

Mathematics for Machine Learning

— Vector Calculus: Linearization & Multivariate Taylor Series

Joseph Chuang-Chieh Lin

Department of Computer Science & Information Engineering,
Tamkang University

Fall 2023

Credits for the resource

- The slides are based on the textbooks:
 - *Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong: Mathematics for Machine Learning. Cambridge University Press. 2020.*
 - *Howard Anton, Chris Rorres, Anton Kaul: Elementary Linear Algebra. Wiley. 2019.*
- We could partially refer to the monograph:
Francesco Orabona: A Modern Introduction to Online Learning.
<https://arxiv.org/abs/1912.13213>

Linear Approximation of a Function

The gradient ∇f of a function f can be used for locally linear approximation of f around \mathbf{x}_0 :

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + (\nabla_{\mathbf{x}} f)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$$

- $(\nabla_{\mathbf{x}} f)(\mathbf{x}_0)$: the gradient of f w.r.t. \mathbf{x} evaluated at \mathbf{x}_0 .

Multivariate Taylor Series

Multivariate Taylor Series

Consider a function $f : \mathbb{R}^D \mapsto \mathbb{R}$ which is smooth (i.e., infinitely differentiable) at \mathbf{x}_0 .

Define the *different vector* $\boldsymbol{\delta} := \mathbf{x} - \mathbf{x}_0$.

The **multivariate Taylor series** of f at \mathbf{x}_0 is

$$f(\mathbf{x}) = \sum_{k=0}^{\infty} \frac{D_{\mathbf{x}}^k f(\mathbf{x}_0)}{k!} \boldsymbol{\delta}^k,$$

where $D_{\mathbf{x}}^k f(\mathbf{x}_0)$ is the k th derivative of f w.r.t. \mathbf{x} evaluated at \mathbf{x}_0 .

Multivariate Taylor Polynomial

Multivariate Taylor Polynomial

The **Taylor polynomial** of **degree n** of f at \mathbf{x}_0 is

$$T_n(\mathbf{x}) = \sum_{k=0}^n \frac{D_{\mathbf{x}}^k f(\mathbf{x}_0)}{k!} \delta^k,$$

where $D_{\mathbf{x}}^k f(\mathbf{x}_0)$ is the k th derivative of f w.r.t. \mathbf{x} evaluated at \mathbf{x}_0 .

- It contains the first $n + 1$ components of the Taylor series.

Notice

- δ^k is not defined for $\mathbf{x} \in \mathbb{R}^D$, $D > 1$ and $k > 1$.

Notice

- δ^k is not defined for $\mathbf{x} \in \mathbb{R}^D$, $D > 1$ and $k > 1$.
- $D_{\mathbf{x}}^k$ and δ^k are k th order tensors (i.e., k -dimensional arrays).

Notice

- δ^k is not defined for $\mathbf{x} \in \mathbb{R}^D$, $D > 1$ and $k > 1$.
- $D_{\mathbf{x}}^k$ and δ^k are k th order tensors (i.e., k -dimensional arrays).
- $\delta^k \in \mathbb{R}^{\overbrace{D \times D \times \cdots \times D}^{k \text{ times}}}$.

Notice

- δ^k is not defined for $\mathbf{x} \in \mathbb{R}^D$, $D > 1$ and $k > 1$.
- $D_{\mathbf{x}}^k$ and δ^k are k th order tensors (i.e., k -dimensional arrays).

- $\delta^k \in \mathbb{R}^{\overbrace{D \times D \times \cdots \times D}^{k \text{ times}}}$.
 - $\delta^2 := \delta \otimes \delta = \delta \delta^\top$.
 - $\delta^2[i, j] = \delta[i] \delta[j]$.
 - $\delta^3 := \delta \otimes \delta \otimes \delta$.
 - $\delta^3[i, j, k] = \delta[i] \delta[j] \delta[k]$.

Notice

- δ^k is not defined for $\mathbf{x} \in \mathbb{R}^D$, $D > 1$ and $k > 1$.
- $D_{\mathbf{x}}^k$ and δ^k are k th order tensors (i.e., k -dimensional arrays).

- $\delta^k \in \mathbb{R}^{\overbrace{D \times D \times \cdots \times D}^{k \text{ times}}}$.

- $\delta^2 := \delta \otimes \delta = \delta \delta^\top$.

- $\delta^2[i, j] = \delta[i] \delta[j]$.

- $\delta^3 := \delta \otimes \delta \otimes \delta$.

- $\delta^3[i, j, k] = \delta[i] \delta[j] \delta[k]$.

