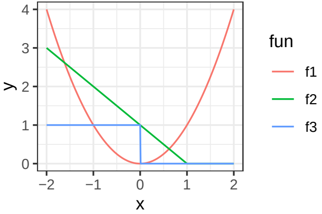
Optimization

Quadratic forms II

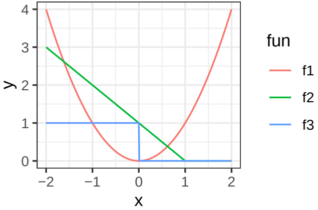


Learning goals

- TODO
- TODO

SMOOTH VS. NON-SMOOTH

- The **smoothness** of a function is a property that is measured by the number of its continuous derivatives.
- We will call a function $f : \mathcal{S} \to \mathbb{R}$ "smooth", if it is at least differentiable for every $\mathbf{x} \in \mathcal{S}$.
- We call a function k-times continuously differentiable, if the k-th derivative exists and is continuous. $\mathcal{C}^k$ denotes the class of k-times continuously differentiable functions.



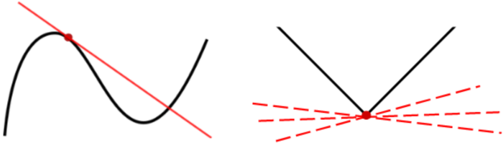
f_1 is smooth, f_2 is continuous but not differentiable, and f_3 is non-continuous.

DIFFERENTIABILITY (UNIVARIATE)

Definition 1: A function $f : \mathcal{S} \subseteq \mathbb{R} \to \mathbb{R}$ is said to be differentiable in $x \in \mathcal{S}$ if the following limit exists:

$$f'(x) := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

A function $f : \mathcal{S} \subseteq \mathbb{R} \to \mathbb{R}$ is said to be differentiable in $x \in \mathcal{S}$, if f can be locally approximated by a linear function in x.



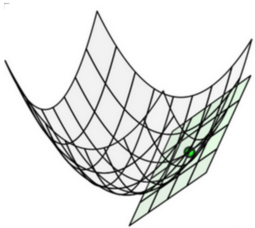
Geometrically: A tangent can be placed on the graph of f through the point $(x, f(x))$. $m = f'(x)$ then indicates the slope of this tangent. The function on the left is differentiable everywhere; the function on the right is not differentiable at the red point.

DIFFERENTIATION (MULTIVARIATE)

A similar definition of differentiability holds for multivariate functions.

Definition: A function $f : \mathcal{S} \subseteq \mathbb{R}^d \to \mathbb{R}$ is differentiable in $\mathbf{x} \in \mathcal{S}$ if there exists a (continuous) linear map $\nabla f(\mathbf{x}) : \mathcal{S} \subseteq \mathbb{R}^d \to \mathbb{R}$ with

$$\lim_{\boldsymbol{h} \to 0} \frac{f(\mathbf{x} + \boldsymbol{h}) - f(\mathbf{x}) - \nabla f(\mathbf{x}) \cdot \boldsymbol{h}}{||\boldsymbol{h}||} = 0$$



Geometrically: The function can be locally approximated by a tangent hyperplane.

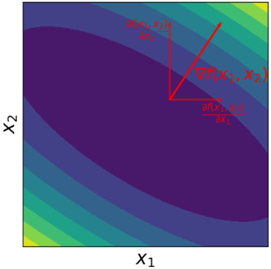
Source: `https://github.com/jermwatt/machine_learning_refined`.

DIFFERENTIATION (MULTIVARIATE)

This local linear approximation is described by the **gradient**: If f is differentiable in $\mathbf{x}$, the **gradient** is defined by

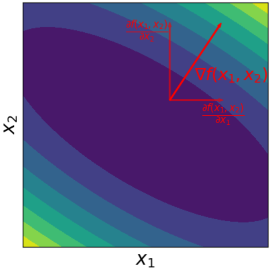
$$\nabla f = \frac{\partial f}{\partial x_1}\boldsymbol{e}_1 + \cdots + \frac{\partial f}{\partial x_n}\boldsymbol{e}_n = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)^\top .$$

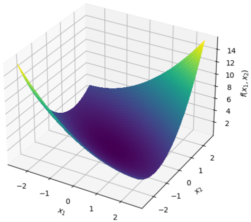
The elements of the gradient are called **partial derivatives**.



DIFFERENTIATION (MULTIVARIATE)

Consider $f(\mathbf{x}) = 0.5x_1^2 + x_2^2 + x_1x_2$. The gradient is $\nabla f(\mathbf{x}) = (x_1 + x_2, 2x_2 + x_1)^\top$.

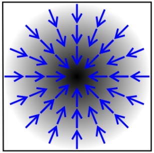


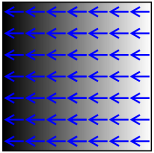


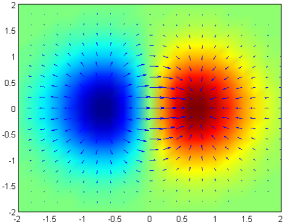
DIFFERENTIATION (MULTIVARIATE)

Properties of the gradient:

- The gradient is orthogonal to level curves and level surfaces of a function
- The gradient points in the direction of greatest increase of f







- The normal vector describing the tangent plane has $n+1$ components, the first n correspond to ∇f and the $(n+1)-$th has the value -1

DIFFERENTIATION (MULTIVARIATE)

We can also compute the instantaneous rate of change of f at $\mathbf{x}$ along an arbitrary direction $\boldsymbol{v}$:

$$D_v f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \boldsymbol{v}.$$

$D_v f(\mathbf{x})$ is called **directional derivative.**
Definition: The directional derivative for direction $\boldsymbol{v}$ for $f : \mathcal{S} \to \mathbb{R}, \mathcal{S} \subseteq \mathbb{R}^d$ is defined as

$$D_v f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\boldsymbol{v}) - f(\mathbf{x})}{h}.$$

For example, the slope in the direction $\boldsymbol{v} = (1, 1)$ is the sum of the first and the second partial derivative:

$$D_v f(\mathbf{x}) = \nabla f(\mathbf{x})^\top \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}$$

DIFFERENTIATION (MULTIVARIATE)

Definition (Hessian): The **Hessian matrix** is analogous to the second derivative in a multivariate setting. The Hessian matrix consists of the second partial derivatives:

$$H(\boldsymbol{x}) = \nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}\right)_{i,j=1\ldots d}$$

- The Hessian indicates the local curvature (2nd derivative) at a point $\mathbf{x}$ of the function f.
- The eigenvector corresponding to the largest absolute eigenvalue indicates the direction of the strongest curvature.
- The eigenvector corresponding to the smallest absolute eigenvalue indicates the direction of the lowest curvature.
- The corresponding eigenvalues specify the strength of the curvature.