STAT3007/7007 Deep Learning, Tutorial 1
2022 Semester 1

This tutorial contains some warm-up exercises in calculus, linear algebra, statistics and programming. While you do not need to submit your answers, you are encouraged to work through the questions below, and fill in any knowledge gaps that you find.

Notations. Unless otherwise stated, a vector $\mathbf{x}$ denotes a column vector. The notation $(x_1, \ldots, x_d)$ denotes a row vector, and the notation $(x_1, \ldots, x_d)^\top$ or $\begin{pmatrix} x_1 \\ \ldots \\ x_d \end{pmatrix}$ denotes a column vector — we will usually use the former as it is more compact.

1. Consider the function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} + b^\top \mathbf{x}$, where $\mathbf{x} = (x_1, \ldots, x_d)^\top \in \mathbf{R}^d$, $A \in \mathbf{R}^{d\times d}$, and $b \in \mathbf{R}^d$.

- **(a)** What is the gradient $\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_d}\right)^\top$?
- **(b)** What is the Hessian $\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \ldots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_d} \\ & \ldots & \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_d \partial x_1} & \ldots & \frac{\partial^2 f(\mathbf{x})}{\partial x_d \partial x_d} \end{pmatrix}$?
- **(c)** If the Hessian of f is positive definite, that is, $\mathbf{x}^\top \nabla^2 f(\mathbf{x})\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbf{R}$, then f has a unique global minimum. Assume that $\nabla^2 f$ is positive definite, what is the minimizer of f? In addition, assume that A is symmetric, find the minimum of f.

2. Consider a function $f : \mathbf{R}^d \to \mathbf{R}$. Let $g(\mathbf{x}) = f(A\mathbf{x})$ for some matrix $A \in \mathbf{R}^{d\times d}$.

- **(a)** What is the relationship between $\nabla g(\mathbf{x})$ and $\nabla f(\mathbf{x})$?
- **(b)** If A is non-singular, show that if $\mathbf{x}$ is a stationary point of g, then $A\mathbf{x}$ is a stationary point of f.

3. Consider the matrix $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Write Python code to answer the following questions.

- **(a)** What are the eigenvalues and their corresponding eigenvectors for A?
- **(b)** What are A^2 and A^5?
- **(c)** Assume that $X \sim N(\mu, \Sigma)$, where $\mu = (1, 2)^\top$, and $\Sigma = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$, that is, X follows a multivariate normal distribution with mean μ and covariance Σ. Sample 100 observations of $Y = AX$ and draw a scatter plot for them. What is the mean of these observations? What is its difference with the true mean, that is, $\mathbb{E}\, Y$?