

# 행렬대수 1 - 최소제곱법, 벡터미분, 정규방정식

시립대학교 통계학과

2019년 6월 3일

## 1 단순선형회귀분석

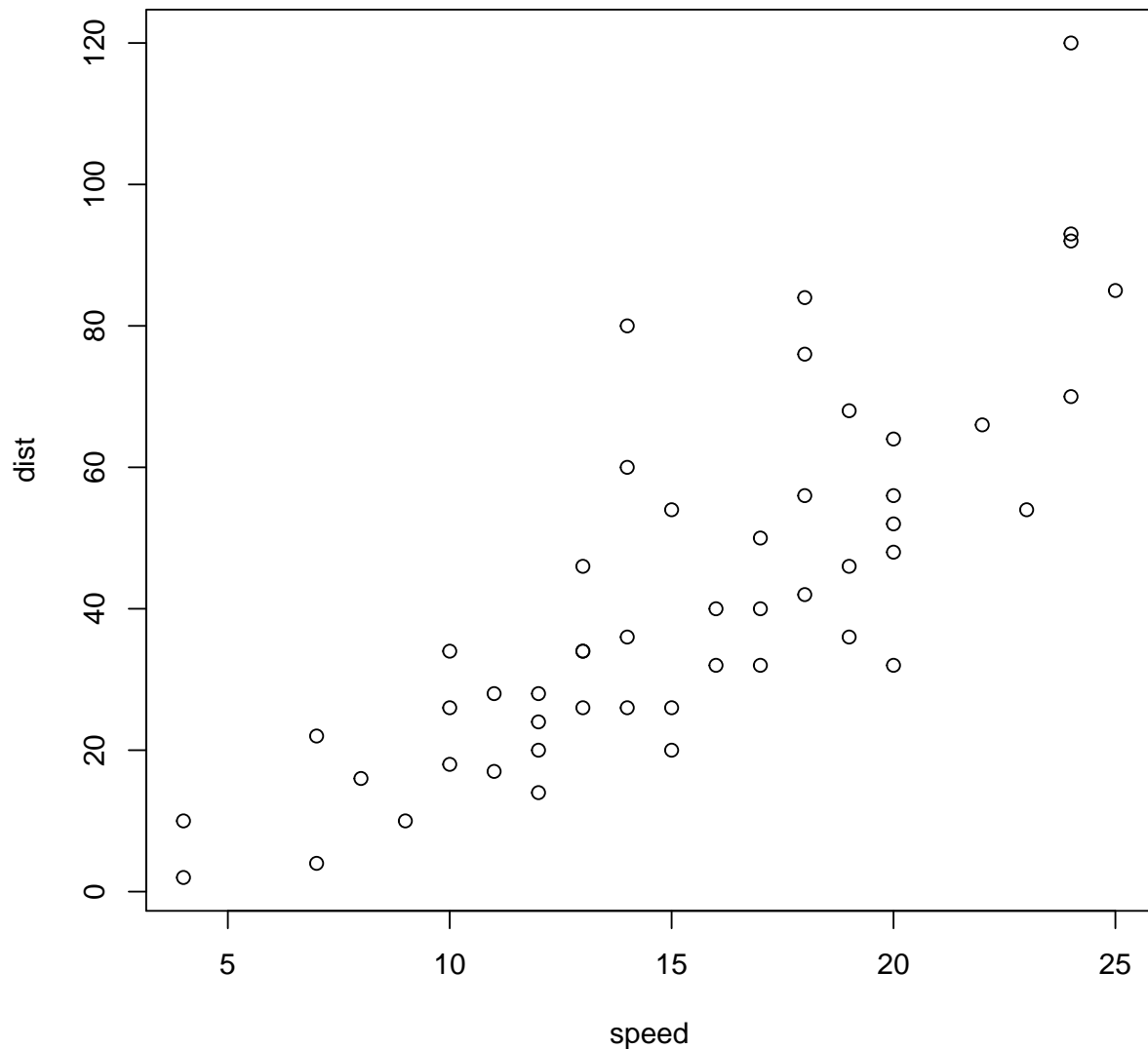
### 1.1 예제: 자동차의 속도와 제동거리

자동차가 달리는 속도(speed, mph)와 제동거리(dist, ft)의 관계를 알아보기 위하여 50대의 자동차로 실험한 결과 자료와 산포도가 아래와 같다.

```
head(cars)
```

```
##      speed dist
## 1         4    2
## 2         4   10
## 3         7    4
## 4         7   22
## 5         8   16
## 6         9   10
```

```
plot(cars)
```



