① $\forall x \in R^n$, $y \in \{1,2,3,4,\cdots,k\}$。数据集是一个由 m 个 x/y 组成的

② 任务：设计一种 softmax 机制来实现标签 y 的预测任务

即寻找 $H: x \to y$,

③ 引入线性变化和 Softmax：

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & & \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \times \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1k} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2k} \\ \vdots & & & \\ \theta_{n1} & \theta_{n2} & \cdots & \theta_{nk} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1k} \\ h_{21} & h_{22} & \cdots & h_{2k} \\ \vdots & & & \\ h_{m1} & h_{m2} & \cdots & h_{mk} \end{bmatrix}$$

$$X_{mn} \times \theta_{nk} \longrightarrow H_{mk}$$

再引入 Softmax $(H) \longrightarrow$

其中 $z_{ij} = \dfrac{e^{h_{ij}}}{\sum_{a=1}^{k} e^{h_{ia}}}$

$$\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1k} \\ z_{21} & z_{22} & \cdots & z_{2k} \\ \vdots & & & \\ z_{m1} & z_{m2} & \cdots & z_{mk} \end{bmatrix}$$

④ z_{ij} 可以看作 $P(\text{label}=j \mid x_i)$，一种预测类别为 j 的 p

定义损失函数 $\text{loss}(H; x_i, y^i) = -\log(p(\text{label}=j \mid x_i))$

loss 小

那么整体 $Loss = \frac{1}{m}\sum_{i=1}^{m} -\log p(label=y^i|x_i,y_i)$

⑤ 现在开始研究 Loss 和 θ 的关系

$$Loss = \frac{1}{m}\sum_{i=1}^{m}\left(-\log\frac{e^{h_{iy_i}}}{\sum_{a=1}^{k} e^{h_{ia}}}\right) = \frac{1}{m}\sum_{i=1}^{m}\left(-h_{iy_i} + \log\sum_{a=1}^{k} e^{h_{ia}}\right)$$

Loss 与 h 有关，h 与 θ 有关，那么引入偏导

$$\frac{dLoss}{d\theta_{r1}} = \frac{1}{m}\cdot\frac{d\sum_{i=1}^{m}\left(-h_{iy_i} + \log\sum_{a=1}^{k} e^{h_{ia}}\right)}{d\ \theta_{i1}}$$

⑥ θ_{r1} 会影响哪些 h？ $h_{ab} = x_{a1}\theta_{1b} + x_{a2}\theta_{2b} + \cdots + x_{an}\theta_{nb}$

观察发现 θ_{r1} 和 $h_{11}, h_{21}, h_{31}, \cdots, h_{m1}$ 有关

$$\begin{cases} h_{11} = x_{11}\theta_{11} + x_{12}\theta_{21} + \cdots + x_{1r}\theta_{r1} + \cdots + x_{1n}\theta_{n1} \\ h_{21} = x_{21}\theta_{11} + x_{22}\theta_{21} + \cdots + x_{2r}\theta_{r1} + \cdots + x_{2n}\theta_{n2} \\ \vdots \\ h_{m1} = x_{m1}\theta_{11} + x_{m2}\theta_{21} + \cdots + x_{mr}\theta_{r1} + \cdots + x_{mn}\theta_{n1} \end{cases}$$

⑦ 记 $f_1 = -h_{iy_i}$，$f_2 = \log\sum_{a=1}^{k} e^{h_{ia}}$

$$\frac{\partial Loss}{\partial \theta_{r1}} = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{\partial f_1}{\partial \theta_{r1}} + \frac{\partial f_2}{\partial \theta_{r1}}\right)$$

$$\frac{\partial f_1}{\partial \theta_{r1}} = \begin{cases} -x_{ir} & y^i = 1 \\ 0 & y^i \neq 1 \end{cases} \quad \left(h_{i1} = x_{i1}\theta_{11} + x_{i2}\theta_{21} + \cdots + x_{ir}\theta_{r1} + \cdots + x_{ik}\theta_{k1}\right)$$

$$\frac{\partial f_2}{\partial \theta_{r1}} = \frac{\partial f_2}{\partial h_{i1}} \cdot \frac{\partial h_{i1}}{\partial \theta_{r1}} = \frac{1}{e^{h_{i1}} + e^{h_{i2}} + \cdots + e^{h_{ik}}} \cdot e^{h_{i1}} \cdot x_{ir}$$

$$= \frac{e^{h_{i1}}}{\sum_{\alpha=1}^{k} e^{h_{i\alpha}}} \cdot x_{ir}$$

$$= z_{i1} \cdot x_{ir}$$

所以 $$\frac{\partial Loss}{\partial \theta_{r1}} = \begin{cases} \frac{1}{m}\sum_{i=1}^{m}\left(-x_{ir} + z_{i1} \cdot x_{ir}\right) & y^i = 1 \\ \frac{1}{m}\sum_{i=1}^{m}\left(0 + z_{i1} \cdot x_{ir}\right) & y^i \neq 1 \end{cases}$$

