UNIVERSITY OF BERN

Calculus Review

Student name: *Fabienne Guertler 12-935-508*

Course: *Machine Learning*
Due date: *October 9th, 2019*

Question 1

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

$$\forall j = 1, ..., n \quad \frac{\partial x^T x}{\partial x_j} = \frac{\partial x_j^2}{\partial x_j} = 2x_j \to \nabla g(x) = 2x$$

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 Jakobian of $a(x)$ is $Da(x) = a$. Given this, answer the following questions.

(*a*) Consider the function $h : \mathbb{R}^n \to \mathbb{R}$ and $h(x) = x^T Q x$, where $Q \in \mathbb{R}^{n \times n}$ is a symmetric matrix. Calculate $\nabla h(x)$.

(*b*) Consider the function $h : \mathbb{R}^n \to \mathbb{R}$ where $h(x) = \|Ax - b\|^2$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Calculate $\nabla h(x)$.

(*c*) Consider the 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

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

where $X \in \mathbb{R}^{n \times n}$ and has eigenvalues $\lambda_1, ..., \lambda_i$.

(*e*) Show that

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

where $X \in \mathbb{R}^{n \times n}$ and has eigenvalues $\lambda_1, ..., \lambda_i$.

Solution:

(*a*) Using product rule

$$\nabla_x (f(x)^T g(x)) = \nabla_x (f) g + \nabla_x (g) f$$

Let $f(x) = x$ and $g(x) = Qx$ and calculate gradient

$$\begin{aligned}
\nabla h(x) &= \frac{\partial f(x)}{\partial x}^T g + \frac{\partial g(x)}{\partial x}^T f \\
&= 1^T \cdot Qx + Q^T x \\
&= x(Q + Q^T) \qquad \text{Q symmetric so} Q^T = Q \\
&= x(Q + Q) = 2Qx
\end{aligned}$$

(*b*) Let $f(x) = Ax - b$ and $g(z) = \|z\|^2 = z^T z$. Then $Df(x) = A$ and $Dg(z) = 2z$.

$$Dh(x) = Dg(f(x)) = Dg(z)^T Df(x) = 2z^T A = 2(Ax - b)^T A$$

$$\nabla h(x) = Dh(x)^T = (2(Ax - b)^T A)^T = 2A^T (Ax - b)$$

(c)

(d)

$$\frac{\partial}{\partial X}\sum_{i=1}^{n}\lambda_i = \frac{\partial}{\partial X}tr(X) = tr(\frac{\partial}{\partial X}X) = tr(J) = I$$

(e)

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

Question 2 (15 points)

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

Solution:

$$\nabla_X tr(AX^TB) = tr(\nabla_X(AX^TB)) = tr(A(\nabla_X X)^TB) = tr(A(J^{ij})^TB) = tr(AJ^{ji}B) = BA$$

Question 3 (30 points)

Solve the following equality constrained optimization problem:

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

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

Solution: Form Lagrangian $\mathcal{L}(x,\lambda) = f(x) - \lambda(g(x))$ with $f(x) = x^TAx$ and $g(x) = b^Tx - 1$.

$$\mathcal{L}(x,\lambda) = x^TAx - \lambda(b^Tx - 1)$$

The gradient of Lagrangian has to be zero at x^*.

$$\nabla_x\mathcal{L}(x,\lambda) = \nabla_x(x^TAx - \lambda(b^Tx - 1)) = 2Ax - \lambda b = 0$$

This gives us $Ax = \frac{\lambda}{2}b$. Multiply with A^{-1} on the left gives

$$x = \frac{\lambda}{2}A^{-T}b$$