위와 같은 자료를 이용하여 자동차의 속도가 주어졌을 경우 제동거리의 평균에 대한 예측을 하려고 한다면 어떤 방법을 사용해야 할까? 회귀분석(regression analysis)은 여러 가지 변수들의 관계를 분석하는 통계적 방법이다. 일반적으로 한 개 또는 여러 가지의 설명변수들(explanatory variables)이 관심있는 종속변수(response variable)에 어떤 형태로 영향을 미치는지에 파악하고 설명변수와 종속변수의 관계를 통계적으로 추론하는 것이 회귀분석의 목적이다.

자동차의 속도를  $x$  라고 하고 제동거리를  $y$  라고 하면 다음과 같은 선형식으로 자동차의 속도와 제동거리의 관계를 나타내는 것을 단순선형회귀식(simple linear regression equation)이라고 한다.

$$E(y|x) = \beta_0 + \beta_1 x \quad (1)$$

식 (1)은  $y$ 의 평균이  $x$ 의 선형식으로 나타나는 관계를 가정한 것이며 절편  $\beta_0$ 와 기울기  $\beta_1$ 은 모르는 모수로서

자료를 통하여 추정해야 한다.

위에서 본 cars 예제와 같이  $n$  개의 자료를 독립적으로 추출하였다면 자료의 생성 과정을 다음과 같은 선형 회귀모형(linear regression model)로 나타낸다. 종속변수  $y$ 는 설명변수  $x$ 의 선형식으로 나타내어지는 결정적인 요인과 확률 변수로 나타내어지는 임의의 오차항  $e$ 의 합으로 나타내어진다.

$$y_i = E(y_i|x_i) + e_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, 2, \dots, n \quad (2)$$

위에서 오차항  $e_i$  평균이 0이고 분산이  $\sigma^2$  인 임의의 확률분포를 따르며 서로 독립이다.

$$E(e_i) = 0, \quad V(e_i) = \sigma^2 \quad i = 1, 2, \dots, n$$

## 1.2 최소제곱법

단순회귀식 (2)에서 모수  $\beta_0$ 와  $\beta_1$ 를 회귀계수(regression coefficient) 라고 하며 자료(observation; data)를 수집하여 추정해야 한다.  $n$ 개의 자료를 이용하여 회귀계수  $\beta_0$ 와  $\beta_1$ 를 추정하려고 할 때 가장 쉽고 오래되었으며 또한 가장 유용한 방법인 최소제곱법(least square method)을 사용할 수 있다. 일단 위의 식 (2)에서 종속변수의 관측값  $y_i$ 을 대응하는 설명변수  $x = x_i$ 를 이용하여 예측한 값은  $\beta_0 + \beta_1 x_i$ 이다. 여기서 실제 관측하여 얻어진 값  $y_i$ 와 예측값  $\beta_0 + \beta_1 x_i$ 사이에는 차이가 발생한다. 그 차이를 잔차(residual)라고 하며 표현하면 다음과 같다.

$$r_i = y_i - E(y_i|x_i) = y_i - (\beta_0 + \beta_1 x_i)$$

잔차는 위에 식에서 알 수 있듯이 관측값과 회귀식을 통한 예측값의 차이를 나타낸 것이다. 그러면 자료를 가장 잘 설명할 수 있는 회귀직선을 얻기 위해서는 잔차  $r_i$ 를 가장 작게하는 회귀모형을 세워야 한다. 잔차들을 최소로 하는 방법들 중 하나인 최소제곱법은 잔차들의 제곱합을 최소로 하는 회귀계수를 추정하는 방법이다. 잔차들의 제곱합은 다음과 같이 표현된다.

$$S(\beta_0, \beta_1) = \sum_{i=1}^n r_i^2 = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2 \quad (3)$$

위의 잔차 제곱합  $S(\beta_0, \beta_1)$  을 최소화하는  $\beta_0$ 와  $\beta_1$ 의 값을 구하는 방법은 잔차 제곱합이  $\beta_0$ 와  $\beta_1$ 의 미분 가능한 2차 함수이고 아래로 볼록한 함수(convex function)임을 이용한다. 각각의 회귀계수에 대해서 편미분을 하고 0으로 놓으면 아래와 같이 정리된다.

