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
 - $\cdot H(x) = Wx + b$
- 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 (X1,X2,X3)
Multi-variable/feature

Test Scores for General Psychology

x ₁	X ₂	X3	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

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

$$H(X_{1}, X_{2}, X_{3}) = W_{1}X_{1} + W_{2}X_{2} + W_{3}X_{3} + b$$

Cost Function

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

cost(W,b) =
$$\frac{1}{m} \sum_{i=1}^{m} (H(X1^{(i)}, X2^{(i)}, X3^{(i)}) - Y^{(i)})^2$$

Multi-variable

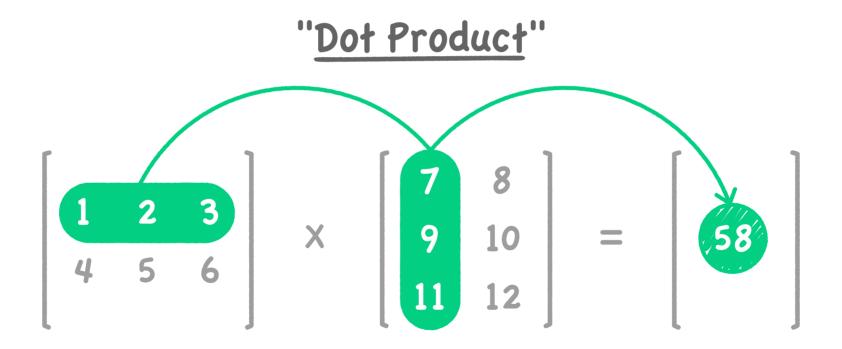
$$H(X_1, X_2, X_3) = W_1X_1 + W_2X_2 + W_3X_3 + b$$

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

Matrix

$$W1X1 + W2X2 + W3X3 + \cdots + WnXn$$

Matrix Multiplication



$$W_1X_1 + W_2X_2 + W_3X_3 + \cdots + W_nX_n$$

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

$$H(X) = XW$$

Many x Instances

x ₁	X ₂	X3	Y
73	80	75	152
93	88	93	185
89	91	90	180
96	98	100	196
73	66	70	142

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

x ₁	X ₂	X3	Y
73	80	75	152
93	88	93	185
89	91	90	180
96	98	100	196
73	66	70	142

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

x ₁	X ₂	X3	Y
73	80	75	152
93	88	93	185
89	91	90	180
96	98	100	196
73	66	70	142

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

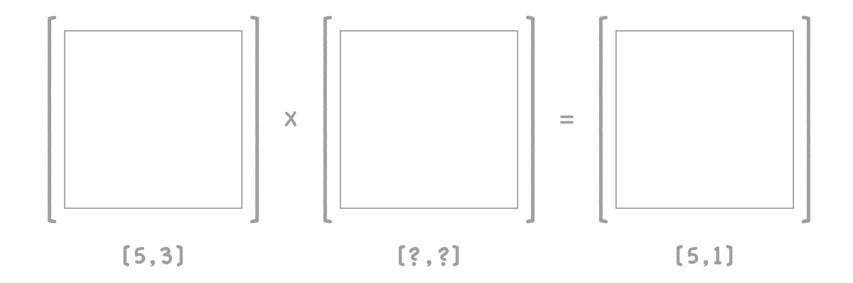
$$H(X) = XW$$

$$W_1X_1 + W_2X_2 + W_3X_3 + \cdots + W_nX_n$$

x ₁	X ₂	X3	Y
73	80	75	152
93	88	93	185
89	91	90	180
96	98	100	196
73	66	70	142

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

$$H(X) = XW$$



$$H(X) = XW$$

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

$$\begin{bmatrix} 5, 3 \end{bmatrix}$$

$$\begin{bmatrix} 3, 1 \end{bmatrix}$$

$$\begin{bmatrix} 5, 1 \end{bmatrix}$$

$$H(X) = XW$$

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

Hypothesis Using Matrix (n output)

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

$$\begin{bmatrix} n, 3 \end{bmatrix} \begin{bmatrix} ?,? \end{bmatrix}$$

$$H(X) = XW$$

Hypothesis Using Matrix (n output)

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

$$[n,3]$$

$$[3,2]$$

$$[n,2]$$



WX vs XW

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

$$H(x)=XW$$

Lecture (Theory)

Implementation (TensorFlow)

Hypothesis without b

$$\begin{bmatrix} b & W_1 & W_2 & W_3 \end{bmatrix} \times \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} b \times 1 + W_1 \times X_1 + W_2 \times X_2 + W_3 \times X_3 \end{bmatrix}$$

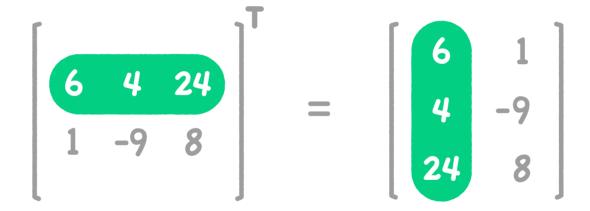
$$H(X) = WX$$

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



Hypothesis Using Transpose

$$\begin{bmatrix} b & W_1 & W_2 & W_3 \end{bmatrix} \times \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} b \times 1 + W_1 \times X_1 + W_2 \times X_2 + W_3 \times X_3 \end{bmatrix}$$

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

NEXT LECTURE

LOGISTIC REGRESSION (CLASSIFICATION)