Machine Learning: Multivariate Calculus

Hao Tang

January 24, 2024
The derivative of a function $f : \mathbb{R} \rightarrow \mathbb{R}$ at $x_0$ is defined as

$$(D_x f)(x_0) = \left( \frac{d}{dx} f \right)(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$  

(1)
Derivative as linear approximation

\[ f(x_0 + h) = f(x_0) + (D_x f)(x_0)h + O(h^2) \]
Derivative as linear approximation
Derivative as linear approximation

\[ f(x_0 + h) = f(x_0) + (D_x f)(x_0)h \]
Consider the line $g_{x_0}(x) = f(x_0) + (D_x f)(x_0)(x - x_0)$.

The term $E(x) = |f(x) - g_{x_0}(x)|$ defines the vertical distance between the line and the function.

Think of approximating the function with the line, and $E(x)$ tells us how bad this approximation is.

The error has to become small as we get close to $x_0$, i.e.,

$$\lim_{x \to x_0} \frac{E(x)}{x - x_0} = 0.$$  \hspace{1cm} (2)
Derivative as linear approximation

\[ E(x) = \frac{|f(x) - g_{x_0}(x)|}{x - x_0} (x - x_0) = \frac{|f(x) - f(x_0) - (D_x f)(x_0)(x - x_0)|}{x - x_0} (x - x_0) \] (3)

\[ = \left| \frac{f(x) - f(x_0)}{x - x_0} - (D_x f)(x_0) \right| (x - x_0) \] (4)
Suppose we have a function $T$ that is linear. Show that if

$$\lim_{h \to 0} \frac{f(x + h) - [f(x) + T(x)h]}{h} = 0$$

for all $x$, we have

$$T(x) = (D_x f)(x)$$

for all $x$. 
Directional derivative

The directional derivative of $f : \mathbb{R}^d \to \mathbb{R}$ along the direction $v$ at $x_0 \in \mathbb{R}^d$ is defined as

$$(D_v f)(x_0) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t}. \quad (7)$$
The directional derivative of $f : \mathbb{R}^d \to \mathbb{R}$ along the direction $v$ at $x_0 \in \mathbb{R}^d$ is defined as

$$(D_v f)(x_0) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t}. \quad (7)$$

If we let $g_{x_0}(t) = f(x_0 + tv)$, then

$$(D_t g)(0) = \lim_{t \to 0} \frac{g(0 + t) - g(0)}{t} = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t} = (D_v f)(x_0) \quad (8)$$
Example

- Consider the function $f(x, y) = x^2 - y^2$.

- If we are at $(2, 0)$, the directional derivative along $(1, 0)$ is 4.
• Consider the function $f(x, y) = x^2 - y^2$.

• If we are at $(2, 0)$, the directional derivative along $(1, 0)$ is 4.

• If we take a line at $\{(x, y) : (x, y) = (2, 0) + t(1, 0) = (2 + t, 0) \text{ for } t \in \mathbb{R}\}$, we have $g(t) = f(2 + t, 0) = (2 + t)^2$. The derivative $(D_t g)(t) = 2(2 + t)$, and $(D_t g)(0) = 2 \cdot (2 + 0) = 4$. 
Partial derivatives

• A partial derivative is a directional derivative along the direction of coordinate axes.
Partial derivatives

• A partial derivative is a directional derivative along the direction of coordinate axes.

• In a three-dimensional space, the direction of the axes are

\[ (1, 0, 0) \quad (0, 1, 0) \quad (0, 0, 1). \] \hspace{1cm} (9)

For a function \( f : \mathbb{R}^3 \rightarrow \mathbb{R} \), the partial derivatives along the axes are

\[ \frac{\partial}{\partial x} f \quad \frac{\partial}{\partial y} f \quad \frac{\partial}{\partial z} f. \] \hspace{1cm} (10)
Example

• Given a function $f(x, y) = x^2 - y^2$, show that

$$
\left( \frac{\partial}{\partial x} f \right)(x, y) = 2x \quad \left( \frac{\partial}{\partial y} f \right)(x, y) = -2y.
$$

(11)
Example

• Given a function \( f(x, y) = x^2 - y^2 \), show that

\[
\left( \frac{\partial f}{\partial x} \right)(x, y) = 2x \\
\left( \frac{\partial f}{\partial y} \right)(x, y) = -2y.
\]  

(11)

• The \( x \)-axis is the direction \((1, 0)\). At any point \((x, y)\), the line along that direction is \((x + t, y)\). The function value along that line is \( g(t) = f(x + t, y) = (x + t)^2 - y^2 \). We then have \( (D_t g)(t) = 2(x + t) \), and

\[
\left( \frac{\partial f}{\partial x} \right)(x, y) = (D_t g)(0) = 2x.
\]  

(12)
Example

• Given a function \( f(x, y) = x^2 - y^2 \), show that

\[
\left( \frac{\partial}{\partial x} f \right)(x, y) = 2x \quad \left( \frac{\partial}{\partial y} f \right)(x, y) = -2y.
\]

(11)