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = \sum_{i=1}^n (-2)[y_i - (\beta_0 + \beta_1 x_i)] = 0 \quad (4)$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^n (-2x_i)[y_i - (\beta_0 + \beta_1 x_i)] = 0 \quad (5)$$

위의 연립방정식을 행렬식으로 표시하면 다음과 같이 나타낼 수 있다.

$$\begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}$$

위의 방정식을 풀어서 구한 회귀계수의 추정치를  $\hat{\beta}_0, \hat{\beta}_1$  이라고 하면 다음과 같이 주어진다.

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}\end{aligned}$$

## 2 회귀식의 행렬형식

일반적으로 회귀모형에서 종속변수의 수는 하나인 경우가 많지만 설명변수의 수는 여러 개인 경우가 많다. 이런 경우 중선형회귀식(multiple linear regression)은 다음과 같이 표현할 수 있고,  $p$ 개의 설명변수가 있다고 가정하고  $(x_1, x_2, \dots, x_p)$  표본의 크기  $n$ 인 자료가 얻어지면 선형회귀식을 행렬로 다음과 같이 표현할 수 있다.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e_i \quad (6)$$

$$= \mathbf{x}_i^t \boldsymbol{\beta} + e_i \quad (7)$$

위의 식을 다시 표현하면 다음과 같이 쓸 수 있다.

$$y_i = \begin{bmatrix} 1 & x_{i1} & x_{i2} & \dots & x_{ip} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + e_i$$

$n$ 개의 관측치가 있을 때  $n$ 개의 회귀식을 행렬식으로 표현하면 다음과 같다.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

위의 식을 벡터를 이용하여 표시하면 다음과 같다.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (8)$$

위의 행렬식에서 각 벡터와 행렬의 차원은 다음과 같다.

- $\mathbf{y}$ :  $n \times 1$
- $\mathbf{X}$ :  $n \times (p+1)$
- $\boldsymbol{\beta}$ :  $(p+1) \times 1$
- $\mathbf{e}$ :  $n \times 1$

여기서 회귀분석의 오차항은 서로 독립이고 동일한 분산을 갖는다. 즉, 오차항은 다음의 분포를 따른다. 즉,  $E(e) = 0$  이므로 관측값 벡터  $y$ 의 평균을 보면

$$E(y|X) = E(X\beta + e) = X\beta + E(e) = X\beta \quad (9)$$

### 3 최소제곱추정

최소제곱추정법(least square estimation)은 자료의 관계를 잘 반영하는 회귀식을 구한 다음 실제 관측값  $y_i$ 과 예측값  $x_i^t\beta$  간에 차이인 잔차를 가장 작게 만드는 것이 목적이다. 모든 잔차항의 제곱의 합을 최소화하는 방법을 최소제곱법이라고 하며 이를 이용하여 회귀계수의 추정량을 찾는다.

$$\min_{\beta} \sum_{i=1}^n (y_i - x_i^t\beta)^2 = \min_{\beta} (y - X\beta)^t (y - X\beta) \quad (10)$$

$\hat{\beta}$ 는 잔차의 제곱합 (10) 을 최소로 하는 최소제곱 추정량이다. 잔차의 제곱합을  $S(\beta)$ 이라고 하면

$$\begin{aligned} S(\beta) &= (y - X\beta)^t (y - X\beta) \\ &= y^t y - y^t X\beta - \beta^t X^t y + \beta^t X^t X\beta \\ &= y^t y - 2\beta^t X^t y + \beta^t X^t X\beta \end{aligned} \quad (11)$$

여기서  $S(\beta)$ 를 최소로 하는 회귀계수벡터의 값을 구하기 위하여  $S(\beta)$ 를 회귀계수벡터  $\beta$ 로 미분한후  $0$  으로 놓고 선형 방정식을 풀어야 한다.

다음 절에 나오는 벡터미분을 이용하면

$$\begin{aligned} \frac{\partial S(\hat{\beta})}{\partial \beta} &= \frac{\partial}{\partial \beta} (y^t y - 2\beta^t X^t y + \beta^t X^t X\beta) \\ &= 0 - 2X^t y + 2X^t X\beta \\ &= 0 \end{aligned}$$

최소제곱 추정량을 구하기 위한 정규방정식은 다음과 같이 쓸 수 있다.

$$X^t X\beta = X^t y \quad (12)$$

방정식 (12)를 정규방정식(normal equation)이라고 한다. 만약  $X^t X$ 가 정칙행렬일 경우 최소제곱법에 의한 회귀계수 추정량  $\hat{\beta}$  다음과 같다.

