Theoretical Machine Learning

Week 1 : Catching Up Mathematically

12th May 2024

Introduction

Most of the maths required should have been taught in your MTH courses. Mainly, we would be using basic Calculus and basic Linear Algebra for this week's topic.
Most of you should be familiar with it, as it was mentioned as a pre requisite for this project, but if you have any doubts, let us know.
We will be looking at calculus with matrices and vectors.
We will be looking at x cases :

- Derivative of Scalar wrt scalar variable.
- Derivative of Vector wrt scalar variable.
- Derivative of Matrix wrt scalar variable.
- Derivative of Scalar wrt vector.
- Derivative of Vector wrt vector.
- Derivative of Matrix wrt matrix.
- Derivative of Scalars wrt matrix.
- Derivative of Vector wrt matrix.
- Derivative of Matrix wrt matrix.

Scalar wrt Scalar

You already know how to do this. This is the basic differentiation you people have been taught in calculus.

Examples

- $\frac{d}{dx}[x^2] = 2x$
- $\frac{d}{dx}[sin(x)] = cos(x)$
- $\frac{\partial}{\partial x}[25x^2y - 4y] = 50xy$

Vector wrt Scalar

This is also very similar to Scalar with Scalar case. Here, you just differentiate each element of the vector wrt the scalar variable.

Examples

- $\frac{d}{dx}\left(\begin{bmatrix} 25x \\ 3x^2 \\ \frac{5}{x} \end{bmatrix}\right) = \begin{bmatrix} 25 \\ 6x \\ \frac{-5}{x^2} \end{bmatrix}$
- $\frac{\partial}{\partial x}\left(\begin{bmatrix} 3xy \\ 4y \\ e^x \end{bmatrix}\right) = \begin{bmatrix} 3y \\ 0 \\ e^x \end{bmatrix}$

Matrix wrt Scalar

This is also very similar to Vector with Scalar case. Here, you just differentiate each element of the matrix wrt the scalar variable.

Examples

- $\frac{d}{dx}\begin{bmatrix} 5x & sin(x) & 4 \\ 3 & cos(x^2) & \frac{1}{x} \end{bmatrix} = \begin{bmatrix} 5 & cos(x) & 0 \\ 0 & -2xsin(x^2) & \frac{-1}{x^2} \end{bmatrix}$
- $\frac{\partial}{\partial x}\begin{bmatrix} 5xy & sin(xy^2) & 4x^3 \\ 3y^2 & cos(x^2y) & \frac{1}{xy} \\ xy^2 & 2y & sin^2(x) \end{bmatrix} = \begin{bmatrix} 5y & y^2cos(xy^2) & 12x^2 \\ 0 & -2xysin(x^2y) & \frac{-1}{x^2y} \\ y^2 & 0 & 2sin(x)cos(x) \end{bmatrix}$

Scalar wrt Vector

This is where stuff gets more interesting. So, basically, the result of this operation is the transpose of a vector. Examples will make it clearer. You should be able to relate it with MTH 112.

Examples

- $V = (x, y, z)^T$
  $\frac{d}{dV}[x^3y^2z] = (3x^2y^2z,\ 2x^3yz,\ x^3y^2)$

Effectively, when you get into the denominator, you get transposed. (Kinda)

In general, this is the idea behind this case :

$\frac{dF}{dV} = (\frac{\partial F}{\partial V_1}, \frac{\partial F}{\partial V_2}, \frac{\partial F}{\partial V_3} ..., \frac{\partial F}{\partial V_N})$ for an N dimensional Vector.

Vector wrt Vector

Scalar wrt Vector with more dimensions.
The idea is to divide the to-be-differentiated vector into scalar components, and getting the derivative wrt the vector, and using all of them, to get the final answer. The final output ends up being a matrix.
Say, $V_1 = (x_1, x_2, ..., x_n)^T$ and $V_2 = (y_1, y_2, ..., y_m)^T$
We need $\frac{dV_2}{dV_1}$

Basically, $\frac{dV_2}{dV_1} = \frac{d}{dV_1} \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_m \end{bmatrix}$

This looks similar to the vector wrt scalar case. So, we can directly, write

$\frac{dV_2}{dV_1} = \begin{bmatrix} \frac{dy_1}{dV_1} \\ \frac{dy_2}{dV_1} \\ . \\ . \\ \frac{dy_m}{dV_1} \end{bmatrix}$ Now, since $\frac{dy_k}{dV_1}$ is a vector transpose of form $(\frac{dy_k}{dx_1}, \frac{dy_k}{dx_2}, ..., \frac{dy_k}{dx_n})$, we write

$$\frac{dV_2}{dV_1} = \begin{bmatrix} \frac{dy_1}{dx_1} & \frac{dy_1}{dx_2} & . & . & \frac{dy_1}{dx_n} \\ \frac{dy_2}{dx_1} & \frac{dy_2}{dx_2} & . & . & \frac{dy_2}{dx_n} \\ . & . & . & . & . \\ . & . & . & . & . \\ \frac{dy_m}{dx_1} & \frac{dy_m}{dx_2} & . & . & \frac{dy_m}{dx_n} \end{bmatrix}$$

As an example, try to do $\dfrac{d\begin{bmatrix} 2sin^2(x)cos(y) \\ x^2 + 3e^y \end{bmatrix}}{d\begin{bmatrix} x \\ y \end{bmatrix}}$

Matrix wrt Vector

In this, again, the idea is similar. Take each scalar value in the matrix, and differentiate wrt the vector, or to simplify, take each column of the matrix as a vector, and differentiate wrt the vector, and use that to get the answer. The final output will be a bigger matrix.

Challenge : Suppose you have a matrix of dimension m x n, and you differentiating wrt a k x 1 vector, what is the dimension of the final result?

Scalar wrt Matrix

Here, you will see the transpose rule again.
Say, you have a 2x2 matrix, M.

$$\frac{d}{dM} = \begin{bmatrix} \frac{d}{dM_{1,1}} & \frac{d}{dM_{2,1}} \\ \frac{d}{dM_{1,2}} & \frac{d}{dM_{2,2}} \end{bmatrix}$$

Example

$$F = 3x^2 + 5y - 2xw + 3ysin(z)$$

$$M = \begin{bmatrix} x & y \\ w & z \end{bmatrix}$$

$$\frac{dF}{dM} = \begin{bmatrix} \frac{dF}{dx} & \frac{dF}{dw} \\ \frac{dF}{dy} & \frac{dF}{dz} \end{bmatrix} = \begin{bmatrix} 6x - 2w & -2x \\ 5 + 3sin(z) & 3ycos(z) \end{bmatrix}$$

Now, for Vector and Matrix wrt Matrix, you can use the same idea for how we did wrt Vector, by taking each scalar value, and differentiating.

Double Derivative

We are not going too much into it now, in terms of calculation. I will just give the basic idea on how to calculate it.

$$\frac{\partial^2 A}{\partial B^2} = \frac{\partial}{\partial B^T}\left[\frac{\partial A}{\partial B}\right]$$

Important Identities

Now that we know how to calculate derivatives related to matrices, we can set up some important identities.
Here, take A to be a Constant Matrix, and x to be a vector.

- $\frac{d}{dx}[Ax] = A$
- $\frac{d}{dx}[x^T A x] = x^T(A + A^T)$
- If A is symmetric, $\frac{d}{dx}[x^T A x] = 2x^T A$