FM:

$$y = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \langle v_i, v_j \rangle x_i x_j$$

$$\langle v_i, v_j \rangle = \sum_{j=1}^{K} v_{i,f} \cdot v_{j,f}$$

损失函数: $J(w_0 \ldots w_n, V) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log \hat{y}^{(i)} + (1-y^{(i)}) \log(1-\hat{y}^{(i)})$

公式变形: $\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \frac{1}{2}\sum_{f=1}^{K}\left[\left(\sum_{i=1}^{n} v_{i,f} x_i\right)^2 - \sum_{i=1}^{n}(v_{i,f} x_i)^2\right]$

推导: $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \langle v_i, v_j \rangle x_i x_j$ (矩阵右上角)

$$= \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \langle v_i, v_j \rangle x_i x_j - \sum_{i=1}^{n} \langle v_i, v_i \rangle x_i x_i\right)$$

可理解为(矩阵 − 对角元素) $\times \frac{1}{2}$

$$= \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{f=1}^{K} v_{i,f} v_{j,f} x_i x_j - \sum_{i=1}^{n}\sum_{f=1}^{K} v_{i,f} v_{i,f} x_i x_i\right)$$

$$= \frac{1}{2}\left(\sum_{f=1}^{K}\sum_{i=1}^{n} v_{i,f} x_i \sum_{j=1}^{n} v_{j,f} x_j - \sum_{j=1}^{K}\sum_{i=1}^{n} v_{i,j}^2 x_i^2\right)$$

$$= \frac{1}{2}\sum_{f=1}^{K}\left[\left(\sum_{i=1}^{n} v_{i,f} x_i\right)^2 - \sum_{i=1}^{n} v_{i,f}^2 x_i^2\right]$$

$$= \boxed{\frac{1}{2}\sum_{f=1}^{K}\left[\left(\sum_{i=1}^{n} v_{i,f} x_i\right)^2 - \sum_{i=1}^{n} (v_{i,f} x_i)^2\right]}$$

放入 sigmiod 求梯度.

梯度：

(1) $\dfrac{\partial J}{\partial w_j} =$

(2) $\dfrac{\partial J}{\partial w_0} =$ 同 LR 和 POLY2

(3) $\dfrac{\partial J}{\partial V_{jk}} = -\dfrac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} - \hat{y}^{(i)}\right]\cdot\left[x_j\left(\sum_{i=1}^{n} V_{ik}\, x_i\right) - V_{jk}\, x_j^2\right]$

推导：

$$\frac{\partial J}{\partial V_{j,k}} = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \times \frac{\partial \log \hat{y}^{(i)}}{\partial V_{j,k}} + (1-y^{(i)}) \times \frac{\partial \log(1-\hat{y}^{(i)})}{\partial V_{j,k}} \right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \times \frac{\partial \log \hat{y}^{(i)}}{\partial \hat{y}^{(i)}} \times \frac{\partial \hat{y}^{(i)}}{\partial z^{(i)}} \times \frac{\partial z^{(i)}}{\partial V_{j,k}} \right.$$

$$\left. + (1-y^{(i)}) \times \frac{\partial \log(1-\hat{y}^{(i)})}{\partial (1-\hat{y}^{(i)})} \times \frac{\partial (1-\hat{y}^{(i)})}{\partial \hat{y}^{(i)}} \times \frac{\partial \hat{y}^{(i)}}{\partial z^{(i)}} \times \frac{\partial z^{(i)}}{\partial V_{j,k}} \right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \times (1-\hat{y}^{(i)}) + (1-y^{(i)})(-\hat{y}^{(i)}) \right] \times \frac{\partial z^{(i)}}{\partial V_{j,k}}$$

$$= -\frac{1}{m}\sum_{i=1}^{m}(y^{(i)} - \hat{y}^{(i)}) \boxed{\frac{\partial z^{(i)}}{\partial V_{j,k}}}$$

$$\frac{\partial\left(w_0 + \sum_{i=1}^{m} w_i x_i + \frac{1}{2}\sum_{f=1}^{k}\left[\left(\sum_{i=1}^{n} V_{if}\, x_i\right)^2 - \sum_{i=1}^{n}(V_{if}\, x_i)^2\right]\right)}{\partial V_{jk}}$$

$$= \frac{1}{2}\left(\frac{\partial \sum_{f=1}^{k}\left(\sum_{i=1}^{n} V_{if}\, x_i\right)^2}{\partial V_{jk}} - \frac{\partial \sum_{f=1}^{k}\sum_{i=1}^{n}(V_{if}\, x_i)^2}{\partial V_{jk}} \right)$$

$$\boxed{\begin{aligned} &2(\cdots + V_{jk}\, x_j + \cdots)\, x_j \\ &= 2\, x_j \sum_{i=1}^{n} V_{ik}\, x_i \end{aligned}} \qquad 2 V_{jk}\, x_j^2$$

只看和 $V_{j,k}$ 有关的项，$\sum_{i=1}^{n}$ 可去

$$= \boxed{x_j\left(\sum_{i=1}^{k} V_{ik}\, x_i\right) - V_{jk}\, x_j^2}$$

最终

$$\frac{\partial J}{\partial v_{jk}} = -\frac{1}{m}\sum_{i=1}^{m}(y^{(i)}-\hat{y}^{(i)})\cdot\left[x_j\left(\sum_{i=1}^{n}v_{ik}x_i\right)-v_{jk}x_j^2\right]$$

$$w_0 = w_0 + \eta\frac{1}{m}\sum_{i=1}^{m}(y^{(i)}-\hat{y}^{(i)})$$

$$w_j = w_j + \eta\frac{1}{m}\sum_{i=1}^{m}(y^{(i)}-\hat{y}^{(i)})\, x_j^{(i)}$$

$$v_{jk} = v_{jk} + \eta\frac{1}{m}\sum_{i=1}^{m}(y^{(i)}-\hat{y}^{(i)})\left[x_j\left(\sum_{i=1}^{n}v_{jk}x_i\right)-v_{jk}x_j^2\right]$$