$$\hat{\beta} = (X^t X)^{-1} X^t y \quad (13)$$

예측값 벡터  $\hat{y}$  는  $E(y|X)$ 의 추정치로서 다음과 같다.

$$\hat{E}(y|X) = \hat{y} = X\hat{\beta} = X(X^tX)^{-1}X^ty$$

만약  $X^tX$ 가 정칙행렬이 아닐 경우 최소제곱법에 의한 회귀계수 추정량  $\hat{\beta}$ 은  $X^tX$ 의 일반화 역행렬  $(X^tX)^{-}$ 를 이용하여 다음과 같이 구한다. 이 경우 일반화 역행렬이 유일하지 않기 때문에 회귀계수 추정량도 유일하지 않다.

$$\hat{\beta} = (X^tX)^{-}X^ty$$

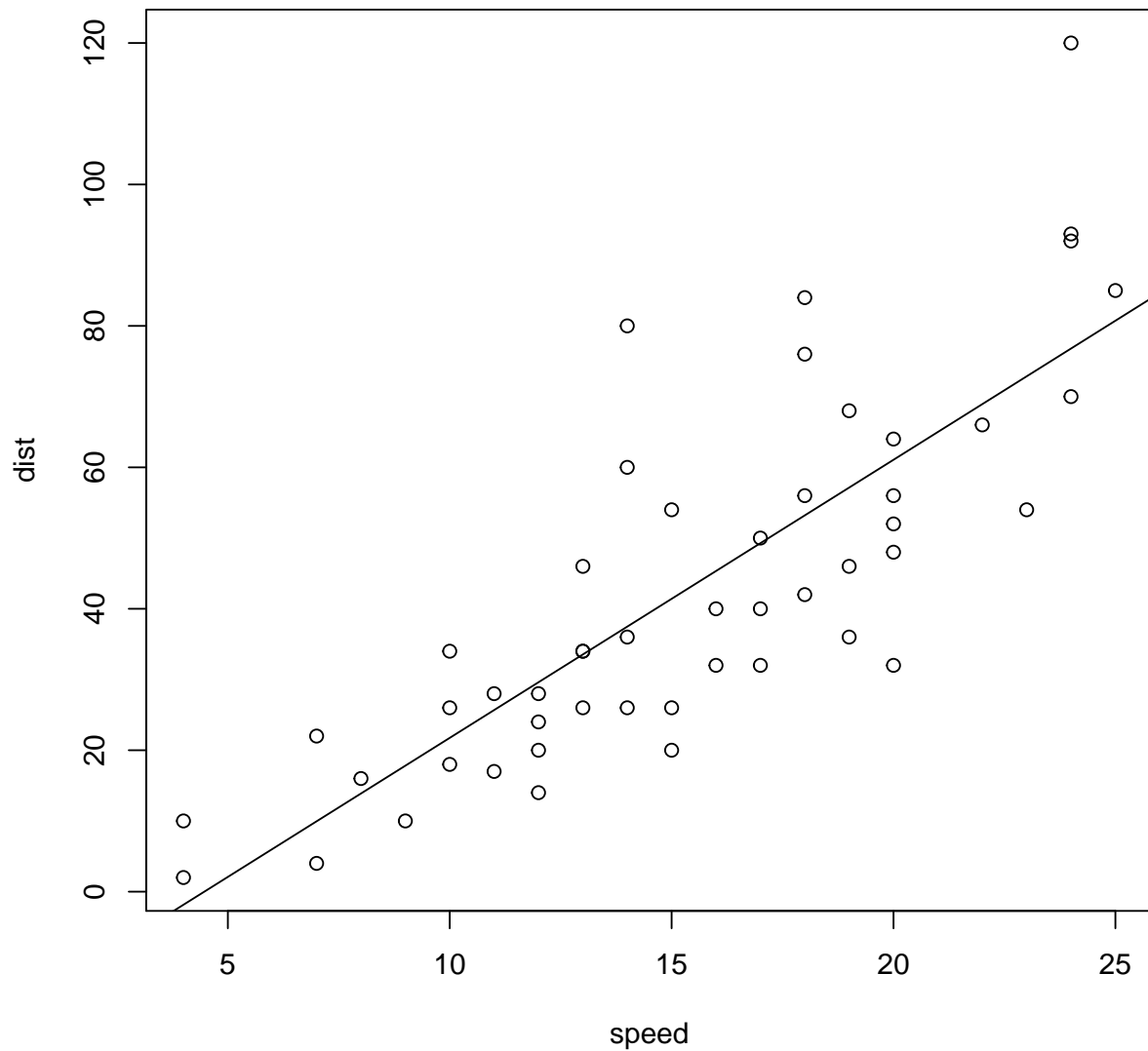
위에서 살펴본 cars 자료에 대한 회귀분석의 결과와 추정된 회귀직선은 다음과 같다.

