Exer 9 ML

Higor Gabriel de Freitas

October 2025

1 Hessiana*

A dedução foi em sua maior parte feita pelo ChatGPT, apenas acompanhei e fiz a transferência para o LATEX

Derivação da Hessiana da Regressão Logística Multinomial

A partir do gradiente da função de custo média

$$\nabla_W L(W) = \frac{1}{m} X^\top (P - Y),$$

onde $P = \text{softmax}(XW)$, queremos obter a matriz Hessiana.

Primeiro, lembramos que cada componente do gradiente para a classe r é

$$g_r = \frac{\partial L}{\partial w_r} = \frac{1}{m} \sum_{i=1}^{m} (p_{ir} - y_{ir}) x_i,$$

onde $p_{ir} = P(y = r \mid x_i)$.

Para obter a segunda derivada, derivamos g_r em relação a w_s:

$$\frac{\partial g_r}{\partial w_s} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial p_{ir}}{\partial w_s} x_i^\top.$$

Sabemos que a derivada do termo de probabilidade em relação às ativações é

$$\frac{\partial p_{ir}}{\partial z_{is}} = p_{ir}(\delta_{rs} - p_{is}),$$

onde δ_{rs} é o delta de Kronecker. Como $z_{is} = x_i^\top w_s$, temos

$$\frac{\partial p_{ir}}{\partial w_s} = \frac{\partial p_{ir}}{\partial z_{is}} \frac{\partial z_{is}}{\partial w_s} = p_{ir}(\delta_{rs} - p_{is}) x_i.$$

Substituindo de volta na expressão da Hessiana, obtemos o bloco (r, s):

$$H_{rs} = \frac{\partial^2 L}{\partial w_s \, \partial w_r^\top} = \frac{1}{m} \sum_{i=1}^{m} p_{ir}(\delta_{rs} - p_{is}) x_i x_i^\top.$$

Em notação matricial, para cada amostra i, definimos

$$S_i = \mathrm{diag}(p_i) - p_i p_i^\top,$$

onde $p_i = (p_{i1}, \ldots, p_{iK})^\top$. Assim, a Hessiana completa pode ser expressa de forma compacta como

$$H = \frac{1}{m} \sum_{i=1}^{m} \left(S_i \otimes (x_i x_i^\top) \right),$$

onde $\otimes$ é o produto de Kronecker.

Essa expressão mostra que a Hessiana é uma matriz bloco $dK \times dK$, em que cada bloco H_{rs} mede a curvatura entre as direções de w_r e w_s, ponderada pelas probabilidades p_{ir} e p_{is} do modelo softmax.