• The \( x \)-axis is the direction \((1, 0)\). At any point \((x, y)\), the line along that direction is \((x + t, y)\). The function value along that line is \( g(t) = f(x + t, y) = (x + t)^2 - y^2 \). We then have \((D_t g)(t) = 2(x + t)\), and

\[
\left( \frac{\partial}{\partial x} f \right)(x, y) = (D_t g)(0) = 2x.
\]

(12)

• Treat other variables as constants and take 1D derivatives.
Example

- Given a function $f(x, y, z) = (x + 2y - 3z)^2$, show that

\[
\frac{\partial}{\partial x} f(x, y, z) = 2(x + 2y - 3z) \quad \text{(13)}
\]

\[
\frac{\partial}{\partial y} f(x, y, z) = 2(x + 2y - 3z) \cdot 2 \quad \text{(14)}
\]

\[
\frac{\partial}{\partial z} f(x, y, z) = 2(x + 2y - 3z) \cdot (-3) \quad \text{(15)}
\]
Example

- Given a function

\[ f(w, b) = \frac{1}{1 + \exp(-(\mathbf{w}^\top \mathbf{x} + b))}, \quad (16) \]

show that

\[ \left( \frac{\partial}{\partial b} f \right)(w, b) = f(w, b)(1 - f(w, b)). \quad (17) \]
Gradients

- The gradient of a function is the vector consisting of all partial derivatives.
The gradient of a function is the vector consisting of all partial derivatives.

For a function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$, its gradient is

$$(\nabla f)(x, y, z) = \begin{bmatrix} \frac{\partial f}{\partial x}(x, y, z) \\ \frac{\partial f}{\partial y}(x, y, z) \\ \frac{\partial f}{\partial z}(x, y, z) \end{bmatrix}. \quad (18)$$
Example

• Given a function $f(x, y, z) = (x + 2y - 3z)^2$, show that its gradient is

$$\nabla f(x, y, z) = \begin{bmatrix}
2(x + 2y - 3z) \\
2(x + 2y - 3z) \cdot 2 \\
2(x + 2y - 3z) \cdot (-3)
\end{bmatrix}.$$  

(19)
Example

• Given a function $f(a) = b^\top a$, show that its gradient is

$$(\nabla f)(a) = b.$$  \hfill (20)
Example

• Given a function \( f(a) = b^\top a \), show that its gradient is

\[
(\nabla f)(a) = b.
\]

(20)

• Given a function \( f(a) = b^\top Aa \), show that its gradient is

\[
(\nabla f)(a) = A^\top b.
\]

(21)
Example

- Given a function $f(a) = \|a\|^2_2$, show that its gradient is

$$\nabla f(a) = 2a.$$  \hfill (22)
Example

Given a function $f(w) = (w^T x + b - y)^2$, show that

$$(\nabla f)(w) = 2(w^T x + b - y)x.$$  \hspace{1cm} (23)
Example

- Given a function

\[ f(w) = \frac{1}{1 + \exp(-(w^\top x + b))}, \tag{24} \]

show that its gradient is

\[ (\nabla f)(w) = f(w)(1 - f(w))x. \tag{25} \]
Theorem

• For a function $f : \mathbb{R}^d \to \mathbb{R}$ and any direction $\nu$ at any point $x$, show that

$$ (D_{\nu} f)(x) = (\nabla f)(x)^{\top} \nu. $$

(26)

• Once we know the gradient, we know all directional derivatives.
Second-order derivative

For a function $f : \mathbb{R} \rightarrow \mathbb{R}$, its second-order derivative is defined and written as

$$
\frac{\partial^2}{\partial x^2} f = \frac{\partial}{\partial x} \left( \frac{\partial}{\partial x} f \right).
$$

(27)
Example

• Given a function $f(x) = x^2$, its second-order derivative is 2.
Example

• Given a function $f(x) = x^2$, it’s second-order derivative is 2.

• The second-order derivative tells us whether the function looks like a cup or an upside-down cup.
The Hessian of a function $f : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$
\begin{bmatrix}
\frac{\partial^2}{\partial x_1 \partial x_1} f & \frac{\partial^2}{\partial x_1 \partial x_2} f & \cdots & \frac{\partial^2}{\partial x_1 \partial x_d} f \\
\frac{\partial^2}{\partial x_2 \partial x_1} f & \frac{\partial^2}{\partial x_2 \partial x_2} f & \cdots & \frac{\partial^2}{\partial x_2 \partial x_d} f \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2}{\partial x_d \partial x_1} f & \frac{\partial^2}{\partial x_d \partial x_2} f & \cdots & \frac{\partial^2}{\partial x_d \partial x_d} f 
\end{bmatrix}.
$$

(28)

Because

$$
\frac{\partial^2}{\partial x_j \partial x_i} f = \frac{\partial^2}{\partial x_i \partial x_j} f,
$$

(29)

the Hessian matrix is always symmetric.
Example

• Given a function \( f(x, y) = x^2 - y^2 \), show that its Hessian is \[
\begin{bmatrix}
2 & 0 \\
0 & -2
\end{bmatrix}.
\]