khan academy의 강의를 듣고 간략하게 정리해보았습니다.

gradient

- store all partial derivative information of a multivariate function
- vector-valued function

$$\nabla f = [\frac{\partial f}{\partial x}\ \frac{\partial f}{\partial y}\ \frac{\partial f}{\partial z} ...]^T$$

- 특징
  - gradient vector $\nabla f(x_0, y_0, ...)$는 함수식 f 위의 점 $(x_0, y_0, ...)$에서 f 에 perpendicular 하다.
  - $\nabla f(x_0, y_0, ...)$는 해당 점에서 f 가 가장 빠르게 커지게하는 방향을 의미한다.

문제상황

- maximize $f(x, y) = x^2 y$
- constraint : $x^2 + y^2 = 1$

두 함수가 만나는 지점에서 f를 최대로 하는 점을 찾아야한다. 여기서 우리는 gradient를 이용한다.

- 해당 점을 x_m, y_m 라고 하자.
- $g(x, y) = x^2 + y^2$ 라고 하자.
- λ : Lagrange multiplier

$$\nabla f(x_m, y_m) = \lambda \nabla g(x_m, y_m)$$

해당 점에서 각 함수의 gradient는 어떤 상수배(λ)를 통해 같아질 수 있다. 해당함수(targent)에 perpendicular하기 때문이다. 그래프를 머리속으로 그려서 생각해보자!

$$\nabla f = [2xy\ x^2]^T, \nabla g = [2x\ 2y]^T$$

이므로 우리는 식 3개와 변수 3개

- $2xy = \lambda 2x$
- $x^2 = \lambda 2y$
- $x^2 + y^2 = 1$ (제약식)

를 얻을 수 있고 이를 통해 x_m, y_m 을 구할 수 있다.

Lagrange method

- Lagrangian
  - optimize f
  - constraint $g = C$

$$L(x, y, \lambda) = f(x, y) - \lambda(g(x, y) - C)$$

이를 미분해서 0이 되는 x^*, y^*, λ^* 를 구하면 된다.

- Lagrange multiplier의 의미

$M^* = f(x^*, y^*)$라고 하자. 그런데 이 식은 C에 따라 달라지기에

$$M^*(C) = f(x^*(C), y^*(C))$$

이를 이용하여 λ^*에 대해 정리하면 (증명 생략)

$$\lambda^* = \frac{dM^*}{dC}$$