

1 Odvodenie účelovej funkcie a jej gradientu

V tejto časti sa budeme venovať odvodzovaniu účelovej funkcie a jej gradientu, ktorú v neskorších častiach budeme minimalizovať, pomocou čoho vytvoríme model na binárnu klasifikáciu.

Do logistickej funkcie $g(z) = \frac{1}{1+e^{-z}}$, ktorá bude odhadovať pravdepodobnosť solventnosti klienta, budeme dosádzať hodnoty $z = x^T u^i$ pre vektor parametrov $x = (x_0, \dots, x_3)$ a vektor údajov o klientovi $u^i = (1, u_1^i, u_2^i, u_3^i)$, pre $i = 1, \dots, m$.

Chceme odhadnúť zložky vektora x tak, aby čo najvierohodnejšie predpovedal solventnosť vzhľadom na naše dáta. To vedie k optimalizačnej úlohe:

$$\min_{x \in \mathbb{R}^4} J(x)$$

kde

$$J(x) = - \sum_{i=1}^m v^i \ln(g(x^T u^i)) + (1 - v^i) \ln(1 - g(x^T u^i))$$

Z predpisu funkcie si môžeme všimnúť, že suma nadobúda záporné hodnoty, čiže $J(x)$ nadobúda kladné hodnoty. Taktiež si môžeme všimnúť, že ak je klient i solventný, čiže $v^i = 1$ a pre nejaký vektor parametrov x je hodnota $g(x^T u^i)$ blízka nule, má to za následok „výrazné“ zvyšovanie hodnoty účelovej funkcie. Podobnou logikou vidíme zvyšovanie hodnoty účelovej funkcie pre nesolventných klientov, ak pomocou vektora x mu prisúdime veľkú pravdepodobnosť solventnosti hodnotou $g(x^T u^i)$. Chceme teda nájsť taký vektor x , že $g(x^T u^i)$ bude blízke 1 pre solventného klienta a blízke 0 pre nesolventného.

1.1 Kompaktnejší tvar účelovej funkcie

Pre lepšiu manipuláciu a neskoršiu implementáciu si zjednodušíme tvar účelovej funkcie nasledovne:

$$\begin{aligned} J(x) &= - \sum_{i=1}^m v^i \ln(g(x^T u^i)) + (1 - v^i) \ln(1 - g(x^T u^i)) \\ &= - \sum_{i=1}^m v^i \ln\left(\left(1 + e^{-x^T u^i}\right)^{-1}\right) + (1 - v^i) \ln\left(\frac{e^{-x^T u^i}}{1 + e^{-x^T u^i}}\right) \\ &= - \sum_{i=1}^m -v^i \ln(1 + e^{-x^T u^i}) + (1 - v^i) \left(\ln(e^{-x^T u^i}) - \ln(1 + e^{-x^T u^i})\right) \\ &= - \sum_{i=1}^m -v^i \ln(1 + e^{-x^T u^i}) - (1 - v^i)x^T u^i - (1 - v^i) \ln(1 + e^{-x^T u^i}) \\ &= \sum_{i=1}^m (1 - v^i)x^T u^i + \ln(1 + e^{-x^T u^i}) \end{aligned}$$

S takýmto vyjadrením funkcie $J(x)$ budeme pracovať v nasledujúcich častiach.

1.2 Gradient účelovej funkcie

Vyjadríme si najprv parciálnu deriváciu podľa x_0 a potom podľa $x_j, j = 1, 2, 3$, keďže tie sa správajú symetricky.

$$\begin{aligned}
 \frac{\partial}{\partial x_0} J(x) &= \frac{\partial}{\partial x_0} \sum_{i=1}^m (1 - v^i) x^T u^i + \ln(1 + e^{-x^T u^i}) \\
 &= \sum_{i=1}^m \frac{\partial}{\partial x_0} \left((1 - v^i)(x_0 + x_1 u_1^i + x_2 u_2^i + x_3 u_3^i) + \ln(1 + e^{-x^T u^i}) \right) \\
 &= \sum_{i=1}^m (1 - v^i) - \frac{e^{-x^T u^i}}{1 + e^{-x^T u^i}} \\
 &= \sum_{i=1}^m 1 - v^i - \frac{1}{1 + e^{x^T u^i}}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial}{\partial x_j} J(x) &= \frac{\partial}{\partial x_j} \sum_{i=1}^m (1 - v^i) x^T u^i + \ln(1 + e^{-x^T u^i}) \\
 &= \sum_{i=1}^m \frac{\partial}{\partial x_j} \left((1 - v^i)(x_0 + x_1 u_1^i + x_2 u_2^i + x_3 u_3^i) + \ln(1 + e^{-x^T u^i}) \right) \\
 &= \sum_{i=1}^m (1 - v^i) u_j^i - u_j^i \frac{e^{-x^T u^i}}{1 + e^{-x^T u^i}} \\
 &= \sum_{i=1}^m \left(1 - v^i - \frac{1}{1 + e^{x^T u^i}} \right) u_j^i \quad j = 1, 2, 3
 \end{aligned}$$

Toto vieme kompaktno zapísať nasledovne:

$$\nabla J(x) = \sum_{i=1}^m \begin{pmatrix} 1 \\ u_1^i \\ u_2^i \\ u_3^i \end{pmatrix} \left(1 - v^i - \frac{1}{1 + e^{x^T u^i}} \right)$$