SCALER

MATHS ML CHEAT-SHEET

01. Straight Line

I. Slope-Intercept Form

$$y = mx + c$$

m = slope ,c = intercept

II. Point - Slope Form

$$y - y_1 = m(x - x_1)$$

m = slope , (x_1, y_1) *= a point on line*

III. Two - Point Form

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

Where, (x_1, y_1) *and* (x_2, y_2) *are two points on the line.*

IV. Intercept Form

$$\frac{x}{a} + \frac{y}{b} = 1$$

a, b are the intercepts on the x and y-axis respectively.

V. General Form

$$ax + by + c = 1$$

a, b, c → real numbers

VI. Lines

Lines $y = m_2x + c_2$

and $y = m_1x + c_1$ are :

Parallel : $m_1 = m_2$

Perpendicular : $m_1.m_2 = -1$

02. Hyperplane

I. Vector Form

$$w^Tx + w_0 = 0$$

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ . \\ . \\ . \\ w_n \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

II. Half Spaces

It is either of the two parts into which a plane divides the 3-D Euclidean space.

$w^Tx + w_0 > 0$ **+ve** Half space

$w^Tx + w_0 < 0$ **-ve** Half space

III. Distance from the origin

$$d = \frac{w_0}{||\vec{w}||}$$

IV. Distance from the Point x_0

$$d = \frac{|w^Tx_0 + w_0|}{||w||}$$

03. Vectors

I. Dot Product

$$\vec{a}.\vec{b} = |\vec{a}|.|\vec{b}|.\cos(\theta)$$

$$\vec{a}.\vec{b} = a_1b_1 + a_2b_2 + ... + a_nb_n$$

II. Unit Vector

$$\hat{u} = \frac{\vec{u}}{||\vec{u}||}$$

III. Distance between two points

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

IV. Norm (Magnitude)

represents the length of a vector

$$||\vec{a}|| = \sqrt{\vec{a}.\vec{a}}$$

V. Angle between two vectors

$$\theta = cos^{-}\left(\frac{\vec{a}.\vec{b}}{||\vec{a}||.||\vec{b}||}\right)$$

VI. Projection

$$\vec{a} \;\; on \;\; \vec{b} = \frac{\vec{a}.\vec{b}}{||\vec{b}||}$$

04. Limit

A limit is a value toward which an expression converges as one or more variables approach certain values.

$$\lim_{x \to a} f(x)$$

f(x) is **continuous** at a point x = a if:

$$\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x) = \lim_{x \to a} f(x)$$

05. Equation of a Circle

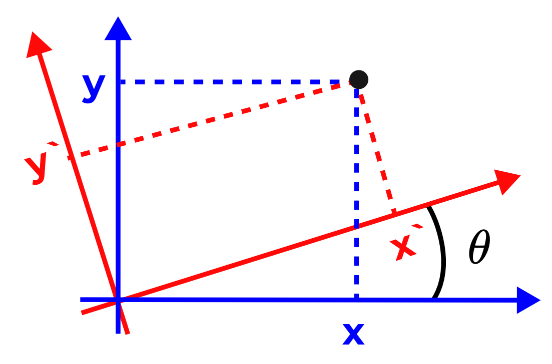
$$(x - x_0)^2 + (y - y_0)^2 = r^2$$

where (x_0, y_0) *→ Centre coordinates and* r *→ Radius*

Points inside the circle give **-ve** values when substituted in the circle equation and points outside the circle give **+ve** values.

06. Rotation of Coordinate system

Let's say we have a coordinate system x-y initially and a point



If the coordinate system is rotated by an angle θ in anticlockwise direction. P in new system would be:

$$x_0' = x_0cos(\theta) + y_0sin(\theta) \;,\;\; y_0' = -x_0sin(\theta) + y_0cos(\theta)$$

07. Differentiation using first principles

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

Rules of Differentiation

Sum/Difference rule:

$$\frac{d}{dx}[f(x) \pm g(x)] = \frac{d}{dx}[f(x)] \pm \frac{d}{dx}[g(x)]$$

Constant multiple rule:

$$\frac{d}{dx}[k.\,f(x)] = k.\,\frac{d}{dx}[f(x)]$$

Product rule:

$$\frac{d}{dx}[f(x).\,g(x)] = f(x).\,\frac{d}{dx}[g(x)] + g(x).\,\frac{d}{dx}[f(x)]$$

