4.10 다변량 회귀분석 - 수학적 모델

- 입력에 사용되는 특징이 하나가 아니라 n개라면 선형 회귀 모델은 다음과 같은 식으로 표현할 수 있으며 다변량multivariate 회귀분석이라 함

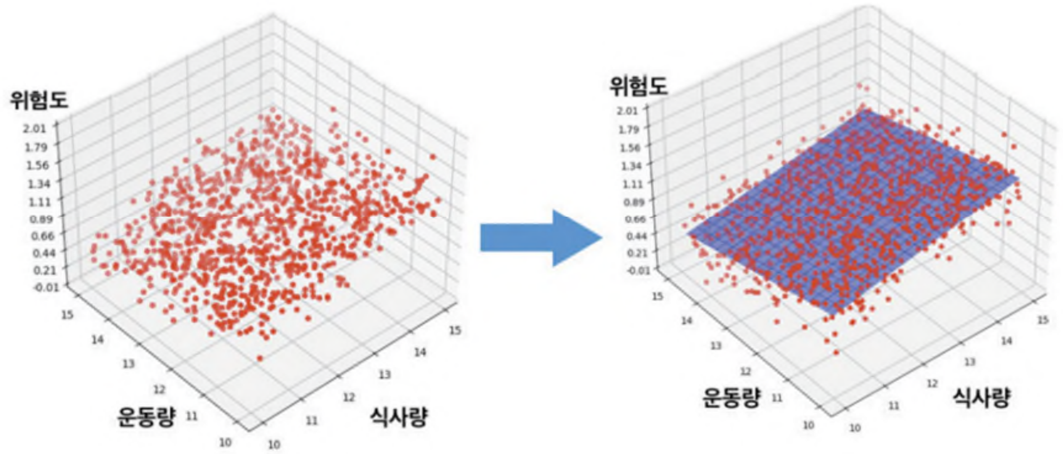
$$\hat{y} = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$$

$$\hat{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & \dots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} & \dots & x_n^{(m)} \end{bmatrix} \quad w = \begin{bmatrix} b \\ w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

$$\boxed{\hat{y} = XW}$$

사람들의 식사량과 운동량에 따른 건강 위험도(건강검진 결과) 관계 분석

- 식사량과 운동량이 기록된 사람들에 대해 건강검진을 통해 건강위험도 정보를 측정할 수 있었다면 아래 그림 왼쪽과 같이 운동량과 식사량에 따라 건강위험도가 결정되는 어떤 관계가 있을 것이라고 추측
- 이때 이 관계가 선형 관계라는 가설을 세웠다면 선형 회귀는 이 데이터를 설명할 수 있는 평면을 찾는 것



- 수식으로 표현하면 식사량을 x_1, 운동량을 x_2라고 할 수 있고, 위험도 y가 다음과 같은 식으로 설명될 수 있다는 것

$$\hat{y} = w_1 x_1 + w_2 x_2 + b = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + b$$

- 이것을 m개의 데이터로 표현하면…

$$X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} \\ x_1^{(2)} & x_2^{(2)} \\ \vdots & \\ x_1^{(m)} & x_2^{(m)} \end{bmatrix} \quad W = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \quad b = b \quad Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

$$Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} = X \cdot W + b = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ x_1^{(3)} & x_2^{(3)} & \cdots & x_n^{(3)} \\ \vdots & \vdots & \cdots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} + b$$

- 수식으로 표현하면 식사량을 x_1, 운동량을 x_2라고 할 수 있고, 위험도 y가 다음과 같은 식으로 설명될 수 있다는 것

$$\hat{y} = w_1 x_1 + w_2 x_2 + \boxed{b} = \begin{bmatrix} x_1 & x_2 & \boxed{1} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \boxed{b} \end{bmatrix}$$

- 이것을 m개의 데이터로 표현하면…

$$X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & 1 \\ x_1^{(2)} & x_2^{(2)} & 1 \\ \vdots & & \\ x_1^{(m)} & x_2^{(m)} & 1 \end{bmatrix} \quad W = \begin{bmatrix} w_1 \\ w_2 \\ \underbrace{\boxed{b}}_{w_0} \end{bmatrix} \quad Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

$$Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ \vdots \\ y^{(m)} \end{bmatrix} = X \cdot W = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & \cdots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{bmatrix} \begin{bmatrix} \boxed{w_0} \\ w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

◆ 수식으로 표현하면 식사량을 x_1, 운동량을 x_2라고 할 수 있고, 위험도 y가 다음과 같은 식으로 설명될 수 있다는 것

$$\boxed{\hat{y} = w_1x_1 + w_2x_2 + b = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b}$$

$$\hat{y} = w_1x_1 + w_2x_2 + b = [w_1\ w_2\ b]\begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

◆ 이때 w는 w_1, w_2를 원소로 하는 벡터이고, x는 입력 특징 x_1, x_2를 원소로 하는 벡터

◆ 다변량 회귀분석은 이때 특징 벡터 x가 임의의 차원을 가질 수 있는 것을 의미하며, 찾아야 하는 것은 n차원 공간의 초평면hyperplane이 되는 것

