2 param model gradient derivation

$$p(u_k) = \frac{e^{(a+e^B u_k)}}{1+e^{a+e^B u_k}} = \frac{e^{(a+e^b u_k)}}{1+e^{(a+e^b u_k)}} \Rightarrow \log p(u_k) = \log \frac{e^{(a+e^b u_k)}}{1+e^{(a+e^b u_k)}}$$

$$\operatorname{argmax} \log p(D|a,b) = \operatorname{argmax} \sum_{k=1}^{K} \{ n_k \log p(u_k) + (N_k - n_k) \log(1 - p(u_k))$$

$$\frac{d}{da} \log p(D|a,b) = \sum_{k=1}^{K} \frac{n_k}{e^{a+e^b u_k}+1} - \frac{N_k e^{a+e^b u_k}}{e^{a+e^b u_k}+1} + \frac{n_k e^{a+e^b u_k}}{e^{a+e^b u_k}+1}$$

$$\frac{d}{da} p(u_k) = \frac{e^{a+e^b d}}{(e^{a+e^b d}+1)^2}$$

$$= \sum_{k=1}^{K} \frac{n_k - N_k e^{a+e^b u_k} + n_k e^{a+e^b u_k}}{e^{a+e^b u_k}+1}$$

$$= \sum_{k=1}^{K} \frac{n_k + (n_k - N_k) e^{a+e^b u_k}}{e^{a+e^b u_k}+1}$$

$$\frac{d}{da} \log p(D|a,b) = \sum_{k=1}^{K} \frac{n_k + (n_k - N_k) e^{a+e^b u_k}}{e^{a+e^b u_k}+1}$$

gradient wrt a (shared param)

$$= \boxed{\sum_{k=1}^{K} \frac{N_k}{e^{a+e^b u_k}+1} + n_k - N_k}$$

$$p(u_k) = \frac{e^{(a+e^b d)}}{1+e^{(a+e^b d)}}$$

$$\frac{d}{db} p(u_k) = \frac{d\, e^{a+e^b d + b}}{(e^{a+e^b d}+1)^2}$$

$$\operatorname{argmax} \log p(D|a,b) = \operatorname{argmax} \sum_{k=1}^{K} \{ n_k \log p(u_k) + (N_k - n_k) \log(1 - p(u_k))$$

$$\frac{d}{db} \log p(D|a,b) = \sum_{k=1}^{K} \frac{n_k}{p(u_k)} \cdot \frac{d}{db} p(u_k) - \frac{N_k p'(u_k)}{1-p(u_k)} + \frac{n_k p'(u_k)}{1-p(u_k)}$$

gradient wrt b

$$= \sum_{k=1}^{K} \frac{n_k d e^b}{e^{a+e^b d}+1} - \frac{N_k d e^{a+e^b d + b}}{e^{a+e^b d}+1} + \frac{n_k d e^{a+e^b d + b}}{e^{a+e^b d}+1}$$

$$= \sum_{k=1}^{K} n_k d e^b - \frac{N_k d e^{a+e^b d + b}}{e^{a+e^b d}+1}$$

$$= \sum_{k=1}^{K} d e^b \left(n_k - \frac{N_k e^{a+e^b d}}{e^{a+e^b d}+1} \right)$$

$$= \boxed{\sum_{k=1}^{K} u_k e^b (n_k - N_k \cdot p(u_k))}$$