Problem 2: Gradient Descent Algorithm and Logistic Regression (40 points)

(1) In logistic regression method, please derive the derivative of the negative logarithm of the likelihood function with respect to parameter w. You need to show the detailed steps to obtain the following results.

Answer:

Problem 2: Gradient Descent Algorithm and Logistic Regression

(1) In logistic regression method, please derive the derivative of negative logarithm of the likelihood function with respect to parameter w. You need to show the detailed steps to obtain the following result:

$$\nabla_{\omega} E(\omega) = \sum_{n=1}^{N} (f(x_n) - y_n) x_n .$$

WKT,

$$P(C_1|x) = \sigma(\omega^T x + \omega_0) = f(x) \quad \text{— (1)}$$

The likelihood function is,

$$L(\omega) = \prod_{n=1}^{N} P(C_1|x_n)^{y_n} (1 - P(C_1|x_n))^{1-y_n} \quad \text{— (2)}$$

$$L(\omega) = \prod_{n=1}^{N} f(x_n)^{y_n} . (1 - f(x_n))^{1-y_n}$$

The derivative of the logistic sigmoid function:

$$\frac{\partial}{\partial a}\sigma(a) = \frac{\partial}{\partial a}\left(\frac{1}{1+e^{-a}}\right) = \frac{e^{-a}}{(1+e^{-a})^2} = \frac{1}{1+e^{-a}} \cdot \frac{e^{-a}}{1+e^{-a}}$$

$$\Rightarrow \frac{1}{1+e^{-a}}\left(1 - \frac{1}{1+e^{-a}}\right) = \sigma(a).(1-\sigma(a)) \quad \text{— (3)}$$

- Taking the negative logarithm of the likelihood function (cross entropy): Equation (2) becomes,

$$E(w) = -\ln L(w) \quad \text{--- (4)}$$

$$= -\ln \prod_{n=1}^{N} p(C_1|x_n)^{y_n} \cdot (1-p(C_1|x_n))^{1-y_n}$$

$$= -\sum_{n=1}^{N} y_n(\ln(f(x_n))) + (1-y_n)(\ln(1-f(x_n)))$$

$$E(w) = -\sum_{n=1}^{N} y_n(\ln(\sigma(w^T x + w_0))) + (1-y_n)(\ln(1-\sigma(w^T x + w_0)))$$

Now, taking the derivative w.r.t w,

$$\frac{\partial E(w)}{\partial w} = \frac{\partial}{\partial w}\Big(\sum_{n=1} y_n(\ln(\sigma(w^T x + w_0))) + (1-y_n)(\ln(1-\sigma(w^T x + w_0)))\Big)$$

Using chain rule,

$$\text{RHS} = \sum_{i=1}^{N}\left[\frac{y_n}{\sigma(w^T x + w_0)} + \frac{1-y_n}{1-\sigma(w^T x+w_0)}\right]\frac{\partial \sigma(w^T x + w_0)}{\partial w}$$

$$= \sum_{i=1}^{N}\left[\frac{y_n}{\sigma(w^T x + w_0)} + \frac{1-y_n}{1-\sigma(w^T x+w_0)}\right]\sigma(w^T x + w_0).(1-\sigma(w^T x+w_0))x$$

$$= \sum_{i=1}^{N}\left[\frac{y_n - \sigma(w^T x + w_0)}{\sigma(w^T x + w_0)(1-\sigma(w^T x+w_0))}\right]\sigma(w^T x + w_0)(1-\sigma(w^T x+w_0))x$$

$$\frac{\partial L}{\partial w} = \sum_{i=1}^{N}\left[y_n - \sigma(w^T x + w_0)\right]x_n$$

$$\frac{\partial L}{\partial \omega} = \sum_{i=1}^{N} [y_n - f(x_n)] x_n$$

Substituting in eqn (4), we get,

$$\nabla_\omega E(\omega) = -\left[\sum_{i=1}^{N} [y_n - f(x_n)]\right] x_n$$

$$\therefore \boxed{\nabla_\omega E(\omega) = \sum_{n=1}^{N} (f(x_n) - y_n) x_n}$$