데이터과학에서 가장 자주 사용되는 수학 도구는 시그마(Σ)이다. 시그마의 성질을 몇 가지만 기억해두자.

cost 함수에 시그마가 사용됨.

$$\sum_{k=1}^{n}(a_k + b_k) = \sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k$$

$$\sum_{k=1}^{n} ca_k = c\sum_{k=1}^{n} a_k$$

$$\sum_{k=1}^{n} c = cn$$

$$\frac{\partial}{\partial x_i}(\sum_{i=1}^{n} f_i) = \sum_{k=1}^{n} \frac{\partial f_i}{\partial x_i}$$

이 중 마지막 식만 말로 바꾸면, 시그마(합)의 편미분은 편미분의 시그마(합)이다(Partial derivative of a sum is just the sum of the partial derivatives).

미분과 시그마를 선형회귀에 적용해보자.

회귀식에 의한 예측은 $\hat{y}_i = b + ax_i$. 오차는 관측값과 회귀방정식에 의한예측값과의 차이 $e = y_i - \hat{y}_i$. 최소제곱법에 의하면, 오차제곱의 합(Sum of Squares of Errors, SSE)을 최소화하는 계수(b, a)를 구해야 한다. 그럴려면, SSE을 각 회귀계수(와 절편)에 대하여 미분한 값이 0이 되어야 한다.

$$SSE = \sum_{i=1}^{n} e^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - b - ax)^2 \cdots (3)$$

먼저 절편 b에 대해 편미분하면,

$$\frac{\partial SSE}{\partial b} = \frac{\partial}{\partial b}\sum_{i=1}^{n}(y_i - b - ax_i)^2 = \sum_{i=1}^{n}\frac{\partial}{\partial b}(y_i - b - ax_i)^2 = \sum_{i=1}^{n} 2(y_i - b - ax_i)(-1) = -2\sum_{i=1}^{n}(y_i - b - ax_i) = 0$$

다음 a에 대해 편미분하면,

$$\frac{\partial SSE}{\partial a} = \frac{\partial}{\partial a}\sum_{i=1}^{n}(y_i - b - ax_i)^2 = \sum_{i=1}^{n}\frac{\partial}{\partial a}(y_i - b - ax_i)^2 = \sum_{i=1}^{n}2(y_i - b - ax_i)(-x_i) = -2\sum_{i=1}^{n}(y_i - b - ax_i)x_i = 0$$

이 식을 정리하면,

$$(\sum_{i=1}^{n} x_i)b + (\sum_{i=1}^{n} x_i^2)a = \sum_{i=1}^{n} x_i y_i \cdots (5)$$

(4)와 (5)를 정규방정식(normal equation)이라고 한다. 데이터를 가지고 이 연립방정식을 풀어서 절편과 기울기를 구한다.

이상으로 미분과 시그마에 대한 복습을 마친다. 데이터과학을 수행하다 보면 여기서 복습한 내용보다 더 복잡한 수식이 필요하곤 하지만 그때 그때 인터넷 등을 참고하면 이해할 수 있을 것이다.