

LECTURE 4

MULTIVARIABLE LOGISTIC REGRESSION

Sung Kim <hunkim+ml@gmail.com>
<http://hunkim.github.io/ml>

NAVER | Clova



Recap

01. Hypothesis

02. Cost function

03. Gradient descent algorithm

Recap

01. Hypothesis

- $H(x) = Wx + b$

02. Cost function

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

03. Gradient descent algorithm

Predicting Exam Score

Regression using one input(X)
One-variable/feature

X (hours)	Y (score)
10	90
9	80
3	50
2	60
11	40

Predicting Exam Score

Regression using three inputs (X_1, X_2, X_3)
Multi-variable/feature

Test Scores for General Psychology

X_1	X_2	X_3	Y
73	80	75	152
93	88	93	185
89	91	90	180
96	98	100	196
73	66	70	142

Hypothesis

$$h(x) = wx + b$$

Hypothesis

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

$$H(X_1, X_2, X_3) = W_1X_1 + W_2X_2 + W_3X_3 + b$$

Cost Function

$$H(X_1, X_2, X_3) = W_1X_1 + W_2X_2 + W_3X_3 + b$$

$$\text{cost}(W, b) = \frac{1}{m} \sum_{i=1}^m (H(X_1^{(i)}, X_2^{(i)}, X_3^{(i)}) - Y^{(i)})^2$$

Multi-variable

$$H(X_1, X_2, X_3) = W_1X_1 + W_2X_2 + W_3X_3 + b$$

$$H(X_1, X_2, X_3, \dots, X_n) = W_1X_1 + W_2X_2 + W_3X_3 \dots + W_nX_n + b$$

Matrix

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

Matrix Multiplication

"Dot Product"

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

Hypothesis Using Matrix

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

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} x_1w_1 + x_2w_2 + x_3w_3 \end{bmatrix}$$

$$H(X) = XW$$

Many x Instances

Test Scores for General Psychology

X ₁	X ₂	X ₃	Y
73	80	75	152
93	88	93	185
89	91	90	180
96	98	100	196
73	66	70	142

Hypothesis Using Matrix

$$H(X_1, X_2, X_3) = X_1W_1 + X_2W_2 + X_3W_3$$

Test Scores for General Psychology

X ₁	X ₂	X ₃	Y
73	80	75	152
93	88	93	185
89	91	90	180
96	98	100	196
73	66	70	142

Hypothesis Using Matrix

$$H(X_1, X_2, X_3) = X_1W_1 + X_2W_2 + X_3W_3$$

Test Scores for General Psychology

X_1	X_2	X_3	Y
73	80	75	152
93	88	93	185
89	91	90	180
96	98	100	196
73	66	70	142

$$\begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix} \cdot \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} X_1W_1 + X_2W_2 + X_3W_3 \end{bmatrix}$$

$$H(X) = XW$$

Hypothesis Using Matrix

$$W_1X_1 + W_2X_2 + W_3X_3 + \dots + W_nX_n$$

Test Scores for General Psychology

X ₁	X ₂	X ₃	Y
73	80	75	152
93	88	93	185
89	91	90	180
96	98	100	196
73	66	70	142

$$\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \\ X_{41} & X_{42} & X_{43} \\ X_{51} & X_{52} & X_{53} \end{bmatrix} \cdot \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} X_{11}W_1 + X_{12}W_2 + X_{13}W_3 \\ X_{21}W_1 + X_{22}W_2 + X_{23}W_3 \\ X_{31}W_1 + X_{32}W_2 + X_{33}W_3 \\ X_{41}W_1 + X_{42}W_2 + X_{43}W_3 \\ X_{51}W_1 + X_{52}W_2 + X_{53}W_3 \end{bmatrix}$$

$$H(X) = XW$$

Hypothesis Using Matrix

A diagram illustrating matrix multiplication. It shows three matrices represented by rectangles with brackets on their sides. The first rectangle is labeled $[5, 3]$ below it. To its right is a multiplication symbol \times . The second rectangle is labeled $[?, ?]$ below it. To its right is an equals sign $=$. The third rectangle is labeled $[5, 1]$ below it.

$$H(X) = XW$$

Hypothesis Using Matrix

$$\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \\ X_{41} & X_{42} & X_{43} \\ X_{51} & X_{52} & X_{53} \end{bmatrix} \times \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} X_{11}W_1 + X_{12}W_2 + X_{13}W_3 \\ X_{21}W_1 + X_{22}W_2 + X_{23}W_3 \\ X_{31}W_1 + X_{32}W_2 + X_{33}W_3 \\ X_{41}W_1 + X_{42}W_2 + X_{43}W_3 \\ X_{51}W_1 + X_{52}W_2 + X_{53}W_3 \end{bmatrix}$$

