Regresja Logistyczna

Dominik Jur

25 marca 2025

Konwencje notacyjne

W prezentacji stosujemy następujące oznaczenia:

- Wektory cech oznaczane są jako $x^{(i)}$ (dla i-tej próbki) lub x (dla ogólnego przypadku)
- Etykiety klas: $y^{(i)} \in \{0, 1\}$ (dla próbek) lub y (ogólnie)
- Przewidywania modelu: $\hat{y}(x^{(i)})$ lub $\hat{y}^{(i)}$
- Parametry modelu: $\theta = [\theta_0, \theta_1, ..., \theta_n]^T$
- θ_0 to wyraz wolny (bias)
- Funkcja sigmoidalna: $\sigma(z) = \frac{1}{1+e^{-z}}$
- n - liczba próbek w zbiorze danych
- m - liczba cech (wymiarowość wektorów x)

Dodatkowe konwencje:

- Indeksy górne w nawiasach (i) odnoszą się do numeru próbki
- Indeksy dolne (np. x_j) odnoszą się do konkretnej cechy
- Operacje na wektorach domyślnie traktujemy jako operacje kolumnowe

Metoda Gradientu Prostego

Algorithm 1 Algorytm gradientu prostego

Inicjalizujemy wektor parametrów θ, stałą uczenia η i tolerancję ε.
while true **do**
    Oblicz gradient funkcji kosztu $\nabla J(\theta)$.
    if $||\nabla J(\theta)|| < \varepsilon$ **then**
        przerwij pętlę
    end if
    Zaktualizuj parametry: $\theta \leftarrow \theta - \eta \nabla J(\theta)$
end while

Gdzie $||\nabla J(\theta)||$ to norma gradientu $J(\theta)$.

Regresja Logistyczna

Funkcja hipotezy liniowej dla wektora cech x dana jest wzorem:

$$h_\theta(x) = \theta_0 + \sum_{i=1}^{n} \theta_i x_i.$$

gdzie θ to wektor parametrów, a θ_0 jest wyrazem wolnym. Wynik regresji logistycznej definiuje się jako:

$$\text{RegresjaLogistyczna}(x) = \sigma\left(h_\theta(x)\right) = \frac{1}{1 + e^{-h_\theta(x)}},$$

gdzie x oznacza wektor cech wejściowych.

Kwestia metryk do ucznia modelu

Regresja logistyczna przewiduje prawdopodobieństwa ($\hat{y} \in [0, 1]$), podczas gdy MSE jest typowo używany w zadaniach regresji liniowej (gdzie $y \in \mathbb{R}$).
Główne powody wyboru Log Loss:
MSE zakłada liniową zależność błędu od parametrów, co prowadzi do **niewypukłej funkcji kosztu** dla regresji logistycznej:

$$J_{\text{MSE}}(\theta) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}(x^{(i)}))^2.$$

Powoduje to wiele minimów lokalnych, utrudniając optymalizację.
Log Loss jest wypukła i nie ma problemów tej natury:

$$J_{\text{Log}}(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right].$$

Błąd logistyczny dla pojedyńczego wektora cech

Funkcja błędu logistycznego dla pojedyńczego wektora cech $x^{(i)}$:

$$J(\theta) = -y^{(i)} \log\left(\hat{y}(x^{(i)})\right) - \left(1 - y^{(i)}\right) \log\left(1 - \hat{y}(x^{(i)})\right).$$

Jako, że $\forall_i \ y^{(i)} \in \{0, 1\}$, możemy zapisać błąd logistyczny jako funkcję klamrową:

$$J(\theta) = \begin{cases} -\log\left(\hat{y}(x^{(i)})\right), & \text{dla} \ \ y^{(i)} = 1, \\ -\log\left(1 - \hat{y}(x^{(i)})\right), & \text{dla} \ \ y^{(i)} = 0. \end{cases}$$

Pochodna funkcji kosztu po parametrze θ_j

Dla funkcji hipotezy w regresji logistycznej:

$$\hat{y}^{(i)} = \sigma(z^{(i)}) = \frac{1}{1 + \exp(-z^{(i)})}, \quad \text{gdzie } z^{(i)} = h_\theta(x^{(i)})$$

Pochodna po parametrze θ_j jest zadana wzorem:

$$\begin{aligned}
\frac{\partial}{\partial \theta_j} \hat{y}^{(i)} &= \frac{\partial}{\partial \theta_j} \frac{1}{1 + \exp(-z^{(i)})} = \frac{\partial}{\partial z^{(i)}} \frac{1}{1 + \exp(-z^{(i)})} \cdot \frac{\partial}{\partial \theta_j} z^{(i)} \\
&= \frac{1}{[1 + \exp(-z^{(i)})]^2} \exp(-z^{(i)}) \cdot \frac{\partial}{\partial \theta_j} z^{(i)} \\
&= \frac{1}{1 + \exp(-z^{(i)})} \left[1 - \frac{1}{1 + \exp(-z^{(i)})}\right] \frac{\partial}{\partial \theta_j} z^{(i)} \\
&= \hat{y}^{(i)} \left[1 - \hat{y}^{(i)}\right] \frac{\partial}{\partial \theta_j} z^{(i)}.
\end{aligned}$$

Pochodna funkcji kosztu po parametrze θ_j

Pochodna funkcji kosztu po parametrze θ_j dla pojedyńczego wektora cech:

$$\begin{aligned}
\frac{\partial}{\partial \theta_j} J(\theta) &= -\left[y^{(i)} \frac{\partial}{\partial \theta_j} \log \hat{y}^{(i)} + (1 - y^{(i)}) \frac{\partial}{\partial \theta_j} \log(1 - \hat{y}^{(i)}) \right] \\
&= -\left[\frac{y^{(i)}}{\hat{y}^{(i)}} - \frac{1 - y^{(i)}}{1 - \hat{y}^{(i)}} \right] \frac{\partial \hat{y}^{(i)}}{\partial \theta_j} \\
&= -\left[\frac{y^{(i)}(1 - \hat{y}^{(i)}) - (1 - y^{(i)})\hat{y}^{(i)}}{\hat{y}^{(i)}(1 - \hat{y}^{(i)})} \right] \hat{y}^{(i)}(1 - \hat{y}^{(i)}) \frac{\partial z^{(i)}}{\partial \theta_j} \\
&= -\left[y^{(i)} - y^{(i)}\hat{y}^{(i)} - \hat{y}^{(i)} + y^{(i)}\hat{y}^{(i)} \right] x_j^{(i)} \\
&= -\left[y^{(i)} - \hat{y}^{(i)} \right] x_j^{(i)}.
\end{aligned}$$

Ostatecznie otrzymujemy wzór:

$$\frac{\partial}{\partial \theta_j} J(\theta) = (\hat{y}^{(i)} - y^{(i)}) x_j^{(i)}$$

Gradient funkcji kosztu

Zatem całkowita wartość pochodnej $J(\theta)$ po θ_j jest postaci:

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}^{(i)} - y^{(i)}) x_j^{(i)}$$

Gradient jest postaci:

$$\nabla J(\theta) = \begin{bmatrix} \frac{\partial}{\partial \theta_0} \\ \frac{\partial}{\partial \theta_1} \\ \vdots \\ \frac{\partial}{\partial \theta_m} \end{bmatrix}.$$

Koszystając z tak policzonego gradientu możemy nauczyć model regresji liniowej stosując metodę gradientu prostego.