Logistic Regression

Tuesday, October 25, 2022 7:47 AM

$$p(y=k|x) = \frac{\exp(w_k^T x)}{\sum_{j=1}^{K} \exp(w_j^T x)}$$

$$p(y=1|x) = \frac{\exp(w_1^T x)}{\sum_{j=1}^{3} \exp(w_j^T x)}$$

$$= \frac{\exp(w_1^T x)}{\exp(w_1^T x) + \exp(w_2^T x) + \exp(w_3^T x)}$$

$$p(y=2|x) = \frac{\exp(w_2^T x)}{\sum_{j=1}^{3} \exp(w_j^T x)}$$

$$p(y=3|x) = \frac{\exp(w_3^T x)}{\sum_{j=1}^{3} \exp(w_j^T x)}$$

$\Rightarrow$ max $\Rightarrow$ class

$$p(y_n|x_n) = \prod_{k=1}^{K} p(y_n = k | x_n)^{y_{nk}}$$

$$p(y_1, \dots, y_N | x_1, \dots, x_N) = \prod_{n=1}^{N} p(y_n|x_n) = \prod_{n=1}^{N}\prod_{k=1}^{K} p(y_n = k|x_n)^{y_{nk}}$$

$$\mathcal{L}(w) = \log \prod_{n=1}^{N} p(y_n|x_n)$$

$$= \sum_{n=1}^{N} \log p(y_n|x_n)$$

$$= \sum_{n=1}^{N} \log \prod_{k=1}^{K} p(y_n = k|x_n)^{y_{nk}}$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} \log p(y_n = k|x_n)^{y_{nk}}$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} y_{nk} \log \frac{\exp(w_k^T x_n)}{\sum_{j=1}^{K}\exp(w_j^T x)} \quad (\max)$$

$$E: \quad \log\left[\exp(w_1^T x_n) + \exp(w_2^T x_n) + \exp(w_3^T x_n)\right]$$

$$\mathcal{L}(w) = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{3} y_{nk} \log \frac{\exp(w_k^T x_n)}{\sum_{j=1}^{3}\exp(w_j^T x_n)} \Leftarrow E \quad (\min)$$

$$= -\frac{1}{N}\sum_{n=1}^{N}\left[y_{n1} \log \frac{\exp(w_1^T x_n)}{E} + y_{n2} \log \frac{\exp(w_2^T x_n)}{E} + y_{n3} \log \frac{\exp(w_3^T x_n)}{E} \right]$$

$$\dots \left(w^T x - \log E\right) + y_{n3}\left(w_3^T x_n - \log E\right) \Big]$$

$$\mathcal{L}(w) = -\frac{1}{N}\sum_{n=1}^{N}\Big[y_{n1}(w_1^T x_n - \log E) + y_{n2}(w_2^T x_n - \log E) + y_{n3}(w_3^T x_n - \log E) \Big]$$

$$\frac{\partial \mathcal{L}(w)}{\partial w_1} = -\frac{1}{N}\sum_{n=1}^{N}\Big[\Big(y_{n1} x_n^T - \frac{y_{n1}\exp(w_1^T x_n)\, x_n^T}{E}\Big) - \frac{y_{n2}\exp(w_1^T x_n)\, x_n^T}{E} - \frac{y_{n3}\exp(w_1^T x_n)\, x_n^T}{E}\Big]$$

$$= -\frac{1}{N}\sum_{n=1}^{N}\Big[\Big(y_{n1} - \frac{y_{n1}\exp(w_1^T x_n)}{E}\Big) - \frac{y_{n2}\exp(w_1^T x_n)}{E} - \frac{y_{n3}\exp(w_1^T x_n)}{E}\Big] x_n^T$$

$$\frac{\partial \mathcal{L}(w)}{\partial w_2} = -\frac{1}{N}\sum_{n=1}^{N}\Big[-\frac{y_{n1}\exp(w_2^T x_n)\, x_n^T}{E} + y_{n2}\, x_n^T - \frac{y_{n2}\exp(w_2^T x_n)\, x_n^T}{E} - \frac{y_{n3}\exp(w_2^T x_n)\, x_n^T}{E}\Big]$$

$$\frac{\partial \mathcal{L}(w)}{\partial w_3} = -\frac{1}{N}\sum_{n=1}^{N}\Big[y_{n3}\, x_n^T - \frac{y_{n1}\exp(w_3^T x_n)\, x_n^T}{E} - \frac{y_{n2}\exp(w_3^T x_n)\, x_n^T}{E} - \frac{y_{n3}\exp(w_3^T x_n)\, x_n^T}{E}\Big]$$

$$w_1 = w_1 - \eta \cdot \frac{\partial \mathcal{L}(w)}{\partial w_1}, \quad w_2 = w_2 - \eta \cdot \frac{\partial \mathcal{L}(w)}{\partial w_2}, \quad w_3 = w_3 - \eta \cdot \frac{\partial \mathcal{L}(w)}{\partial w_3}$$

$$y_n = [\,0,\ 1,\ 2,\ 2,\ 1,\ 0\,]$$

Encode / decode

$$= \left[\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\0\end{pmatrix} \right]$$

$$W = \begin{bmatrix} w_{10} & w_{20} & w_{30} \\ w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \end{bmatrix}$$

$$W^T = \begin{bmatrix} w_{10} & w_{11} & w_{12} \\ w_{20} & w_{21} & w_{22} \\ w_{30} & w_{31} & w_{32} \end{bmatrix}$$

x (3×150)

$$\begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & n \\ x_1 & \cdots & & & & n \\ x_2 & \cdots & & & & \end{bmatrix}$$

feature 1

feature 2

$$x^T \quad (n\times 3) \quad \begin{bmatrix} 1 & x_1 & x_2 \\ \vdots & \vdots & \vdots \\ n & n & n \end{bmatrix}$$

$$y_{n1} = \begin{bmatrix}1\\0\\0\end{bmatrix} \quad (3\times 1)$$

$$W^T x \qquad (1\times 3)(3\times 150) \qquad (1\times 150)$$