④ 把 θ_{r1} 推广到 θ_{rq}, $q \in \{1, 2, \cdots, k\}$

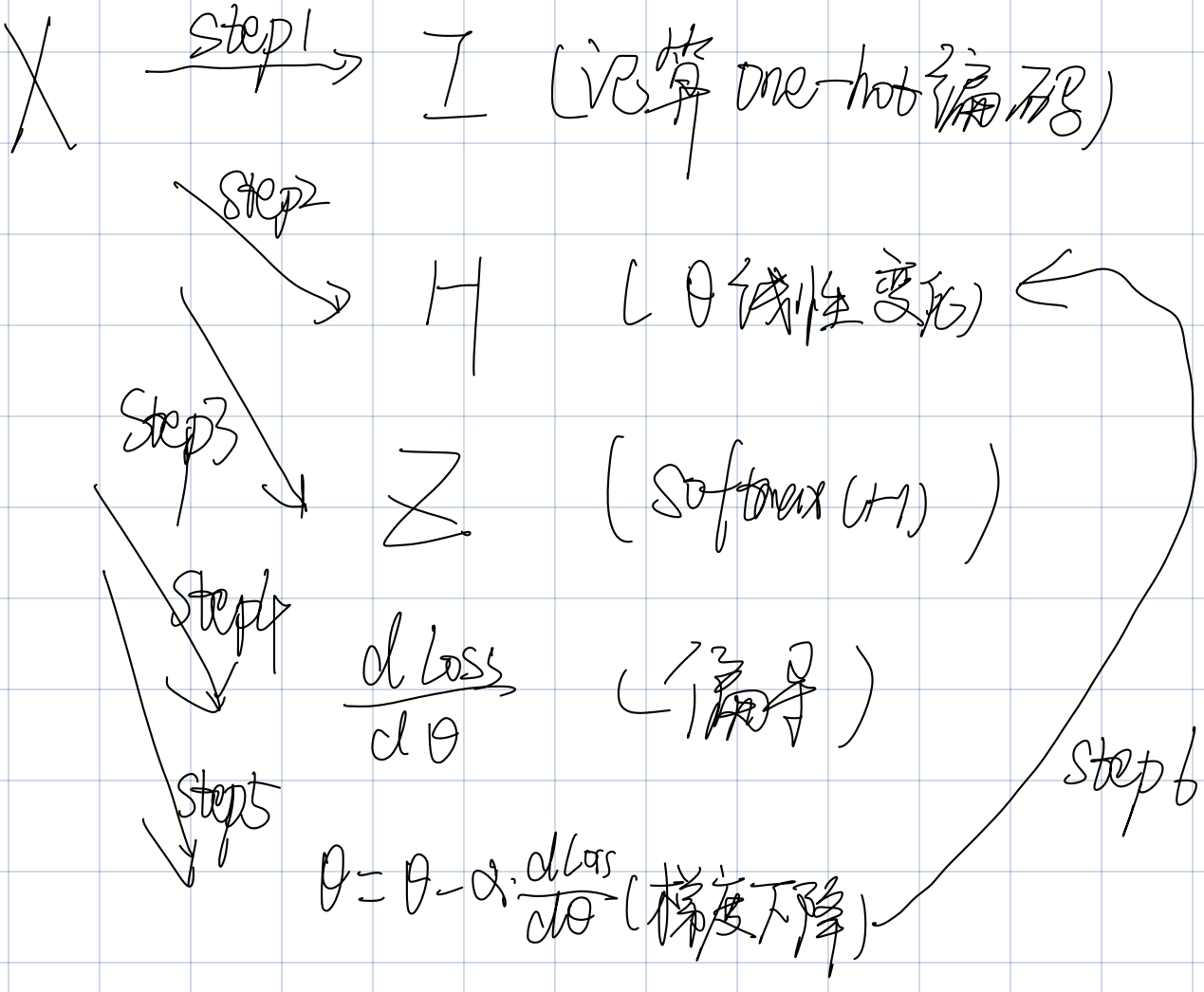
$$\frac{\partial Loss}{\partial \theta_{rq}} = \frac{1}{m}\sum_{i=1}^{m}\left(x_{ir}\left(z_{iq} - I_{iq}\right)\right)$$

其中 $$I_{iq} = \begin{cases} 0 & y^i \neq q \\ 1 & y^i = q \end{cases}$$

⑨ 我们再把 $\frac{dLoss}{d\theta_{rq}}$ 整理成矩阵形式

$$\underset{n\times k}{\frac{dLoss}{d\theta}} = \frac{1}{m}\underset{n\times m}{X^T}\underset{m\times k}{(Z-I)}$$

⑩ 现在有3偏导, everything is okay!



六步求解θ参数!