

# Lecture 4

## Multi-variable linear regression

Sung Kim <[hunkim+mr@gmail.com](mailto:hunkim+mr@gmail.com)>

# Recap

- Hypothesis
- Cost function
- Gradient descent algorithm

# Recap

- Hypothesis

$$H(x) = Wx + b$$

- Cost function

$$cost(W, b) = \frac{1}{m} \sum_{i=1}^m (H(x^{(i)}) - y^{(i)})^2$$

- Gradient descent algorithm

# Predicting exam score: regression using one input ( $x$ )

one-variable  
one-feature

| x (hours) | y (score) |
|-----------|-----------|
| 10        | 90        |
| 9         | 80        |
| 3         | 50        |
| 2         | 60        |
| 11        | 40        |

# Predicting exam score: regression using two inputs ( $x_1$ , $x_2$ )

multi-variable/feature

| $x_1$ (hours) | $x_2$ (attendance) | $y$ (score) |
|---------------|--------------------|-------------|
| 10            | 5                  | 90          |
| 9             | 5                  | 80          |
| 3             | 2                  | 50          |
| 2             | 4                  | 60          |
| 11            | 1                  | 40          |

# Hypothesis

$$H(x) = Wx + b$$

# Hypothesis

$$H(x) = Wx + b$$

$$H(x_1, x_2) = w_1x_1 + w_2x_2 + b$$

# Cost function

$$H(x_1, x_2) = w_1x_1 + w_2x_2 + b$$

$$\text{cost}(W, b) = \frac{1}{m} \sum_{i=1}^m (H(x^{(i)}) - y^{(i)})^2$$

# Multi-variable

$$H(x_1, x_2) = w_1x_1 + w_2x_2 + b$$

$$H(x_1, x_2, x_3, \dots, x_n) = w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n + b$$

# Matrix

$$w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n$$

# Matrix

$$w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n$$

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3 \end{bmatrix}$$

# Matrix multiplication

"Dot Product"

The diagram shows the multiplication of two matrices. The first matrix is  $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$  and the second matrix is  $\begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}$ . A yellow arrow labeled "Dot Product" points from the first row of the first matrix to the first column of the second matrix. The result is shown as  $\begin{bmatrix} 58 \\ \phantom{0} \end{bmatrix}$ , with the number 58 highlighted in a yellow circle.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix} = \begin{bmatrix} 58 \\ \phantom{0} \end{bmatrix}$$

<https://www.mathsisfun.com/algebra/matrix-multiplying.html>

# Hypothesis

$$[w_1 \quad w_2 \quad w_3] \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = [w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3] :$$

$$H(X) = WX + b$$

# Hypothesis without $b$

$$[b \quad w_1 \quad w_2 \quad w_3] \times \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = [b \times 1 + w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3]$$

$$H(X) = WX$$

# W vs X

$$[b \quad w1 \quad w2 \quad w3] \times \begin{bmatrix} 1 \\ x1 \\ x2 \\ x3 \end{bmatrix} = [b \times 1 + w1 \times x1 + w2 \times x2 + w3 \times x3]$$

$$H(X) = WX$$

# Transpose

$$\begin{bmatrix} 6 & 4 & 24 \\ 1 & -9 & 8 \end{bmatrix}^T = \begin{bmatrix} 6 & 1 \\ 4 & -9 \\ 24 & 8 \end{bmatrix}$$

<http://www.mathsisfun.com/algebra/matrix-introduction.html>

# Hypothesis using Transpose

$$[b \quad w_1 \quad w_2 \quad w_3] \times \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = [b \times 1 + w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3]$$

$$H(X) = W^T X$$

**Next**  
**Logistic Regression**  
**(Classification)**