◆ 이것을 파라미터parameter 벡터로 표현할 수도 있음

- 이 모델의 동작을 결정하는 변수는 w 와 b
- 이 둘을 따로 구분하지 않고 하나의 파라미터 벡터 θ로 표현한다면 w_i는 θ_i로 표현할 수 있고, b는 θ_0라고 하면 될 것 (여기서 x_0 = 1)
- 그러면 선형 회귀는 다음과 같은 식으로 표현할 수 있음

$$\boxed{\hat{y} = \theta_0x_0 + \theta_1x_1 + \theta_2x_2 + \cdots + \theta_nx_n = X\theta}$$

- 따라서 선형 회귀분석의 가설 $y = ax + b$ 는 아래와 같이 간략히 표현할 수 있다.

$$Y = X\theta \quad \text{또는} \quad Y = XW$$

- 다변량 회귀분석은 이 때 특징 벡터 x가 임의의 차원을 가질 수 있는 것을 의미하며, 찾아야 하는 것은 n차원 공간의 초평면(hypoerplane)이 되는 것

◆ 데이터의 인스턴스의 개수가 m이라고 하면 i번째 인스턴스는 $x^{(m)}$으로 나타낼 수 있고, $n+1$개의 원소를 가진 다음과 같은 벡터로 표현할 수 있다. 이 데이터 인스턴스에 대응하는 레이블이 $y^{(i)}$이다

$$x^{(i)} = (1, x_1^{(i)}, x_2^{(i)}, x_1^{(i)}, \dots, x_n^{(i)})$$

$$x^{(1)} = (1,\ x_1^{(1)},\ x_2^{(1)},\ x_3^{(1)},\ \dots,\ x_n^{(1)})$$
$$x^{(2)} = (1,\ x_1^{(2)},\ x_2^{(2)},\ x_3^{(2)},\ \dots,\ x_n^{(2)})$$
$$\dots$$
$$x^{(m)} = (1,\ x_1^{(m)},\ x_2^{(m)},\ x_3^{(m)}, \dots, x_n^{(m)})$$

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & \dots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} & \dots & x_n^{(m)} \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & \dots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} & \dots & x_n^{(m)} \end{bmatrix}, \quad Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ \vdots \\ y^{(m)} \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$$

$$Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ \vdots \\ y^{(m)} \end{bmatrix} = X \cdot \theta = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & \dots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} & \dots & x_n^{(m)} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$$

$$y^{(1)} = 1 \cdot \theta_0 + x_1^{(1)}\theta_1 + x_2^{(1)}\theta_2 + \cdots + x_n^{(1)}\theta_n$$
$$y^{(2)} = 1 \cdot \theta_0 + x_1^{(2)}\theta_1 + x_2^{(2)}\theta_2 + \cdots + x_n^{(2)}\theta_n$$
$$\dots$$
$$y^{(m)} = 1 \cdot \theta_0 + x_1^{(m)}\theta_1 + x_2^{(m)}\theta_2 + \cdots + x_n^{(m)}\theta_n$$

행렬 표현으로 바꾸기(Y=WX+b)

- 3개의 분류기를 하나의 수식으로 묶어서 간단하게 표현
  > w들을 열벡터로 표현해 class 수 x 특징 차원 수 행'렬로 표현 (3x2 행렬)

$$\begin{bmatrix} w_{a_1} & w_{a_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b_a = w_{a_1}x_1 + w_{a_2}x_2 + b_a$$

$$\begin{bmatrix} w_{b_1} & w_{b_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b_b = w_{b_1}x_1 + w_{b_2}x_2 + b_b$$

$$\begin{bmatrix} w_{c_1} & w_{c_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b_c = w_{c_1}x_1 + w_{c_2}x_2 + b_c$$

⇩

$$\begin{bmatrix} w_{a_1} & w_{a_2} \\ w_{b_1} & w_{b_2} \\ w_{c_1} & w_{c_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_a \\ b_b \\ b_c \end{bmatrix} = \begin{bmatrix} w_{a_1}x_1 + w_{a_2}x_2 + b_a \\ w_{b_1}x_1 + w_{b_2}x_2 + b_b \\ w_{c_1}x_1 + w_{c_2}x_2 + b_c \end{bmatrix}$$

행렬 표현으로 바꾸기(Y=WX+b)

