Logistic Regression

Wednesday, March 27, 2024 10:14 PM

Let out data be as follows: $(x^i)_{i\in\{1,\dots,N\}}$, $x^i \in \mathbb{R}^D$

$(y^i)_{i\in\{1,\dots,N\}}$ $y^i \in \{0,1\}$

$\theta \in \mathbb{R}^D$

We assume our data is modelled by $P(y^i = 1 | x^i; \theta) = \frac{1}{1+\exp(-\theta^T x^i)}$

To model our data, we find optimal θ, s.t.

$$L(\theta) = -\sum_{i=1} y_i \log \mu_i + (1-y_i)\log(1-\mu_i) \quad \text{minimal}$$

where $\mu_i = \frac{1}{1+\exp(-\theta^T x)}$.

We want to compute $\frac{\partial L}{\partial \theta}(\theta)$. Note $\frac{\partial \mu_i}{\partial \theta_j} = \frac{x_j^i \exp(-\theta^T x^i)}{(1+\exp(-\theta^T x^i))^2}$

$$\frac{\partial L}{\partial \theta_j}(\theta) = -\sum_i y_i \frac{1}{\mu_i} \cdot \frac{\partial \mu_i}{\partial \theta_j} + (1-y_i)\frac{1}{1-\mu_i}\cdot(-1)\frac{\partial \mu_i}{\partial \theta_j} =$$

$$= -\sum_i \left(\frac{y_i}{\mu_i} + \frac{y_i - 1}{1-\mu_i}\right)\frac{\partial \mu_i}{\partial \theta_j} = -\sum_i \left(y_i(1+\exp(-\theta^T x^i)) + (y_i - 1)\frac{1+\exp(-\theta^T x^i)}{\exp(-\theta^T x^i)}\right)\frac{x_j^i \exp(-\theta^T x^i)}{(1+\exp(-\theta^T x^i))^2} =$$

$$= -\sum_i \left(y_i \exp(-\theta^T x^i) + y_i - 1\right)\frac{x_j^i}{1+\exp(-\theta^T x^i)} = \sum (\mu_i - y_i)\, x_j^i$$

$$\boxed{\Rightarrow \frac{\partial L}{\partial \theta} = \sum_i (\mu_i - y_i) x^i}$$

!!