Quotient rule:

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x).\,f'(x) - f(x).\,g'(x)}{[g(x)]^2}$$

Chain rule:

$$\frac{d}{dx}[f(g(x))] = f'(g(x)).\,g'(x)$$

1. The **first derivative** gives the slope of the tangent line to the function at a point.
   - *+ve first derivative → the function is increasing*
   - *-ve first derivative → the function is decreasing*
2. The **Second derivative** of a function represents its concavity.
   - *If the second derivative is +ve → concave upwards*
   - *If the second derivative is -ve → concave downwards.*

Common Derivatives

$$\frac{dC}{dx} = 0 \quad \text{(C = Constant)}$$

$$\frac{d(cosx)}{dx} = -sinx$$

$$\frac{dx^n}{dx} = n.\,x^{n-1}$$

$$\frac{db^x}{dx} = b^x.\,ln(b)$$

$$\frac{d(sinx)}{dx} = cosx$$

$$\frac{d(ln(x))}{dx} = \frac{1}{x}$$

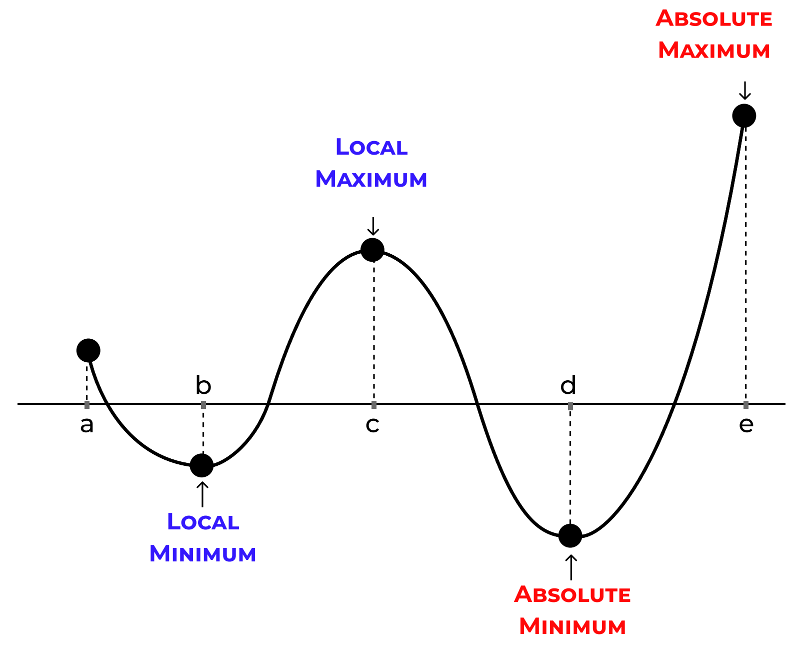
Steps to find the optima

1. Given a function f(x), calculate its derivative. i.e. **f$'$(x)**
2. Put **f$'$(x) = 0** to obtain the stationary points x = c.
3. Calculate **f$'$(x)** at each stationary points x = c (i.e **f$''$(c)**)
4. Choose on of the following ;

i. If ***f$''$(c) > 0****, then f(x) has a minimum value at x = c.*

ii. If ***f$''$(c) < 0****, then f(x) has a maximum value at x = c.*

iii. If ***f$''$(c) = 0****, then f(x) may or may not have a maxima or minima at x = c.*



Partial Derivative

Derivative with respect to one variable, with the others held constant.

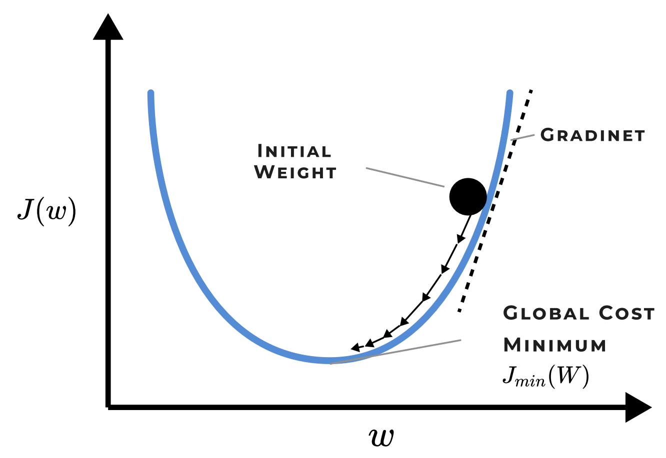
Delta Operator

$$\Delta f(w_0, w_1, w_2, \ldots, w_n) = \begin{bmatrix} \frac{\delta f}{\delta w_0} \\ \frac{\delta f}{\delta w_1} \\ \frac{\delta f}{\delta w_2} \\ . \\ . \\ . \\ \frac{\delta f}{\delta w_n} \end{bmatrix}$$