- 클래스 3, 데이터 4개인 경우 → 데이터를 열벡터로 표기
- 효과적인 행렬 연산을 위해 X와 W에 대한 곱셈 위치 변경

$$\begin{bmatrix} w_{a1} & w_{a2} \\ w_{b1} & w_{b2} \\ w_{c1} & w_{c2} \end{bmatrix} \begin{bmatrix} x_{11} & x_{21} & x_{31} & x_{41} \\ x_{12} & x_{22} & x_{32} & x_{42} \end{bmatrix} + \begin{bmatrix} b_1 & b_1 & b_1 & b_1 \\ b_2 & b_2 & b_2 & b_2 \\ b_3 & b_3 & b_3 & b_3 \end{bmatrix}$$

shape : $[\boxed{3} \times \boxed{2] \times [2} \times \boxed{4}] + [3 \times 4] = [3 \times 4]$

$$= \begin{bmatrix} x_{11}w_{a_1} + x_{12}w_{a_1} + b_1 & x_{21}w_{a_1} + x_{22}w_{a_1} + b_1 & x_{31}w_{a_1} + x_{32}w_{a_1} + b_1 & x_{41}w_{a1} + x_{42}w_{a_2} + b_1 \\ x_{11}w_{b1} + x_{12}w_{b2} + b_2 & x_{21}w_{b1} + x_{22}w_{b2} + b_2 & x_{31}w_{b_1} + x_{32}w_{b2} + b_2 & x_{41}w_{b1} + x_{22}w_{b_2} + b_2 \\ x_{11}w_{c1} + x_{12}w_{c2} + b_3 & x_{21}w_{c1} + x_{22}w_{c_2} + b_3 & x_{31}w_{c1} + x_{32}w_{c_2} + b_3 & x_{41}w_{c1} + x_{22}w_{c_2} + b_3 \end{bmatrix}$$

행렬 표현으로 바꾸기(Y=XW+b)

- 클래스 3, 데이터 4개인 경우 → 데이터를 행벡터로 표기
- 열벡터로 정의된 w는 샘플 수에 따라서 행렬의 크기가 변화되지 않고 고정됨

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{41} & x_{42} \end{bmatrix} \begin{bmatrix} w_{a_1} & w_{b_1} & w_{c_1} \\ w_{a_2} & w_{b_2} & w_{c_2} \end{bmatrix} + \begin{bmatrix} b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \end{bmatrix}$$

$$[4 \times 2] \times [2 \times 3] + [4 \times 3] = [4 \times 3]$$

$$= \begin{bmatrix} x_{11}w_{a1} + x_{12}w_{a2} + b_1 & x_{11}w_{b1} + x_{12}w_{b2} + b_2 & x_{11}w_{c1} + x_{12}w_{c2} + b_3 \\ x_{21}w_{a1} + x_{22}w_{a2} + b_1 & x_{21}w_{b1} + x_{22}w_{b2} + b_2 & x_{21}w_{c1} + x_{22}w_{c2} + b_3 \\ x_{31}w_{a1} + x_{32}w_{a2} + b_1 & x_{31}w_{b1} + x_{32}w_{b2} + b_2 & x_{31}w_{c1} + x_{32}w_{c2} + b_3 \\ x_{41}w_{a1} + x_{42}w_{a2} + b_1 & x_{41}w_{b1} + x_{42}w_{b2} + b_2 & x_{41}w_{c1} + x_{42}w_{c2} + b_3 \end{bmatrix}$$

머신러닝(선형회귀)를 통해 하고자 하는 일은 아래의 가설에서 W 값을 찾는 일이며

$$Y = XW$$

```
y_hat = np.matmul(X, W)
```

여기서 가설을 H(X) 라고 표현한다면 아래와 같이 쓸 수 있고…

$$H(X) = XW$$

```
hypothesis = np.dot(X, W)
```

가설을 평가하기 위한 Cost 함수를 오차제곱평균(Mean Squared Error)로 사용할 경우 아래와 같이 정의할 수 있다.

$$MSE(W) = \frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)}W - y^{(i)}\right)^2$$

$$= mean\{(XW - y)^2\}$$

```
mse = ((y_hat - y)**2).mean()
```

- 가설은 파라미터 θ에 의해 동작이 결정되므로, m개의 데이터 인스턴스에 대해 현재 가설의 평균 제곱 오차mean squared error를 다음과 같이 구할 수 있음
  - 각 데이터 인스턴스에 대해 구한 오차를 제곱하여 평균한 값
  - 다변량 선형 회귀 분석은 결국 이 오차가 최소가 되는 파라미터 θ를 찾는 문제

$$E_{mse}(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)}\theta - y^{(i)}\right)^2$$

4.11 회귀분석의 학습, 혹은 최적화 방법 - 정규 방정식

- 다변량 선형 회귀에서 가설과 레이블의 평균 제곱 오차를 구했다면, 오차를 줄이는 방향으로 파라미터를 수정하는 것이 학습
  - 특정 파라미터 θ_j에 대해 이 오차를 편미분하면 다음을 얻을 수 있음.

$$\frac{\partial E_{mse}(\theta)}{\partial \theta_j} = \frac{2}{m}\sum_{i=1}^{m}\left(x^{(i)}\theta - y^{(i)}\right)x_j^{(i)}$$

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & \cdots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \quad Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

- 오차 곡면이 파라미터에 대해 가지는 기울기 $\nabla_\theta \boldsymbol{E}_{mse}$의 j번째 원소가 됨
  - 파라미터는 다음과 같이 변경할 수 있음

$$\theta \leftarrow \theta - \eta \nabla_\theta E_{mse}$$