Simple Linear Models

Itthi Chatnuntawech

Optional: MLR with matrices

$$y_i = w_0 + \underline{w_1 x_{i1} + w_2 x_{i2} + \ldots + w_p x_{ip}}$$

features of sample i

parameters

$$\mathbf{x_i} = \begin{array}{cl} \left[\begin{matrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{matrix}\right] & \begin{matrix} \text{feature 1} \\ \text{feature 2} \\ \\ \text{feature p} \end{matrix} \end{array} \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$$

$$\mathbf{x_i}^T \mathbf{w} = \begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{ip} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} = w_1 x_{i1} + w_2 x_{i2} + \ldots + w_p x_{ip}$$

Brain Code Camp

Optional: MLR with matrices

$$y_i = w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_p x_{ip}$$

features of sample i

parameters

$$\mathbf{x_i} = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} \begin{array}{l} \text{feature 1} \\ \text{feature 2} \\ \\ \text{feature p} \end{array} \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$$

$$\mathbf{x_i}^T\mathbf{w} = \begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{ip} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} = w_1 x_{i1} + w_2 x_{i2} + \ldots + w_p x_{ip}$$

features of sample i

parameters

$$\mathbf{x_i} = \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} \begin{array}{l} \\ \text{feature 1} \\ \text{feature 2} \\ \\ \text{feature p} \end{array} \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$$

$$\mathbf{x_i}^T\mathbf{w} = \begin{bmatrix} 1 & x_{i1} & x_{i2} & \cdots & x_{ip} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} = w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_p x_{ip}$$

Optional: MLR with matrices

target of sample i	features of sample i	parameters
y_i	$\mathbf{x_i} = \begin{bmatrix} 1 \\ x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{bmatrix}$	$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$

sample 1 $\quad y_1 = w_0 + w_1 x_{11} + w_2 x_{12} + \ldots + w_p x_{1p} \quad \leftrightarrow \quad y_1 = \mathbf{x_1}^T \mathbf{w}$

$$\vdots \qquad\qquad \vdots$$

sample i $\quad y_i = w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_p x_{ip} \quad \leftrightarrow \quad y_i = \mathbf{x_i}^T \mathbf{w}$

$$\vdots \qquad\qquad \vdots$$

sample n $\quad y_n = w_0 + w_1 x_{n1} + w_2 x_{n2} + \ldots + w_p x_{np} \quad \leftrightarrow \quad y_n = \mathbf{x_n}^T \mathbf{w}$

Brain Code Camp

Optional: MLR with matrices

target of sample i	features of sample i	parameters
y_i	$\mathbf{x_i} = \begin{bmatrix} 1 \\ x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{bmatrix}$	$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$

$$\begin{array}{lllccc}
 & & & \mathbf{y} & & \mathbf{X} \\
\text{sample 1} & y_1 = w_0 + w_1 x_{11} + w_2 x_{12} + \ldots + w_p x_{1p} & \leftrightarrow & y_1 & = & \mathbf{x_1}^T \mathbf{w} \\
 & \vdots & & & & \vdots \\
\text{sample i} & y_i = w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_p x_{ip} & \leftrightarrow & y_i & = & \mathbf{x_i}^T \mathbf{w} \\
 & \vdots & & & & \vdots \\
\text{sample n} & y_n = w_0 + w_1 x_{n1} + w_2 x_{n2} + \ldots + w_p x_{np} & \leftrightarrow & y_n & = & \mathbf{x_n}^T \mathbf{w}
\end{array}$$

Optional: MLR with matrices

targets of all samples

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

features of sample i

$$\mathbf{x_i} = \begin{bmatrix} 1 \\ x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{bmatrix}$$

features of all samples

$$\mathbf{X} = \begin{bmatrix} — & \mathbf{x_1}^T & — \\ — & \mathbf{x_2}^T & — \\ & \vdots & \\ — & \mathbf{x_n}^T & — \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

features of sample n

parameters

$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$$

$$y_i = w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_p x_{ip} \quad \text{for } i = 1,2,\ldots,n$$

$$\updownarrow$$

$$\mathbf{y} = \mathbf{X}\mathbf{w}$$

Brain Code Camp

Optional: MLR with matrices

$$\min_{\hat{w}_0,\dots,\hat{w}_p} MSE(Y, \hat{Y}) = \min_{\hat{w}_0,\dots,\hat{w}_p} \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (\hat{w}_0 + \hat{w}_1 x_{i1} + \ldots + \hat{w}_p x_{ip})\right)^2$$

$$\Updownarrow$$

$$\min_{\hat{\mathbf{w}}} \frac{1}{n}\|\mathbf{y} - \mathbf{X}\hat{\mathbf{w}}\|_2^2$$

$$\Updownarrow$$

$$\min_{\hat{\mathbf{w}}} \|\mathbf{y} - \mathbf{X}\hat{\mathbf{w}}\|_2^2$$

Brain Code Camp

Optional: MLR with matrices

Compute the gradient of $||\mathbf{y} - \mathbf{X}\hat{\mathbf{w}}||_2^2$ with respect to $\hat{\mathbf{w}}$ and set it to 0

$$\nabla_{\hat{\mathbf{w}}} ||\mathbf{y} - \mathbf{X}\hat{\mathbf{w}}||_2^2 = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\hat{\mathbf{w}}) = 0$$

$$-\mathbf{X}^T\mathbf{y} + \mathbf{X}^T\mathbf{X}\hat{\mathbf{w}} = 0$$

$$\mathbf{X}^T\mathbf{X}\hat{\mathbf{w}} = \mathbf{X}^T\mathbf{y}$$

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$