<Vectorization with code 2> - Gradient Descent

$\vec{w} = (w_1 \; w_2 \cdots w_{16})$ ⇒ Parameter w

$\vec{d} = (d_1 \; d_2 \cdots d_{16})$ ⇒ derivative d

features = 16 , b = 0

code :

```
w=np.array([0.5, 1.3, ...3.4])
d=np.array([0.3, 0.2, ... 0.4])
```

compute ⇒ $w_j := w_j - \underbrace{0.1}_{\text{learning rate}} d_j$ (for $j = 1 \ldots 16$)

① Without Vectorization

$w_1 := w_1 - 0.1 d_1$

$w_2 := w_2 - 0.1 d_2$

$\vdots$

$w_{16} := w_{16} - 0.1 d_{16}$

code :

```
for j in range(0,16):
    w[j] = w[j] - 0.1 * d[j]
```

② With Vectorization

$\vec{w} := \vec{w} - 0.1 \vec{d}$

code :

```
w = w - 0.1 d
```

⇓

<One feature>

Model: $f_{w,b}(x) = wx + b$

(univariate linear regression)

Gradient Descent :

repeat {

$$w := w - \alpha \underbrace{\frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right) x^{(i)}}_{= \frac{\partial}{\partial w} J(w,b)}$$

$$b := b - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right)$$

simultaneously update w, b

}

<multiple features>

Model: $f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$

(multivariate linear regression)

Gradient Descent :

repeat {

$$(j=1) \quad w_1 := w_1 - \alpha \underbrace{\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right) x_1^{(i)}}_{= \frac{\partial}{\partial w_1} J(\vec{w},b)}$$

$\vdots \qquad \vdots$

$$(j=n) \quad w_n := w_n - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right) x_n^{(i)}$$

$$b := b - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)$$

simultaneously update

w_j (for $j = 1, \cdots, n$) and b

}