1.1.(i)

$$\frac{\partial y_i}{\partial \gamma} = \hat{x_i}$$
$$\frac{\partial y_i}{\partial \beta} = 1$$

1.1.(ii)

$$\frac{\partial y_j}{\partial x_i} = \begin{cases} 0, i \neq j \vee r_j < p \\ 1/(1-p), i = j \wedge r_j \geq p \end{cases}$$

1.1.(iii)

$$\frac{\partial g(z)_j}{\partial z_i} = \begin{cases} \frac{e^{z_i}(\Sigma_{j=1}^k e^{z_j}) - e^{2z_i}}{(\Sigma_{j=1}^k e^{z_j})^2}, i = j \\ -\frac{e^{z_i+z_j}}{(\Sigma_{j=1}^k e^{z_j})^2}, i \neq j \end{cases}$$

1.2.(i)

$$z_{FC_{1a}} = \theta_{1a}x + b_{1a}$$
$$a_{FC_{1a}} = ReLU(z_{FC_{1A}})$$
$$a_{DP_{1a}} = M \bigodot a_{FC_{1a}}$$
$$\hat{y}_a = a_{FC_{2a}} = \theta_{2a}a_{DP_{1a}} + b_{2a}$$
$$z_{FC_{1b}} = \theta_{1b}x + b_{1b}$$
$$a_{FC_{1b}} = ReLU(z_{FC_{1b}})$$
$$a_{BN_{1b}} = BN_{\gamma,\beta}(a_{FC_{1b}})$$
$$z_{FC_{2b}} = \theta_{2b}(a_{BN_{1b}} \bigoplus a_{FC_{2a}}) + b_{2b}$$
$$\hat{y}_b = a_{FC_{2b}} = Softmax(z_{FC_{2b}})$$
$$L(x, y_a, y_b; \theta) = \frac{1}{m}\Sigma_{i=1}^m[\frac{1}{2}||(\hat{y}_{ai} - y_{ai})||_2^2 - \Sigma_{j=1}^{n_{yb}} y_{bi}^j log(\hat{y}_{bi}^j)]$$

1.2.(ii)

$$\frac{\partial L}{\partial z_{FC_{2b}}} = \frac{1}{m}\Sigma_{i=1}^m(\hat{y}_b^{(i)} - y_b^{(i)}), \text{得到残差}\delta^{(FC_{2b})}$$
$$\frac{\partial L}{\partial \theta_{2b}} = \delta^{(FC_{2b})}(a_{BN_{1b}} \oplus \hat{y}_a)^T$$

$$\frac{\partial L}{\partial a_{BN_{1b}}} = \delta^{(FC_{2b})}\theta_{2b}^T, \text{得到残差}\delta^{(BN_{1b})}\frac{\partial L}{\partial \gamma} = \delta^{(BN_{1b})}\hat{a}_{FC_{1b}}^T$$
$$\frac{\partial L}{\partial \beta} = \Sigma_{i=1}^{n_{ya}}\delta_i^{(BN_{1b})}$$

对于BN层，记输出$a_{BN_{1b}}$为y，记输入$a_{FC_{1b}}$为x，则有

$$\frac{\partial y_j}{\partial x_i} = \begin{cases} \gamma(\sigma_b^2 + \epsilon)^{-3/2}((1 - \frac{1}{m})(\sigma_b^2 + \epsilon) - \frac{1}{m}(x_j - \sigma_B)(x_i - \sigma_B), i = j \\ \gamma(\sigma_b^2 + \epsilon)^{-3/2}(-\frac{1}{m}(\sigma_b^2 + \epsilon) - \frac{1}{2}(x_j - \sigma_B)(x_i - \sigma_B), i \neq j \end{cases}$$
$$\frac{\partial L}{\partial a_{FC_{1b}i}} = \Sigma_{j=1}^m \delta^{(BN_{1b})}\frac{\partial y_j}{\partial x_i}$$

$$\frac{\partial a_{FC_{1b}i}}{\partial z_{FC_{1b}i}} = sgn(z_{FC_{1b}i})$$

$$\frac{\partial L}{\partial z_{FC_{1b}i}} = \Sigma_{j=1}^{m} \delta^{(BN_{1b})} \frac{\partial y_j}{\partial x_i} \bigodot sgn(z_{FC_{1b}i}), \text{记为残差} \delta^{(FC_{1b})}$$

$$\frac{\partial L}{\partial \theta_{1b}} = \delta^{(FC_{1b})} x^T$$

$$\frac{\partial L}{\partial \hat{y}_a} = \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_a^{(i)} - y_a^{(i)}) + \delta^{(FC_{2b})} \theta_{2b}^T, \text{组合得到残差} \delta^{(ya)}$$

$$\frac{\partial L}{\partial \theta_{2a}} = \delta^{(ya)} a_{DP_{1a}}^T$$

$$\frac{\partial L}{\partial a_{DP_{1a}}} = \delta^{(ya)} \theta_{2a}^T, \text{组合得到残差} \delta^{(dp)}$$

$$\frac{\partial L}{\partial a_{FC_{1a}} i} = \begin{cases} 0, r_i < p \\ \frac{1}{1-p} \delta_i^{(dp)}, r_i \geq p \end{cases}$$

$$\frac{\partial L}{\partial z_{FC_{1a}} i} = \frac{\partial L}{\partial a_{FC_{1a}} i} * Relu'(z_{FC_{1a}i}) = \begin{cases} 0, r_i < p \\ \frac{1}{1-p} \delta_i^{(dp)} sgn(z_{FC_{1a}i}), r_i \geq p \end{cases}, \text{组合得到残差} \delta^{(FC_{1a})}$$

$$\frac{\partial L}{\partial \theta_{1a}} = \delta^{(FC_{1a})} x^T$$