$y_i \sim \text{Bernoulli}(p_i)$

$p_i = \text{sigmoid}(h_i^T \beta) = \text{sigmoid}\left(\sum_{k=1}^{d} \beta_k h_{ik}\right)$

$\cancel{h_{ik} = \text{sigmoid}(x_i^T \alpha_k) = \text{sigmoid}\left(\sum_{j=1}^{p} \alpha_{kj} x_{ij}\right)}$

$h_{ik} = \text{ReLU}(x_i^T \alpha_k) = \text{ReLU}\left(\sum_{j=1}^{p} \alpha_{kj} x_{ij}\right)$

$P = \prod_i p_i^{y_i} (1-p_i)^{1-y_i}$

$\log P = \sum_i \left(y_i \log p_i + (1-y_i)\log(1-p_i)\right)$

$= \sum_{i=1}^{n} \left(y_i (A - \log(1+\exp(A))) + (1-y_i)(\log 1 - \log(1+\exp(A))) \right)$

$= \sum_{i=1}^{n} \left(y_i A - \log(1+\exp(A)) \right), \quad A = \sum_{k=1}^{d} \beta_k h_{ik}$

$\therefore L(\beta, \alpha) = \sum_{i=1}^{n} \left(y_i \sum_{k=1}^{d} \beta_k h_{ik} - \log\left(1 + \exp\left(\sum_{k=1}^{d} \beta_k h_{ik}\right)\right)\right).$

$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} (y_i - p_i) h_i$

$\frac{\partial L}{\partial h_{ik}} = \frac{\partial}{\partial h_{ik}} \sum_{i=1}^{n} \left(y_i \sum_{k=1}^{d} \beta_k h_{ik} - \log\left(1 + \exp\left(\sum_{k=1}^{d} \beta_k h_{ik}\right)\right)\right)$

$= y_i \beta_k - \frac{\exp\left(\sum_{k=1}^{d} \beta_k h_{ik}\right)\beta_k}{1 + \exp\left(\sum_{k=1}^{d} \beta_k h_{ik}\right)} = (y_i - p_i)\beta$

$\frac{\partial h_{ik}}{\partial \alpha_k} = 1(x_i^T \alpha_k > 0)\, x_i$

$\frac{\partial L}{\partial \alpha_k} = (y_i - p_i)\, \beta_k\, 1(x_i^T \alpha_k > 0)\, x_i$