$$E(dist|speed) = -17.5791 + 3.9324 * (speed)$$

```
res <- lm(dist~speed, data=cars)
summary(res)

##
## Call:
## lm(formula = dist ~ speed, data = cars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601   0.0123 *
## speed        3.9324     0.4155   9.464 1.49e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.38 on 48 degrees of freedom
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF, p-value: 1.49e-12

plot(cars)
abline(res)
```



## 4 벡터미분

### 4.1 스칼라미분

벡터미분(Vector differential) 또는 행렬미분(Matrix differential)은 벡터와 행렬의 미분식에 대한 표기법을 정의하는 방법이다. 보통 스칼라(scalar)에 대한 미분은 일변수 함수  $f : \mathbb{R}^1 \rightarrow \mathbb{R}^1$  또는 다변수 함수(function of several variables)  $f : \mathbb{R}^n \rightarrow \mathbb{R}^1$ 에서 쉽게 정의된다. 만약  $y = f(x)$  또는  $y = f(\mathbf{x})$ 라고 하면 다음과 같이 미분이 주어진다.

$$\frac{\partial y}{\partial x} = \frac{\partial f(x)}{\partial x} = f'(x)$$

$$\frac{\partial y}{\partial \mathbf{x}} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \left( \frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_p} \right) = \nabla f(\mathbf{x})$$

함수가 다변수함수일 경우 함수의 값을 각 축의 변수로 미분한 것(partial derivative)을 벡터로 표시하는 것을 gradient 라고 한다.

## 4.2 벡터미분의 표기 방법

이제 다변량함수(multivariate function),  $f: \mathbb{R}^p \rightarrow \mathbb{R}^q$ 에 대한 미분을 생각해보자. 앞 절에서 본것과 같이 스칼라 함수를 여러 변수로 미분하여 partial derivative를 구한 뒤 gradient를 만드는 경우 열벡터와 행벡터 중 하나를 선택해야 한다. 이러한 선택은 절대적인 것이 아니며 각 분야의 특성과 편의에 따라 다르게 선택 될 수 있다.

이제 간단한 예제를 고려해 보자. 두 열벡터  $\mathbf{x} = (x_1, x_2)^t \in \mathbb{R}_2$ ,  $\mathbf{y} = (y_1, y_2, y_3)^t \in \mathbb{R}_3$ 를 고려하고 다음과 같은 함수로 두 벡터의 관계가 정의된다고 하자.

$$y_1 = x_1^2 + x_2, \quad y_2 = \exp(x_1) + 3x_2, \quad y_3 = \sin(x_1) + x_2^3$$

일단 각각의 partial derivative  $\partial y_i / \partial x_j$ 를 구해야 하며 이는 scalar 미분으로 쉽게 구해진다.

$$\begin{array}{lll} \frac{\partial y_1}{\partial x_1} = 2x_1, & \frac{\partial y_2}{\partial x_1} = \exp(x_1), & \frac{\partial y_3}{\partial x_1} = \cos(x_1) \\ \frac{\partial y_1}{\partial x_2} = 1, & \frac{\partial y_2}{\partial x_2} = 3, & \frac{\partial y_3}{\partial x_2} = 3x_2^2 \end{array}$$

통계학에서는 벡터  $\mathbf{y}$ 를 벡터  $\mathbf{x}$ 로 미분하려면 다음과 같이 분모 표기법 (Denominator layout)을 사용하여 표기한다.

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \equiv \frac{\partial \mathbf{y}^t}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \frac{\partial y_3}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_3}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 & \exp(x_1) & \cos(x_1) \\ 1 & 3 & 3x_2^2 \end{bmatrix}$$

즉 분모표기법은 분모를 열벡터로, 분자를 행벡터로 보고 각각 위치에 있는 변수들에 대하여 미분을 표기하는 방법이다.

