

# Mathematics for Machine Learning

## Multivariate Calculus

### *Formula sheet*

Dr Samuel J. Cooper  
Prof. David Dye  
Dr A. Freddie Page

#### Definition of a derivative

$$f'(x) = \frac{df(x)}{dx} = \lim_{\Delta x \rightarrow 0} \left( \frac{f(x + \Delta x) - f(x)}{\Delta x} \right)$$

#### Time saving rules

- *Sum Rule:*

$$\frac{d}{dx} (f(x) + g(x)) = \frac{d}{dx} (f(x)) + \frac{d}{dx} (g(x))$$

- *Power Rule:*

$$f(x) = ax^b$$
$$f'(x) = abx^{(b-1)}$$

- *Product Rule:*

$$A(x) = f(x)g(x)$$
$$A'(x) = f'(x)g(x) + f(x)g'(x)$$

- *Chain Rule:*

If  $h = h(p)$  and  $p = p(m)$

then  $\frac{dh}{dm} = \frac{dh}{dp} \times \frac{dp}{dm}$

- *Total derivative:*

For the function  $f(x, y, z, \dots)$ , where each variable is a function of parameter  $t$ , the total derivative is

$$\frac{df}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial z} \frac{dz}{dt} + \dots$$

#### Derivatives of named functions

$$\frac{\partial}{\partial x} \frac{1}{x} = -\frac{1}{x^2} \quad (1)$$

$$\frac{\partial}{\partial x} \sin x = \cos x \quad (2)$$

$$\frac{\partial}{\partial x} \cos x = -\sin x \quad (3)$$

$$\frac{\partial}{\partial x} \exp x = \exp x \quad (4)$$

#### Derivative structures

$$f = f(x, y, z)$$

- *Jacobian:*

$$\mathbf{J}_f = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right]$$

- *Hessian:*

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

## Taylor Series

- *Univariate:*

$$\begin{aligned} f(x) &= f(c) + f'(c)(x - c) + \frac{1}{2}f''(c)(x - c)^2 + \dots \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x - c)^n \end{aligned}$$

- *Multivariate:*

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{c}) + \mathbf{J}_f(\mathbf{c})(\mathbf{x} - \mathbf{c}) + \dots \\ &\quad \frac{1}{2}(\mathbf{x} - \mathbf{c})^t \mathbf{H}_f(\mathbf{c})(\mathbf{x} - \mathbf{c}) + \dots \end{aligned}$$

## Optimization and Vector Calculus

- *Newton-Raphson:*

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

- *Grad:*

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial z} \end{bmatrix}$$

- *Directional Gradient:*

$$\nabla f \cdot \hat{r}$$

- *Gradient Descent:*

$$s_{n+1} = s_n - \gamma \nabla f$$

- *Lagrange Multipliers  $\lambda$ :*

$$\nabla f = \lambda \nabla g$$

- *Least Squares -  $\chi^2$  minimization:*

$$\chi^2 = \sum_i^n \frac{(y_i - y(x_i; a_k))^2}{\sigma_i}$$

$$\text{criterion: } \nabla \chi^2 = 0$$

$$\begin{aligned} a_{\text{next}} &= a_{\text{cur}} - \gamma \nabla \chi^2 \\ &= a_{\text{cur}} + \gamma \sum_i^n \frac{(y_i - y(x_i; a_k))}{\sigma_i} \frac{\partial y}{\partial a_k} \end{aligned}$$