Vector and Matrix Calculus Review

Agenda

- Review of Partial Derivatives
- Gradient of a scalar valued function
- Hessian of a scalar valued function
- Gradient with respect to a matrix

Partial Derivatives

- For a bivariate function $f(x, y)$, the partial derivative with respect to x is defined as $\frac{\partial f(x, y)}{\partial x} = \lim_{h \to 0^+} \frac{f(x + h, y) - f(x, y)}{h}$. Also represented by $\nabla_x f(x, y)$.

- Second (partial) derivative $\frac{\partial^2 f}{\partial x^2}$ (also represented by $\nabla_x^2 f(x, y)$) is the partial derivative of $\frac{\partial f}{\partial x}$. Similarly, we can define $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)$ and $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$.

Example

- Let $f(x, y) = x^3y^2 + e^{x+y}$.

  - Compute $\dfrac{\partial^2 f}{\partial x \partial y}$ and $\dfrac{\partial^2 f}{\partial y \partial x}$.

- Turns out that $\dfrac{\partial^2 f}{\partial x \partial y} = \dfrac{\partial^2 f}{\partial y \partial x}$ is true more generally (when both derivatives exist and are continuous).

Gradient of a scalar valued function

- For a scalar function $f : \mathbb{R}^d \to \mathbb{R}$, the gradient of f (denoted by $\nabla f(x)$) is

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_d} \right]^\intercal \in \mathbb{R}^d.$$

- Some rules of derivatives apply for gradients as well (can be derived via scalar derivatives).
  - (Scalar Multiplication) Let $h(x) = t \cdot f(x)$. Then, $\nabla h(x) = t \cdot \nabla f(x)$.
  - (Addition) Let $h(x) = f(x) + g(x)$. Then, $\nabla h(x) = \nabla f(x) + \nabla g(x)$.

Gradient of a Linear Function

- Let $f(x) = w^\intercal x + b$ for some fixed $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$. Compute $\nabla f(x)$.

Gradient of a Quadratic Function

- Let $f(x) = \frac{1}{2}x^\intercal Ax + b^\intercal x + c$, where $A \in \mathbb{R}^{d\times d}$, $b \in \mathbb{R}^d$, and $c \in \mathbb{R}$. Compute $\nabla f(x)$.

Gradient of Logistic Loss

- Let $f(x) = \ln(1 + \exp(-w^{\intercal} x))$, where $w \in \mathbb{R}^d$.
  - Compute $\nabla f(x)$ by the definition of gradient.
  - Compute $\nabla f(x)$ by the Chain Rule.

Hessian of a scalar valued function

- For a scalar function $f : \mathbb{R}^d \to \mathbb{R}$, the Hessian of f (represented by $\nabla^2 f(x)$) is a matrix $\in \mathbb{R}^{d \times d}$ of which the (i, j)-th entry is given by $\dfrac{\partial^2 f}{\partial x_i \partial x_j}$.
- Rules:
  - Let $h(x) = t \cdot f(x)$. Then, $\nabla^2 h(x) = t \cdot \nabla^2 f(x)$.
  - Let $h(x) = f(x) + g(x)$. Then, $\nabla^2 h(x) = \nabla^2 f(x) + \nabla^2 g(x)$.

Hessian of a Quadratic Function

- Let $f(x) = \frac{1}{2}x^\intercal Ax + b^\intercal x + c$, where $A \in \mathbb{R}^{d \times d}$, $b \in \mathbb{R}^d$, and $c \in \mathbb{R}$.
    - Compute $\nabla^2 f(x)$.
    - Let $A \succeq 0$ (positive semi definite). Argue that f is convex.

Hessian of Logistic Loss

- Let $f(x) = \ln(1 + \exp(-w^\intercal x))$, where $w \in \mathbb{R}^d$.
  - Compute $\nabla^2 f(x)$.
  - Is f convex?

Gradient with respect to a matrix

- Let $f : \mathbb{R}^{n \times m} \to \mathbb{R}$ be a scalar valued function. Then, $\dfrac{\partial f(X)}{\partial X}$ is the matrix $\in \mathbb{R}^{n \times m}$ with (i, j)-th entry given by $\dfrac{\partial f}{\partial x_{ij}}$.

- Let $f(X) = \text{Trace}(X)$. Compute $\dfrac{\partial f}{\partial X}$.

- Refer to the Matrix Cookbook for more involved gradient and hessian computations.