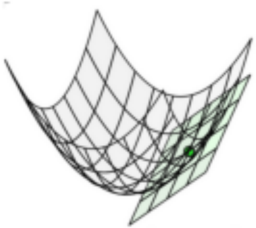
MULTIVARIATE DIFFERENTIABILITY

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

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



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

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

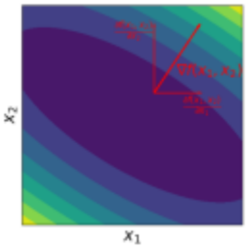
GRADIENT

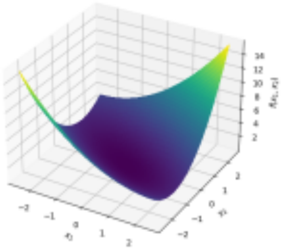
- Linear approximation is given by the **gradient**:

$$\nabla f = \frac{\partial f}{\partial x_1}\mathbf{e}_1 + \cdots + \frac{\partial f}{\partial x_d}\mathbf{e}_d = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_d}\right)^T$$

- Elements of the gradient are called **partial derivatives**.
- To compute $\partial f / \partial x_j$, regard f as function of x_j only (others fixed)

Example: $f(\mathbf{x}) = x_1^2/2 + x_1 x_2 + x_2^2 \Rightarrow \nabla f(\mathbf{x}) = (x_1 + x_2, x_1 + 2x_2)^T$





LOCAL CURVATURE BY HESSIAN

Eigenvector corresponding to largest (resp. smallest) **eigenvalue** of Hessian points in direction of largest (resp. smallest) **curvature**

Example (previous slide): For $\mathbf{a} = (-\pi/2, 0)^T$, we have

$$H(\mathbf{a}) = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$$

and thus $\lambda_1 = 4$, $\lambda_2 = 1$, $\mathbf{v}_1 = (0, 1)^T$, and $\mathbf{v}_2 = (1, 0)^T$.