq k} a_{i,j}x_ix_j}_{\text{no index is } k} + \underbrace{\sum_{i\neq k} a_{i,k}x_ix_k}_{\text{2nd index is } k} + \underbrace{\sum_{j\neq k} a_{k,j}x_kx_j}_{\text{1st index is } k} + \underbrace{a_{kk}x_k^2}_{\text{both indices are } k}.$$

It then follows that

$$\begin{aligned}
\frac{\partial f(\mathbf{x})}{\partial x_k} &= \sum_{i,j\neq k} \frac{\partial a_{i,j}x_ix_j}{\partial x_k} + \sum_{i\neq k} a_{i,k}\frac{\partial x_ix_k}{\partial x_k} + \sum_{j\neq k} a_{k,j}\frac{\partial x_kx_j}{\partial x_k} + a_{kk}\frac{\partial x_k^2}{\partial x_k} \\
&= 0 + \sum_{i\neq k} a_{i,k}x_i + \sum_{j\neq k} a_{k,j}x_j + 2a_{k,k}x_k \\
&= \sum_{i=1}^{d} a_{i,k}x_i + \sum_{j=1}^{d} a_{k,j}x_j.
\end{aligned}$$

It follows that

$$\begin{aligned}
\begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_d} \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^d a_{i,1}x_i + \sum_{j=1}^d a_{1,j}x_j \\ \sum_{i=1}^d a_{i,2}x_i + \sum_{j=1}^d a_{2,j}x_j \\ \vdots \\ \sum_{i=1}^d a_{i,d}x_i + \sum_{j=1}^d a_{d,j}x_j \end{pmatrix} \\
&= \begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{d,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{d,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,d} & a_{2,d} & \cdots & a_{d,d} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} + \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,d} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d,1} & a_{d,2} & \cdots & a_{d,d} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} = A^\top\mathbf{x} + A\mathbf{x}.
\end{aligned}$$