$[5, 3] \qquad \qquad [3, 1] \qquad \qquad [5, 1]$

$$H(X) = XW$$

Hypothesis Using Matrix

$$\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \\ X_{41} & X_{42} & X_{43} \\ X_{51} & X_{52} & X_{53} \end{bmatrix} \times \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} X_{11}W_1 + X_{12}W_2 + X_{13}W_3 \\ X_{21}W_1 + X_{22}W_2 + X_{23}W_3 \\ X_{31}W_1 + X_{32}W_2 + X_{33}W_3 \\ X_{41}W_1 + X_{42}W_2 + X_{43}W_3 \\ X_{51}W_1 + X_{52}W_2 + X_{53}W_3 \end{bmatrix}$$

$[n, 3] \qquad \qquad [3, 1] \qquad \qquad [n, 1]$

$$H(X) = XW$$

Hypothesis Using Matrix (n output)

$$\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \\ X_{41} & X_{42} & X_{43} \\ X_{51} & X_{52} & X_{53} \end{bmatrix} \times \begin{bmatrix} \phantom{X_{11}W_{11} + X_{12}W_{21} + X_{13}W_{31}} \\ \phantom{X_{21}W_{11} + X_{22}W_{21} + X_{23}W_{31}} \\ \phantom{X_{31}W_{11} + X_{32}W_{21} + X_{33}W_{31}} \\ \phantom{X_{41}W_{11} + X_{42}W_{21} + X_{43}W_{31}} \\ \phantom{X_{51}W_{11} + X_{52}W_{21} + X_{53}W_{31}} \end{bmatrix} = \begin{bmatrix} X_{11}W_{11} + X_{12}W_{21} + X_{13}W_{31} & X_{11}W_{12} + X_{12}W_{22} + X_{13}W_{32} \\ X_{21}W_{11} + X_{22}W_{21} + X_{23}W_{31} & X_{21}W_{12} + X_{22}W_{22} + X_{23}W_{32} \\ X_{31}W_{11} + X_{32}W_{21} + X_{33}W_{31} & X_{31}W_{12} + X_{32}W_{22} + X_{33}W_{32} \\ X_{41}W_{11} + X_{42}W_{21} + X_{43}W_{31} & X_{41}W_{12} + X_{42}W_{22} + X_{43}W_{32} \\ X_{51}W_{11} + X_{52}W_{21} + X_{53}W_{31} & X_{51}W_{12} + X_{52}W_{22} + X_{53}W_{32} \end{bmatrix}$$

$[n, 3] \qquad \qquad [?, ?] \qquad \qquad [n, 2]$

$$H(X) = XW$$

Hypothesis Using Matrix(n output)

$$\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \\ X_{41} & X_{42} & X_{43} \\ X_{51} & X_{52} & X_{53} \end{bmatrix} \times \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \\ W_{31} & W_{32} \end{bmatrix} = \begin{bmatrix} X_{11}W_{11} + X_{12}W_{21} + X_{13}W_{31} & X_{11}W_{12} + X_{12}W_{22} + X_{13}W_{32} \\ X_{21}W_{11} + X_{22}W_{21} + X_{23}W_{31} & X_{21}W_{12} + X_{22}W_{22} + X_{23}W_{32} \\ X_{31}W_{11} + X_{32}W_{21} + X_{33}W_{31} & X_{31}W_{12} + X_{32}W_{22} + X_{33}W_{32} \\ X_{41}W_{11} + X_{42}W_{21} + X_{43}W_{31} & X_{41}W_{12} + X_{42}W_{22} + X_{43}W_{32} \\ X_{51}W_{11} + X_{52}W_{21} + X_{53}W_{31} & X_{51}W_{12} + X_{52}W_{22} + X_{53}W_{32} \end{bmatrix}$$

$[n, 3] \qquad [3, 2] \qquad [n, 2]$

$$H(X) = XW$$

WX vs XW

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

Lecture
(Theory)

$$H(x) = XW$$

Implementation
(TensorFlow)

Hypothesis without b

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

$$H(X) = WX$$

W vs X

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

$$H(X) = WX$$

Transpose

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

Hypothesis Using Transpose

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

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

NEXT LECTURE

LOGISTIC REGRESSION (CLASSIFICATION)