- Hence,

$$D_{\mathbf{x}}^k f(\mathbf{x}_0) \delta^k = \sum_{i_1=1}^D \cdots \sum_{i_k=1}^D D_{\mathbf{x}}^k f(\mathbf{x}_0)[i_1, \dots, i_k] \delta[i_1] \cdots \delta[i_k].$$

Note & Exercise

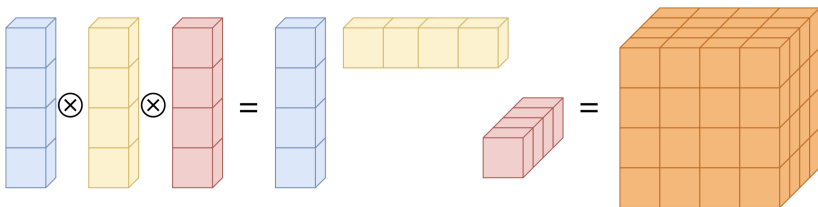
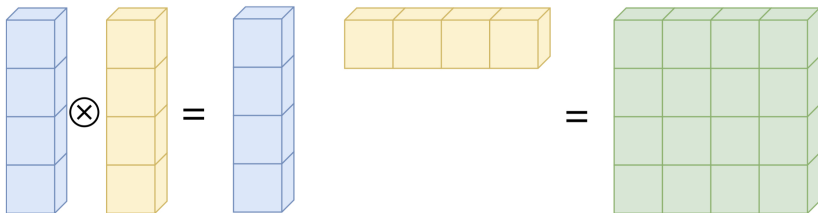
Exercise

Show that

$$D_{\mathbf{x}}^2 f(\mathbf{x}_0) \delta^2 = \delta^\top \mathbf{H}(\mathbf{x}_0) \delta,$$

where

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}.$$

δ^2 & δ^3 

Example (1/3)

Example

Consider the function $f(x, y) = x^2 + 2xy + y^3$ and $(x_0, y_0) = (1, 2)$.

- Note: f is a polynomial of degree 3.

$$f(1, 2) = 13, \quad \delta = [x - 1, y - 2]^T.$$

$$\frac{\partial f}{\partial x} = 2x + 2y \implies \frac{\partial f}{\partial x}(1, 2) = 6.$$

$$\frac{\partial f}{\partial y} = 2x + 3y^2 \implies \frac{\partial f}{\partial y}(1, 2) = 14.$$

Example (1/3)

Example

Consider the function $f(x, y) = x^2 + 2xy + y^3$ and $(x_0, y_0) = (1, 2)$.

- Note: f is a polynomial of degree 3.

$$f(1, 2) = 13, \quad \delta = [x - 1, y - 2]^\top.$$

$$\frac{\partial f}{\partial x} = 2x + 2y \implies \frac{\partial f}{\partial x}(1, 2) = 6.$$

$$\frac{\partial f}{\partial y} = 2x + 3y^2 \implies \frac{\partial f}{\partial y}(1, 2) = 14.$$

$$\therefore D_{x,y}^1 f(1, 2) = \nabla_{x,y} f(1, 2) = \left[\frac{\partial f}{\partial x}(1, 2) \quad \frac{\partial f}{\partial y}(1, 2) \right] = [6 \quad 14] \in \mathbb{R}^{1 \times 2}.$$

Example (1/3)

Example

Consider the function $f(x, y) = x^2 + 2xy + y^3$ and $(x_0, y_0) = (1, 2)$.

- Note: f is a polynomial of degree 3.

$$f(1, 2) = 13, \quad \delta = [x - 1, y - 2]^\top.$$

$$\frac{\partial f}{\partial x} = 2x + 2y \implies \frac{\partial f}{\partial x}(1, 2) = 6.$$

$$\frac{\partial f}{\partial y} = 2x + 3y^2 \implies \frac{\partial f}{\partial y}(1, 2) = 14.$$

$$\therefore D_{x,y}^1 f(1, 2) = \nabla_{x,y} f(1, 2) = \begin{bmatrix} \frac{\partial f}{\partial x}(1, 2) & \frac{\partial f}{\partial y}(1, 2) \end{bmatrix} = [6 \quad 14] \in \mathbb{R}^{1 \times 2}.$$

$$\implies \frac{D_{x,y}^1 f(1, 2)}{1!} \delta = [6 \quad 14] \begin{bmatrix} x - 1 \\ y - 2 \end{bmatrix} = 6(x - 1) + 14(y - 2).$$

Example (2/3)

Example

$$\frac{\partial^2 f}{\partial x^2} = 2 \implies \frac{\partial^2 f}{\partial x^2}(1, 2) = 2$$

$$\frac{\partial^2 f}{\partial y^2} = 6y \implies \frac{\partial^2 f}{\partial y^2}(1, 2) = 12$$

$$\frac{\partial^2 f}{\partial y \partial x} = 2 \implies \frac{\partial^2 f}{\partial y \partial x}(1, 2) = 2$$

$$\frac{\partial^2 f}{\partial x \partial y} = 2 \implies \frac{\partial^2 f}{\partial x \partial y}(1, 2) = 2$$

