Tutorial 1

Fundamentals

CS/SWE 4/6TE3 September 12, 2006

Review of derivatives, gradients and Hessians:

- The gradient extends the notion of derivative, the Hessian matrix – that of second derivative.

- Given a function f of n variables $x_1, x_2, \ldots, x_n$ we define the *partial derivative* relative to variable x_i, written as $\frac{\partial f}{\partial x_i}$, to be the derivative of f with respect to x_i treating all variables except x_i as constant. Let x denote the vector $(x_1, x_2, \ldots, x_n)^T$. With this notation, $f(x) = f(x_1, x_2, \ldots, x_n)$.

- The gradient of f at x, written as $\nabla f(x)$, is

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

- The gradient vector $\nabla f(x)$ gives the direction of steepest ascent of the function f at point x. The gradient acts like the derivative in that small changes around a given point x^* can be estimated using the gradient (see first-order Taylor series expansion).

- Second partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}$ are obtained from $f(x)$ by taking the derivative relative to x_i (this yields the first partial derivative $\frac{\partial f}{\partial x_i}$) and then by taking the derivative of $\frac{\partial f}{\partial x_i}$ relative to x_j. So, we can compute $\frac{\partial^2 f}{\partial x_1 \partial x_1} = \frac{\partial^2 f}{\partial x_1^2}$, $\frac{\partial^2 f}{\partial x_1 \partial x_2}$ and so on. This values are arranged into the *Hessian* matrix:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

The Hessian matrix is a symmetric matrix, that is $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$.

Computing gradients and Hessians:

Example

Compute the gradient and the Hessian of the function $f(x_1, x_2) = x_1^2 - 3x_1x_2 + x_2^2$ at the point $x = (x_1, x_2)^T = (1, 1)^T$.

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2x_1 - 3x_2 \\ -3x_1 + 2x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix}$$

Taylor series expansion:

Second-order Taylor series expansion:

$$f(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

First-order Taylor series expansion:

$$f(x) = f(x_0) + \nabla f(x_0)^T (x - x_0)$$

Example

$f(x_1, x_2) = x_1^2 - 3x_1x_2 + x_2^2$, compute $f(1.01, 1.01)$ using first- and second-order Taylor series expansion at the point $x_0 = (1, 1)^T$.

First-order Taylor series expansion:

$$f(1.01, 1.01) = f(1, 1) + \nabla f(1, 1)^T \begin{pmatrix} 1.01 - 1 \\ 1.01 - 1 \end{pmatrix} = -1 + (-1, -1) \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix} = -1.02$$

Second-order Taylor series expansion:

$$f(1.01, 1.01) = f(1, 1) + \nabla f(1, 1)^T \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix} + \frac{1}{2}(0.01, 0.01) \nabla^2 f(1, 1) \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix} =$$

$$= -1 + (-1, -1) \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix} + \frac{1}{2}(0.01, 0.01) \begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix} = -1.0201$$

Convex functions:

Definition A function f is convex if for any $x^1,\ x^2 \in C$ and $0 \le \lambda \le 1$

$$f(\lambda x^1 + (1-\lambda)x^2) \le \lambda f(x^1) + (1-\lambda)f(x^2).$$

A square matrix A said to be positive definite (PD) if $x^T A x > 0$ for all $x \neq 0$.

A square matrix A said to be positive semidefinite (PSD) if $x^T A x \ge 0$ for all x.

Hessian $\nabla f^2(x)$ is PD $\Longrightarrow$ strictly convex function.

Hessian $\nabla f^2(x)$ is PSD $\Longrightarrow$ convex function.

Gradient $\nabla f(\bar{x}) = 0$ and Hessian $\nabla f^2(\bar{x})$ is PSD $\Longrightarrow$ $\bar{x}$ is a minimum of the function f.

Gradient $\nabla f(\bar{x}) = 0$ and Hessian $\nabla f^2(\bar{x})$ is PD $\Longrightarrow$ $\bar{x}$ is a strict minimum of the function f.

Checking a matrix for PD and PSD:

Leading principal minors D_k, $k = 1, 2, \ldots, n$ of a matrix $A = (a_{ij})_{[n\times n]}$ are defined as

$$D_k = \det \begin{pmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \ldots & a_{kk} \end{pmatrix}$$

A square matrix A is PD $\Leftrightarrow$ $D_k > 0$ for all $k = 1, 2, \ldots, n$.

Example

Consider the function $f(x) = 3x_1^2 + 3x_2^2 + 5x_3^2 - 2x_1x_2$. The corresponding Hessian matrix is

$$\nabla^2 f(x) = 2 \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

Leading principal minors of $\nabla^2 f(x)$ are

$$D_1 = 2\cdot 3 = 6 > 0, \qquad D_2 = 2\cdot\det\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} = 2[3\cdot 3 - (-1)(-1)] = 2\cdot 8 = 16 > 0,$$

$$\begin{aligned} D_3 &= 2\cdot\det\begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} \\ &= 2([3\cdot 3\cdot 5 + 0\cdot 0\cdot(-1) + 0\cdot 0\cdot(-1)] - [0\cdot 0\cdot 3 + 0\cdot 0\cdot 3 + (-1)\cdot(-1)\cdot 5]) \\ &= 2\cdot 40 = 80 > 0 \end{aligned}$$

So, the Hessian is positive definite (PD) and the function is strictly convex.

A square matrix A is PSD $\Leftrightarrow$ all the principal minors of A are ≥ 0.

The *principal minor* is

$$\det \begin{pmatrix} a_{i_1 i_1} & \dots & a_{i_1 i_p} \\ \vdots & & \vdots \\ a_{i_p i_1} & \dots & a_{i_p i_p} \end{pmatrix}, \text{where } 1 \leq i_1 < i_2 < \ldots < i_p \leq n,\ p \leq n.$$