

3. Regularized logistic regression

lunedì 28 ottobre 2024 17:22

La regolarizzazione per la regressione logistica mostra la funzione di costo con regolarizzazione e la regola di aggiornamento dei parametri θ durante la discesa del gradiente.

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left(y^{(i)} \log h_{\theta}(\mathbf{x}^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(\mathbf{x}^{(i)})) \right) + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2$$

$$\begin{cases} \theta_0 = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m \left(h_{\theta}(\mathbf{x}^{(i)}) - y^{(i)} \right) x_0^{(i)} \\ \theta_j = \theta_j - \alpha \left[\frac{1}{m} \sum_{i=1}^m \left(h_{\theta}(\mathbf{x}^{(i)}) - y^{(i)} \right) x_j^{(i)} + \frac{\lambda}{m} \theta_j \right] \quad j = 1, 2, \dots, n \end{cases}$$

- L'aggiornamento per θ_0 non include il termine di regolarizzazione $\frac{\lambda}{m} \theta_j$, poiché θ_0 non viene regolarizzato.
- Per gli altri parametri θ_j (con $j \geq 1$), il termine aggiuntivo $\frac{\lambda}{m} \theta_j$ penalizza valori elevati di θ_j , contribuendo alla regolarizzazione.

- **Primo termine:** È la funzione di costo standard per la regressione logistica, che misura l'errore di classificazione tra le previsioni $h_{\theta}(x^{(i)})$ e i valori effettivi $y^{(i)}$.
- **Secondo termine:** È il termine di regolarizzazione, $\frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2$, che penalizza valori elevati dei parametri θ_j (eccetto θ_0). Questo termine riduce l'overfitting facendo in modo che i valori di θ siano più piccoli.

2. Aggiornamento dei Parametri nella Discesa del Gradiente

La discesa del gradiente modifica i parametri θ fino a convergenza. Gli aggiornamenti dei parametri sono dati da:

$$\theta_0 = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m \left(h_{\theta}(x^{(i)}) - y^{(i)} \right) x_0^{(i)}$$

$$\theta_j = \theta_j - \alpha \left[\frac{1}{m} \sum_{i=1}^m \left(h_{\theta}(x^{(i)}) - y^{(i)} \right) x_j^{(i)} + \frac{\lambda}{m} \theta_j \right] \quad \text{per } j = 1, 2, \dots, n$$

- L'aggiornamento per θ_0 non include il termine di regolarizzazione $\frac{\lambda}{m} \theta_j$, poiché θ_0 non viene regolarizzato.
- Per gli altri parametri θ_j (con $j \geq 1$), il termine aggiuntivo $\frac{\lambda}{m} \theta_j$ penalizza valori elevati di θ_j , contribuendo alla regolarizzazione.

↓

La regolarizzazione riduce il valore di θ_j ad ogni

iterazione, specialmente se λ è grande. Questo porta a un modello meno complesso e meno incline all'overfitting, ma è importante scegliere λ con cura: un valore troppo grande potrebbe causare underfitting, mentre un valore troppo piccolo potrebbe non avere effetto significativo.