Example (2/3)

Example

$$\frac{\partial^2 f}{\partial x^2} = 2 \implies \frac{\partial^2 f}{\partial x^2}(1, 2) = 2$$

$$\frac{\partial^2 f}{\partial y^2} = 6y \implies \frac{\partial^2 f}{\partial y^2}(1, 2) = 12$$

$$\frac{\partial^2 f}{\partial y \partial x} = 2 \implies \frac{\partial^2 f}{\partial y \partial x}(1, 2) = 2$$

$$\frac{\partial^2 f}{\partial x \partial y} = 2 \implies \frac{\partial^2 f}{\partial x \partial y}(1, 2) = 2$$

Hessian:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 6y \end{bmatrix}$$

Example (2/3)

Example

$$\frac{\partial^2 f}{\partial x^2} = 2 \implies \frac{\partial^2 f}{\partial x^2}(1, 2) = 2$$

$$\frac{\partial^2 f}{\partial y^2} = 6y \implies \frac{\partial^2 f}{\partial y^2}(1, 2) = 12$$

$$\frac{\partial^2 f}{\partial y \partial x} = 2 \implies \frac{\partial^2 f}{\partial y \partial x}(1, 2) = 2$$

$$\frac{\partial^2 f}{\partial x \partial y} = 2 \implies \frac{\partial^2 f}{\partial x \partial y}(1, 2) = 2$$

Hessian:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 6y \end{bmatrix}$$

$$\implies \mathbf{H}(1, 2) = \begin{bmatrix} 2 & 2 \\ 2 & 12 \end{bmatrix} \in \mathbb{R}^{2 \times 2}.$$

Example (2/3)

Example

$$\frac{\partial^2 f}{\partial x^2} = 2 \implies \frac{\partial^2 f}{\partial x^2}(1, 2) = 2$$

$$\frac{\partial^2 f}{\partial y^2} = 6y \implies \frac{\partial^2 f}{\partial y^2}(1, 2) = 12$$

$$\frac{\partial^2 f}{\partial y \partial x} = 2 \implies \frac{\partial^2 f}{\partial y \partial x}(1, 2) = 2$$

$$\frac{\partial^2 f}{\partial x \partial y} = 2 \implies \frac{\partial^2 f}{\partial x \partial y}(1, 2) = 2$$

Hessian:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 6y \end{bmatrix}$$

$$\implies \mathbf{H}(1, 2) = \begin{bmatrix} 2 & 2 \\ 2 & 12 \end{bmatrix} \in \mathbb{R}^{2 \times 2}.$$

$$\begin{aligned} \frac{D_{x,y}^2 f(1, 2)}{2!} \delta^2 &= \frac{1}{2} \delta^\top \mathbf{H}(1, 2) \delta \\ &= \dots \end{aligned}$$

Example (2/3)

Example

Hessian:

$$\frac{\partial^2 f}{\partial x^2} = 2 \implies \frac{\partial^2 f}{\partial x^2}(1, 2) = 2$$

$$\frac{\partial^2 f}{\partial y^2} = 6y \implies \frac{\partial^2 f}{\partial y^2}(1, 2) = 12$$

$$\frac{\partial^2 f}{\partial y \partial x} = 2 \implies \frac{\partial^2 f}{\partial y \partial x}(1, 2) = 2$$

$$\frac{\partial^2 f}{\partial x \partial y} = 2 \implies \frac{\partial^2 f}{\partial x \partial y}(1, 2) = 2$$

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 6y \end{bmatrix}$$

$$\implies \mathbf{H}(1, 2) = \begin{bmatrix} 2 & 2 \\ 2 & 12 \end{bmatrix} \in \mathbb{R}^{2 \times 2}.$$

$$\begin{aligned} \frac{D_{x,y}^2 f(1, 2)}{2!} \delta^2 &= \frac{1}{2} \delta^\top \mathbf{H}(1, 2) \delta \\ &= \cdots = (x-1)^2 + 2(x-1)(y-2) + 6(y-2)^2. \end{aligned}$$

Example (3/3)

Now, compute

$$D_{x,y}^3 f =$$

Example (3/3)

Now, compute

$$D_{x,y}^3 f = \begin{bmatrix} \frac{\partial \mathbf{H}}{\partial x} & \frac{\partial \mathbf{H}}{\partial y} \end{bmatrix} \in \mathbb{R}^{2 \times 2 \times 2}$$

Example (3/3)

