

Review Assignment 2

Nalet Meinen
Machine Learning

October 9, 2019

1 Calculus review

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

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

where $x = [x_1 x_2 \dots x_n]^\top$, $f(x) = [f_1(x) f_2(x) \dots f_m(x)]^\top$ 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$), 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

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

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

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

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

$$\forall j = 1, \dots, n \quad \frac{\partial x^\top x}{\partial x_j} = \frac{\partial x^2_j}{\partial x_j} = 2x_j \rightarrow \nabla g(x) = 2x \quad (3)$$

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

We notice that $\nabla g(x) = 2x$, therefore

$$\begin{aligned}\nabla h(x) &= \frac{\partial x^\top Qx^\top}{\partial x} \\ &= 2Qx\end{aligned}$$

- (b) Consider the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$. 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 $\nabla x f(x)$.
-

$$\begin{aligned}\|Ax - b\|^2 &= (Ax - b)^\top (Ax - b) \\ &= x^\top A^\top Ax - 2b^\top Ax + b^\top b\end{aligned}$$

$$\begin{aligned}\nabla h(x) &= \frac{\partial x^\top A^\top Ax - 2b^\top Ax^\top}{\partial x} \\ &= 2Qx\end{aligned}$$

- (d) Show that

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

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

-
- (e) Show that

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

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 $\nabla X \text{tr}(AX^\top B) = BA$

$$\begin{aligned}
 BA &= \nabla X \text{tr}(AX^\top B) \\
 &= \nabla X \text{tr}(X^\top BA) && \text{we can get rid of the trace} \\
 &= ((BA)^\top)^\top && \text{transpose of transpose eliminate each other} \\
 &= BA
 \end{aligned}$$

3. Solve the following equality constrained optimization problem

$$\max_{x \in \mathbb{R}^n} x^\top A x \quad \text{subject to } b^\top x = 1$$

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

A standard way of solving optimization problems with equality constraints is by forming the Lagrangian, an objective function that includes the equality constraints. The Lagrangian in this case is given by

$$\mathcal{L}(x, \lambda) = x^\top A x - \lambda(b^\top x - 1).$$

The parameter λ is called the Lagrangian multiplier associated with the equality constraint. It can be shown that for x^* to be an optimal solution to the problem, the gradient of the Lagrangian w.r.t. x has to be zero at x^* . That is,

$$\nabla_x(\mathcal{L}(x, \lambda)) = \nabla_x(x^\top A x - \lambda b^\top x) \stackrel{!}{=} 2Ax - 2\lambda b$$

This shows that the only points which can be possibly maximize (or minimize) $b^\top A x$ assuming $x^\top b = 1$ are the eigenvectors of A .