The goal is to provide the mathematical foundation to begin learning Machine Learning.

Will cover,

- geometry and intuition
- generalize calc. to multidimensional sys.
- calculus used in neural nets.
- regression.

Formula Sheet

(*) A good resource is: Mathematical Methods in Physical Sciences by Boas, the 3rd Edition Wiley and Sons.

Definition for a derivative

$$f'(x) = \frac{df(x)}{dx}$$

$$= \lim_{\Delta x \to 0} \left(\frac{f(x+\Delta x) - f(x)}{\Delta x} \right)$$

Time saving rules

- Sum Rule:

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

Product Rule:

$$A(x) = f(x)\,g(x)$$

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

- Power Rule:

$$f(x) = ax^b$$

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

Chain Rule:

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

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

- Total Derivative: For the function $f(x, y, z, \ldots)$, where each var is a function of t, the total derivative is $\frac{df}{dt} = \frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt} + \ldots$

Derivatives of named functions

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

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

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

(4) $\frac{\partial}{\partial x} e^x = e^x$

Derivative Structures

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

Jacobian:

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

Hessian:

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

Taylor Series

Univariate:

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

Multivariate:

$$\begin{aligned} f(x) &= f(c) + J_f(c)(x-c) \\ &\quad + \dots \\ &\quad + \tfrac{1}{2}(x-c)^t H_f(c)(x-c) \\ &\quad + \dots \end{aligned}$$

Optimization & Vector Calc.

Newton-Raphson:

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

Gradient:

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

Optimization & Vector Calculus

Directional Gradient:

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

Gradient Descent:

$$S_{n+1} = S_n - \gamma \nabla f$$

Lagrange Multipliers:

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

Least Squares - χ^2 minimization:

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

Criterion:

$$\nabla \chi^2 = 0.$$

$$a_{next} = a_{curr} - \gamma \nabla \chi^2$$

$$= a_{curr} + \gamma \sum_i^n \frac{(y_i - y(x_i; a_k))}{\sigma_i} \frac{\partial y}{\partial a_k}$$