Now, compute

$$D_{x,y}^3 f = \begin{bmatrix} \frac{\partial \mathbf{H}}{\partial x} & \frac{\partial \mathbf{H}}{\partial y} \end{bmatrix} \in \mathbb{R}^{2 \times 2 \times 2}$$

$$\frac{\partial \mathbf{H}}{\partial x} = \begin{bmatrix} \frac{\partial^3 f}{\partial x^3} & \frac{\partial^3 f}{\partial x^2 \partial y} \\ \frac{\partial^3 f}{\partial x \partial y \partial z} & \frac{\partial^3 f}{\partial x \partial y^2} \end{bmatrix}, \quad \frac{\partial \mathbf{H}}{\partial y} = \begin{bmatrix} \frac{\partial^3 f}{\partial y \partial x^2} & \frac{\partial^3 f}{\partial y \partial x \partial y} \\ \frac{\partial^3 f}{\partial y^2 \partial x} & \frac{\partial^3 f}{\partial y^3} \end{bmatrix}$$

Example (3/3)

Now, compute

$$D_{x,y}^3 f = \begin{bmatrix} \frac{\partial \mathbf{H}}{\partial x} & \frac{\partial \mathbf{H}}{\partial y} \end{bmatrix} \in \mathbb{R}^{2 \times 2 \times 2}$$

$$\frac{\partial \mathbf{H}}{\partial x} = \begin{bmatrix} \frac{\partial^3 f}{\partial x^3} & \frac{\partial^3 f}{\partial x^2 \partial y} \\ \frac{\partial^3 f}{\partial x \partial y \partial z} & \frac{\partial^3 f}{\partial x \partial y^2} \end{bmatrix}, \quad \frac{\partial \mathbf{H}}{\partial y} = \begin{bmatrix} \frac{\partial^3 f}{\partial y \partial x^2} & \frac{\partial^3 f}{\partial y \partial x \partial y} \\ \frac{\partial^3 f}{\partial y^2 \partial x} & \frac{\partial^3 f}{\partial y^3} \end{bmatrix}$$

We only need to compute $\frac{\partial^3 f}{\partial y^3} = 6 \implies \frac{\partial^3 f}{\partial y^3}(1, 2) = 6$.

Example (3/3)

Now, compute

$$D_{x,y}^3 f = \begin{bmatrix} \frac{\partial \mathbf{H}}{\partial x} & \frac{\partial \mathbf{H}}{\partial y} \end{bmatrix} \in \mathbb{R}^{2 \times 2 \times 2}$$

$$\frac{\partial \mathbf{H}}{\partial x} = \begin{bmatrix} \frac{\partial^3 f}{\partial x^3} & \frac{\partial^3 f}{\partial x^2 \partial y} \\ \frac{\partial^3 f}{\partial x \partial y \partial z} & \frac{\partial^3 f}{\partial x \partial y^2} \end{bmatrix}, \quad \frac{\partial \mathbf{H}}{\partial y} = \begin{bmatrix} \frac{\partial^3 f}{\partial y \partial x^2} & \frac{\partial^3 f}{\partial y \partial x \partial y} \\ \frac{\partial^3 f}{\partial y^2 \partial x} & \frac{\partial^3 f}{\partial y^3} \end{bmatrix}$$

We only need to compute $\frac{\partial^3 f}{\partial y^3} = 6 \implies \frac{\partial^3 f}{\partial y^3}(1, 2) = 6$.

$$\frac{D_{x,y}^3 f(1, 2)}{3!} \delta^3 = (y - 2)^3.$$

Example (3/3)

Now, compute

$$D_{x,y}^3 f = \left[\frac{\partial \mathbf{H}}{\partial x} \quad \frac{\partial \mathbf{H}}{\partial y} \right] \in \mathbb{R}^{2 \times 2 \times 2}$$

$$\frac{\partial \mathbf{H}}{\partial x} = \begin{bmatrix} \frac{\partial^3 f}{\partial x^3} & \frac{\partial^3 f}{\partial x^2 \partial y} \\ \frac{\partial^3 f}{\partial x \partial y \partial z} & \frac{\partial^3 f}{\partial x \partial y^2} \end{bmatrix}, \quad \frac{\partial \mathbf{H}}{\partial y} = \begin{bmatrix} \frac{\partial^3 f}{\partial y \partial x^2} & \frac{\partial^3 f}{\partial y \partial x \partial y} \\ \frac{\partial^3 f}{\partial y^2 \partial x} & \frac{\partial^3 f}{\partial y^3} \end{bmatrix}$$

We only need to compute $\frac{\partial^3 f}{\partial y^3} = 6 \implies \frac{\partial^3 f}{\partial y^3}(1, 2) = 6$.

$$\frac{D_{x,y}^3 f(1, 2)}{3!} \delta^3 = (y - 2)^3.$$

Check if $f(x) = f(1, 2) + D_{x,y}^1 f(1, 2) \delta + \frac{D_{x,y}^2 f(1, 2)}{2!} \delta^2 + \frac{D_{x,y}^3 f(1, 2)}{3!} \delta^3$.

Discussions