## 4.3 핵심공식

다음은 분모표기법을 이용한 가장 기본적이고 핵심적인 미분 공식들이다. 공식을 유도하는 경우 분모표기법에서는  $\partial \mathbf{y} / \partial \mathbf{x} \equiv \partial \mathbf{y}^t / \partial \mathbf{x}$  임을 이용한다. 변환이 있거나 여러가지 곱이 있는 경우 미분할 대상 벡터를 가장 왼쪽에 전치형태(즉, 행벡터의 형태로)로 놓는 것이 필요하다. 예를 들어

$$\frac{\partial \mathbf{a}^t \mathbf{V} f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial f(\mathbf{x})^t \mathbf{V}^t \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial f(\mathbf{x})^t}{\partial \mathbf{x}} \mathbf{V}^t \mathbf{a} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \mathbf{V}^t \mathbf{a}$$



또한 행렬은 교환법칙이 성립하지 않기 때문에 연산의 순서를 유지해야 하는 것을 유념하자.

1. 기본행렬 미분 벡터  $\mathbf{c}$ 를 상수벡터라고 하자.

$$\frac{\partial \mathbf{c}}{\partial \mathbf{x}} = \mathbf{0}, \quad \frac{\partial \mathbf{x}}{\partial \mathbf{x}} = \mathbf{I}$$

2. 벡터-스칼라 미분

이 경우는  $\mathbf{x} \in \mathbb{R}^1, \mathbf{y} \in \mathbb{R}^q$  인 경우이며 결과는 다음과 같이 행벡터로 결과가 주어진다.

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \frac{\partial \mathbf{y}^t}{\partial \mathbf{x}} = \left[ \frac{\partial y_1}{\partial \mathbf{x}}, \frac{\partial y_2}{\partial \mathbf{x}}, \dots, \frac{\partial y_q}{\partial \mathbf{x}} \right]$$

3. 스칼라-벡터 미분

이 경우는  $\mathbf{x} \in \mathbb{R}^p, \mathbf{y} \in \mathbb{R}^1$  인 경우이며 결과는 다음과 같이 열벡터로 결과가 주어진다.

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_p} \end{bmatrix}$$

4. 상수벡터와 내적에 대한 미분

열벡터  $\mathbf{a}$ 를  $p \times 1$  상수벡터이라고 하고  $y = \mathbf{a}^t \mathbf{x} = \mathbf{x}^t \mathbf{a}$ 라 하자.

$$\frac{\partial y}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^t \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^t \mathbf{a}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{a}^t \mathbf{x}}{\partial x_1} \\ \frac{\partial \mathbf{a}^t \mathbf{x}}{\partial x_2} \\ \vdots \\ \frac{\partial \mathbf{a}^t \mathbf{x}}{\partial x_p} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \mathbf{a}$$

5. 선형변환에 대한 미분

행렬  $\mathbf{A}$ 를  $q \times p$  행렬이라고 하고  $\mathbf{y} = \mathbf{A}\mathbf{x}$ 라 하자. 여기서 행렬  $\mathbf{A}$ 를 다음과 같이 나타내자.

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^t \\ \mathbf{a}_2^t \\ \vdots \\ \mathbf{a}_q^t \end{bmatrix} \quad \text{or} \quad \mathbf{A}^t = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_q]$$

위의 내적에 대한 미분 결과를 이용하면 다음은 결과를 얻는다.

$$\begin{aligned}
 \frac{\partial y}{\partial x} &= \frac{\partial Ax}{\partial x} \\
 &\equiv \frac{\partial x^t A^t}{\partial x} \\
 &= \frac{\partial}{\partial x} [x^t a_1 \ x^t a_2 \ \cdots \ x^t a_q] \\
 &= \left[ \frac{\partial x^t a_1}{\partial x} \ \frac{\partial x^t a_2}{\partial x} \ \cdots \ \frac{\partial x^t a_q}{\partial x} \right] \\
 &= [a_1 \ a_2 \ \cdots \ a_q] \\
 &= A^t
 \end{aligned}$$

위의 결과를 응용하면 다음의 결과를 얻는다.

$$\frac{\partial Ax}{\partial x} = A^t \quad \text{and} \quad \frac{\partial x^t A}{\partial x} = A$$

6. 이차형식

$$\frac{\partial x^t Ax}{\partial x} = \frac{\partial x^t}{\partial x} Ax + \frac{\partial x^t A^t}{\partial x} x = Ax + A^t x$$

만약 행렬  $A$ 가 대칭이면

$$\frac{\partial x^t Ax}{\partial x} = 2Ax$$