Review Assignment 2

Nalet Meinen
Machine Learning

October 8, 2019

1 Calculus review

Recall that the Jacobian of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is an $m \times n$ matrix of partial derivatives

$$Df(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_2(x)}{\partial x_3} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \frac{\partial f_m(x)}{\partial x_2} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}$$

where $x = [x_1 x_2 \dots x_n]^\intercal$, $f(x) = [f_1(x) f_2(x) \dots f_m(x)]^\intercal$ and $\frac{\partial f_i(x)}{\partial x_j}$ is the partial derivative of the i-th output with respect to the j-th input. When f is a scalar-valued function (i.e., when $f : \mathbb{R}^n \to \mathbb{R}$), the Jacobian $Df(x)$ is a $1 \times n$ matrix, i.e., it is a row vector. Its transpose is called the *gradient* of the function

$$\nabla f(x) = Df(x)^\intercal \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \tag{1}$$

Also, recall that the **chain rule** is a tool to calculate gradients of function compositions. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at x and $g : \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $f(x)$. Define the composition $h : \mathbb{R}^m \to \mathbb{R}^p$ by $h(z) = g(f(z))$. Then h is differentiable at x, with Jacobian

$$Dh(x) = Dg(z)\Big|_{z=f(x)} Df(x). \tag{2}$$

1. Consider the function $g : \mathbb{R}^m \to \mathbb{R}$ with $g(x) = x^\intercal x$. We can readily calculate the gradient $\nabla g(x) = 2x$ by noticing that

$$\forall j = 1, \dots, n \qquad \frac{\partial x^\intercal x}{\partial x_j} = \frac{\partial x^2 j}{\partial x_j} = 2x_j \to \nabla g(x) = 2x \tag{3}$$

Consider also the function $a : \mathbb{R}^n \to \mathbb{R}^m$ with $a(x) = Ax$, and $A \in \mathbb{R}^{m\times n}$. The Jacobian of $a(x)$ is $Da(x) = A$. Given this, answer the following questions by using the above definitions (show all the steps of your working)

(a) Consider the function $h : \mathbb{R}^n \to \mathbb{R}$ and $h(x) = x^\intercal Qx$, where $Q \in \mathbb{R}^{n\times n}$ is a symmetric matrix. Calculate $.h(x)$ by using the product rule, the gradient of g in eq. (3), and the Jacobian of the linear function $a(x)$.

(b) Consider the function $f : \mathbb{R}^n \to \mathbb{R}$, where $f(x) = \|Ax - b\|^2$, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$. Calculate $\nabla h(x)$ by using the chain rule in eq. (2), the gradient of g in eq. (3), and the Jacobian of the linear function $a(x)$.

(c) Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. Suppose we have a matrix $A \in \mathbb{R}^{n\times m}$ and a vector $x \in \mathbb{R}^m$. Calculate $\nabla x f(Ax)$ as a function of $\nabla x f(x)$.

(d) Show that

$$\frac{\partial}{\partial X}\sum_{i=1} n\lambda_i = 1$$

where $X \in \mathbb{R}^{m\times n}$ and has eigenvalues $\lambda_1 \dots \lambda_n$

(e) Show that

$$\frac{\partial}{\partial X}\prod_{i=1} n\lambda_i = (det)(X)X^{-\intercal}$$

where $X \in \mathbb{R}^{m\times n}$ and has eigenvalues $\lambda_1 \dots \lambda_n$

2. Assume $A \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^{m\times n}$, and $B \in \mathbb{R}^{m\times m}$. Show that $\nabla X\mathrm{tr}(AX^\intercal B) = BA$

3. Solve the following equality constrained optimization problem

$$\underset{\max}{x \in} \mathbb{R}^n x^\intercal Ax \qquad \text{subject to } b^\intercal x = 1$$

for a symmetric matrix $A \in \mathbb{S}^n$. Assume that A is invertible and $b \neq 0$.