Vectorization in feature scaling

Saturday, December 17, 2022   2:43 PM

$$x_j^{(i)'} = \frac{x_j^{(i)} - \mu_j}{\sigma_j} \qquad \mu_j = \frac{1}{n}\sum_{i=1}^{n} x_j^{(i)} \qquad \sigma_j = SD(x_j^{(1)} \cdots x_j^{(n)})$$

$$X = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_m^{(1)} \\ \vdots & & & \vdots \\ x_1^{(n)} & \cdots & \cdots & x_m^{(n)} \end{pmatrix}$$

$$\begin{matrix} \downarrow & \downarrow & \cdots & \downarrow \\ \mu_1 & \mu_2 & & \mu_m \\ \sigma_1 & \sigma_2 & & \sigma_m \end{matrix}$$

$\mu =$ np.mean(X, axis=0) 可以进行逐列求mean

$\sigma =$ np.std(X, axis=0) 逐列求 stand deviation

为了求 $\frac{x_j^{(i)} - \mu_j}{\sigma_j}$，这里看 broadcasting

$X: (n, m)$ $\mu: (m,)$ $\sigma (m,)$

以 $X-\mu$ 为例： $X:(n,m)$, $\mu:(m,)$ $\rightarrow$ $X:(n,m)$, $\mu:(1,m)$ $\rightarrow$ $X:(n,m)$, $\mu:(n,m)$ $\rightarrow$ $X-\mu = \begin{pmatrix} x_0^{(0)}-\mu_0, & x_1^{(0)}-\mu_1 \\ \vdots & \vdots \\ x_0^{(n-1)}-\mu_0, & x_1^{(n-1)}-\mu_1 \end{pmatrix}$

$$\begin{pmatrix} \mu_0 & \mu_1 \\ \mu_0 & \mu_1 \\ \vdots & \\ \mu_0 & \mu_1 \end{pmatrix}$$

Broadcast rule: 在dim小的array的shape 左边补上"1"，然后将这一些"1"变成 dim大的array中相应位置的值 (如果可以的话)

这是我们想要的，因为 我们想让第一列减去 μ_0，第二列减去 μ_1

Similarly, $\frac{X-\mu}{\sigma}$ 将会令第一列除以 σ_0，第二列除以 σ_1