

Previous notation

Parameters w_1, \dots, w_n
 b

Model $f_{\vec{w},b}(\vec{x}) = w_1 x_1 + \dots + w_n x_n + b$

Cost function $J(\underbrace{w_1, \dots, w_n}_b)$

Vector notation

$\vec{w} = [w_1 \dots w_n]$ ← vector of length n
 b still a number

$f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$
 $J(\vec{w}, b)$ ↑ dot product

Gradient descent

repeat {
 $w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\underbrace{w_1, \dots, w_n}_b)$
 $b = b - \alpha \frac{\partial}{\partial b} J(\underbrace{w_1, \dots, w_n}_b)$
}

repeat {
 $w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{w}, b)$
 $b = b - \alpha \frac{\partial}{\partial b} J(\vec{w}, b)$
}

→ Previous vs vector notation.

Gradient descent

One feature
 repeat {

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

$\frac{\partial}{\partial w} J(w, b)$

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

simultaneously update w, b

}

n features ($n \geq 2$)

repeat {

$$\underbrace{w_1}_j = w_1 - \alpha \frac{1}{m} \sum_{i=1}^m (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}) \underbrace{x_1^{(i)}}_j$$

$\frac{\partial}{\partial w_1} J(\vec{w}, b)$

\vdots

$$\underbrace{w_n}_j = w_n - \alpha \frac{1}{m} \sum_{i=1}^m (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}) \underbrace{x_n^{(i)}}_j$$

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

simultaneously update
 w_j (for $j = 1, \dots, n$) and b

}

→ Gradient Descent, 1 feature vs n feature.

An alternative to gradient descent

→ Normal equation

- Only for linear regression
- Solve for w , b without iterations

Disadvantages

- Doesn't generalize to other learning algorithms.
- Slow when number of features is large ($> 10,000$)

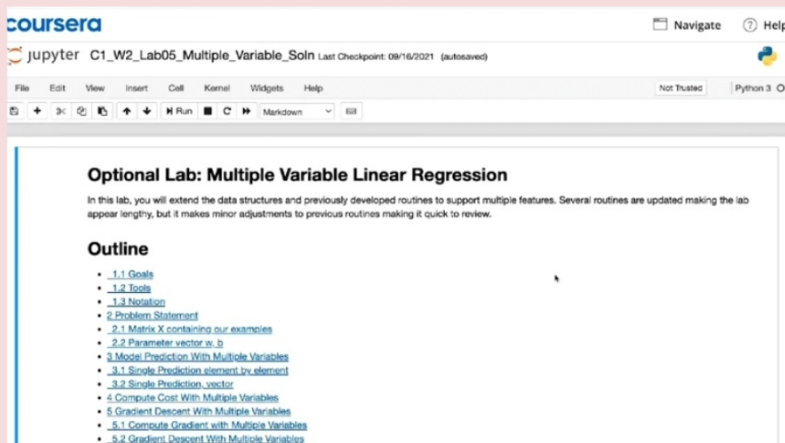
What you need to know

- Normal equation method may be used in machine learning libraries that implement linear regression.
- Gradient descent is the recommended method for finding parameters w, b

→ An alternative of gradient descent, only for linear regression.

$$w = \frac{(x_1 - \bar{x})(y_1 - \bar{y})}{(x_1 - \bar{x})^2}$$

$$b = \bar{y} - w\bar{x}$$



→ optional lab practice.