Foundations of AI Part II

Justin A. Dang

Consider the following derivative:

$$\frac{d}{d\alpha}[f(x)]$$

This derivative is with respect to α, so we assume all other variables (including x) will be a constant unless we define it otherwise.

```
Let y be a function of x.
```

eg.

$$y(x) = x + 10$$

$$y = x + 10$$

Remember that variables can point of functions or function definitions. There are numerous ways to represent these functions.

$$\frac{d}{dx}[y] = \frac{d[y]}{dx} = \frac{dy}{dx}$$

Say that u is a function where

$$u = x^2$$

So with respect to x,

$$\frac{d}{dx}[u] = \frac{du}{dx} = 2x$$

Chain Rule

We consider the role of composition for solving derivatives.

$$f(g(x))$$

$$g(x) = x^2$$

$$f(z) = z + z^2$$

1. Find the derivative of the outer function

$$\frac{d}{d\_}[f(g(n)]$$

2. Find the derivative of the inner function. Let's wrap $g(x)$ under a variable u and rewrite as

$$\frac{d}{du}[f(u)] \cdot \frac{du}{dx}$$

What is the derivative for $f(g(x))$?

$$f(g(x)) = (g(x)) + (g(x))^2$$

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

$$f(g(x)) = (x^2) + (x^2)^2$$

Substitute $g(x)$ with u.

$$\frac{d}{dx}[(x^2) + (x^2)^2] = u + (u)^2$$

$$\frac{d}{du}[u + u^2] = \frac{d}{du}[u] + \frac{d}{du}[u^2]$$

$$[1 + 2u] \cdot \frac{du}{dx}$$

$$\frac{du}{dx} = \frac{d}{dx}[u] = \frac{d}{dx}[x^2] = 2x$$

$$= [1 + 2u] \cdot (2x)$$

$$= [1 + 2x^2] \cdot (2x)$$

in accordance with the chain rule.

To simplify,

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

where $u = g(x)$

$$= \frac{df}{du} \cdot \frac{du}{dx}$$

or the `inner function` times the `outer function`.

Key Points

1. Understand notation.

$$\frac{d}{d\_}[\_]$$

   $d\_$ represents the variable that can be changed and $\_$ represents the function of interest.

2. Derivative of constants is 0.

$$\frac{d}{dx}[C] = \emptyset$$

   Derivative of x^2 is $2x$.

$$\frac{d}{dx}[x^2] = 2x$$

$$\frac{d}{dx}[x + x^2] = \frac{d}{dx}[x] + \frac{d}{dx}[x^2]$$

,

$$\frac{d}{d\beta}[x^2] = 0$$

and the chain rule:

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

.

These few notations lead to state of the art results in the AI discipline.

Partial Derivative

There are many cases where there are multiple inputs needed in a function.

$$f(x_1, x_2, x_3)$$

In this case, solve for one variable, treating the other variables as constants:

$$\frac{2}{2x_1}[f(x_1, x_2, x_3)] = x_1 + x_2 + x_2 + x_3 + x_1^2 + x_3$$

Basic Math Data Structures

1. Scalar. A single value $[x]$.
2. Vector. $[x_1, x_2, \ldots x_p]_{1 \times p}$. Can be a row or column vector.
3. Matrix.

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}$$

4. Tensor. 3-dimensional arrays.

Element wise multiplication

Put two matrices $A \times B$ of the same dimensions and multiply values in corresponding locations.

```
for i to A.length
        for k to A[i].length
                C[i][k] = A[i][k] x B[i][k]
```

Dimensions must be identical.

Matrix Multiplication

For a given matrix A of $n \times m$ size, and a matrix B of $m \times n$ size, the product is matrix C with a dimension of $n \times n$.

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 2 \end{bmatrix} \times \begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 8 \\ 7 & 13 \end{bmatrix}$$

Data

1. Discrete - $0, 1, 2..$.
2. Continuous - $2.1113, 2.11, 3.10...$
3. Categorial - A, B, C, D, F
4. Ordinal - Ranking (1st, 2nd, 3rd...)

Ordinal data should be handled carefully and not ideal to be fed into a machine learning algorithm.

Data Set

A data set ($n \times p$) is comprised of:

1. A feature or co variance ($x_1, x_2, x_3...$).
2. An observation - an intersection between a feature and a record.

To minimize data, we use derivatives.