Lecture 1

Math Basics for DS and ML

Scalar-Vector Operations

Consider a scalar $a \in \mathbb{R}$ and a vector $\bar{x} \in \mathbb{R}^{N \times 1}$. Then,

$$a\bar{x} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_N \end{bmatrix}$$

For scalar-vector addition, since $\bar{x} + a$ is not valid, we consider

$$\bar{x} + a\mathbb{1} = \begin{bmatrix} x_1 + a \\ x_2 + a \\ \vdots \\ x_N + a \end{bmatrix}$$

Vector-Vector Operations

For vectors $\bar{x}, \bar{y} \in \mathbb{R}^{N \times 1}$,

$$\bar{x} + \bar{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_N + y_N \end{bmatrix}$$

Also, we can define an inner product,

$$\bar{x} \cdot \bar{y} = \bar{x}^T \bar{y} = \langle x, y \rangle = x_1 y_1 + \ldots + x_N y_N$$

Similarly,

$$\bar{x} \odot \bar{y} = \begin{bmatrix} x_1 y_1 \\ x_2 y_2 \\ \vdots \\ x_N y_N \end{bmatrix}$$

which is also a $N \times 1$ column vector.

Sub-Spaces

For a vector space V,
If $v_1, v_2 \in V$, then -

1. $cv_i \in V$, where $c \in \mathbb{R}$
2. $v_1 + v_2 \in V$

Now, if $w_1, w_2 \in V$ and $c_1, c_2 \in \mathbb{R}$, then

$$c_1 w_1 + c_2 w_2 \in W \subset V,$$

where W is a subspace described by w_1 and w_2.

Matrix-Matrix Operations

Consider two matrices $X, Y \in \mathbb{R}^{M \times N}$. And column vectors are $x, y \in \mathbb{R}^{N \times 1}$.
Then,

$$X + Y = \begin{bmatrix} x_{11} + y_{11} & \cdots & x_{1N} + y_{1N} \\ \cdots & \cdots & \cdots \\ x_{M1} + y_{M1} & \cdots & x_{MN} + y_{MN} \end{bmatrix}$$

For $Z \in \mathbb{R}^{N \times P}$,

$$X \cdot Z = \begin{bmatrix} x_{11} z_{11} + \ldots + x_{1N} z_{N1} & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & x_{M1} z_{1P} + \ldots + x_{MN} z_{NP} \end{bmatrix}$$

Also,

$$X \odot Y = \begin{bmatrix} x_{11} y_{11} & \cdots & x_{1N} y_{1N} \\ \cdots & \cdots & \cdots \\ x_{M1} y_{M1} & \cdots & x_{MN} y_{MN} \end{bmatrix}$$

Transpose, Determinant and Inverse of a matrix

Let the elements of X^T be b_{ij}. Then,

$$b_{ij} = a_{ji},$$

for all i, j in range and a_{ij} are the elements of X.

$X_{M \times N} X^{-1}_{M \times N} = I_{M \times N}$
If X is invertible $\implies X$ should be Full Rank, where $Rank(X)$ = Number of independent rows and columns of a matrix

Pseudo-Inverse

If, $X^{+}_{N \times M} X_{M \times N} = I_{N \times N}$, then X^{+} is the pseudo-inverse of the non-square matrix X.

The formula for X^+ is,

$$(X^H X)^{-1} X^H,$$

where X^H is the conjugate transpose of X. If all entries of X are real, then X^H becomes X^T (transpose matrix).

Eigen Decomposition

We have a matrix $A(N \times N)$ for which we are trying to find a vector v_i for which,

$$Av_i = \lambda_i v_i,$$

where v_i's are the (normalized, i.e. $||v_i||_2^2 = 1$ and orthogonal, i.e. $v_i^T v_j = \delta_{ij}$) eigenvectors of A, and λ_i are the corresponding eigenvalues.
Implication of $\lambda_i = 0 \implies \text{Rank Deficiency}$

The matrix A can be eigen-decomposed as,

$$A = Q \Lambda Q^{-1}$$

where, $Q = [\bar{v_1}, \bar{v_2}, \cdots, \bar{v_N}]$ and Λ is the matrix with diagonal entries as λ_i .

Tensors

Say, a tensor $T \in \mathbb{R}^{M \times N \times P}$ and a matrix $X \in \mathbb{R}^{M \times N}$.
For transpose of T, we also need to define the order of dimension swap, whereas in matrices, only 1 swap was possible (since, $2D$ matrices).

$$\text{e.g. } transpose(T, [0, 2, 1])$$

Functions

$$f : X \to Y$$

where, $x \in X$ and $f(x) \in Y$.

Continuity

If

$$\underset{\Delta x \to 0}{lt} f(x + \Delta x) = f(x)$$

and

$$\underset{\Delta x \to 0}{lt} f(x - \Delta x) = f(x)$$

then, $f(x)$ is continuous at x.

Smoothness

If the derivative is continuous at x, then function is smooth at x.

Lipschitz Continuity

If

$$|f(x + \Delta x) - f(x)| \leq K\Delta x$$

for some $K \in \mathbb{R}$, then the function is Lipschitz continuous at x.

Derivative of a function

$$\frac{d}{dx} f(x)$$

Critical Points

The points where $f'(x) = 0$, which can lead to three types of critical points:

- Maxima - If $f''(x) < 0$
- Minima - If $f''(x) > 0$
- Inflection Point - If $f''(x) = 0$

Multi-Variate functions

$y = f(x_1, x_2)$, then gradient of the function is

$$\bigtriangledown f = \begin{bmatrix} \frac{df}{dx_1} \\ \frac{df}{dx_2} \end{bmatrix}$$

Thus, for calculating maxima and minima (or saddle point), we equate all entries of $\bigtriangledown f$ to 0.

Hessian Matrix

Now,

$$Hf = \begin{bmatrix} \frac{d^2f}{dx_1^2} & \frac{d^2f}{dx_1dx_2} \\ \frac{d^2f}{dx_2dx_1} & \frac{d^2f}{dx_2^2} \end{bmatrix}$$

All eigenvalues of Hf positive, then minima, if all negative, then maxima, else it is a saddle point.

Constrained Optimization using Lagrange Multiplier

If we want to maximize $f(x)$, then we will try to find the critical points and find the maxima. This would be known as unconstrained optimization.

For (equality) constrained optimization, maximize $f(x)$, subject to $g(x) = 0$.

Then, the direction of normals of $f(x)$ and $g(x)$ should be aligned in graphical representation.
Thus, we define a Lagrangian function

$$L(x) = f(x) + \lambda g(x), \lambda \neq 0$$

Making the gradient of L to be 0, we obtain λ.

$$\bigtriangledown L(x) = 0 \implies \bigtriangledown f(x) = -\lambda \bigtriangledown g(x)$$

where, substituting values of x obtained from $g(x) = 0$, will give the required λ.