Optima for f can be found by putting Δf equal to a **Null matrix** of the same dimensions as Δf

08. Gradient Descent:

Iterative algorithm to reach the optima of a function.



GD Algorithm to optimize $f(x,y)$

Step 1: Initially, pick x_0 and y_0 randomly

Step 2: Compute $\frac{\delta f}{\delta x}$ and $\frac{\delta f}{\delta y}$ at $x = x_0$ and $y = y_0$ respectively.

Step 3: The new values of x_0 and y_0 which are closer to the optima are given as:

$$x_1 = x_0 - \eta.\frac{\delta f}{\delta x} \qquad y_1 = y_0 - \eta.\frac{\delta f}{\delta x}$$

Step 4: Repeat step 3 until $\frac{\delta f}{\delta x} \approx 0$ $\frac{\delta f}{\delta y} \approx 0$

Here, η ⇒ learning rate

If the η value is very small, then the updates will happen **very slowly**. If it is a large value, it may **overshoot** the minima.

Variants of Gradient Descent

Batch GD

Calculates the partial derivative using the **full training set** at each step.

$$\theta^{t+1} = \theta^t - \eta.\sum_{i=1}^{n}\frac{\delta f x_i}{\delta \theta}$$

Mini-Batch GD

Calculates the partial derivative using only a **few data points** (random) from the data set.

$$\theta^{t+1} = \theta^t - \eta.\sum_{x_i \in B}\frac{\delta f x_i}{\delta \theta}$$

(B = Mini - Batch)

Stochastic Gradient Descent:

Updates the parameters for each training example **one by one.**

$$\theta^{t+1} = \theta^t - \eta\frac{\delta f x_k}{\delta \theta}$$

WHERE, K IS A RANDOM NUMBER FROM I TO N

09. Method of Lagrange Multipliers:

A method of finding the local minima or local maxima of a function ($e.g\;\; minf(x,y)$) subject to constraints ($g(x,y) = c$)

The problem can be rewritten as:

$$x^*, y^* = min_{x,y}f(x,y) + \lambda(g(x,y) - c) = L(x,y,\lambda)$$

λ → Lagrange multiplier

$L(x,y)$ → Lagrangian function.

10. Eigen Vector and Eigen Value:

For any matrix, there exists a vector such that when this vector is multiplied with the matrix, we get a new vector in the same direction having a different magnitude.

Eigen Vector of Matrix A

$$A\vec{x} = \lambda\vec{x}$$

Eigen Vector of Matrix A

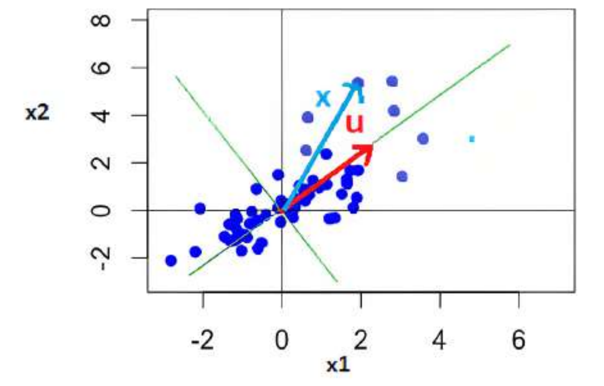
There can be multiple eigen vectors, which are always orthogonal to each other.

The **eigenvector** associated with the largest eigenvalue indicates the direction in which the data has the most variance.

11. Principal Component Analysis (PCA):

The act of finding a new axis to represent the data so that a few principal components may contain most of the information or variance.

Consider a vector **x** in the space representing one of the points in our data and a unit vector **u** representing the direction of the new axis.



The best **u** will be where the summation of the length of projections of all such points(x_i) on the vector **u** is maximum.

Explained Variance

The amount of variability in a data set that can be attributed to each individual principal component.