Minimal energy maximal entropy

A Mean field approach

Minimize

$$E(\delta) = \frac{1}{2} \sum_{m} \sum_{n} \delta_m^T C(m, n) \delta_n$$

Subject to $\sum_{i} \delta_{mi} = 1, \delta_{mi} \in \{0,1\}$

$$\cancel{\frac{\partial E}{\partial \delta_{mi}}}$$

Minimize

$$E(\delta) = \frac{1}{2} \sum_{m} \sum_{n} \delta_m^T C(m,n) \delta_n$$

Subject to $\sum_{i} \delta_{mi} = 1, \delta_{mi} \in \{0,1\}$

Probabilistic approach: expectation maximization

- Let $p_{ij} \propto \exp(u_{ij})$
- $\sum_j p_{ij} = 1$
- Minimal energy $E(\delta)$ and maximal entropy $-\sum_i \sum_j p_{ij} \log p_{ij}$

An auxiliary variable

p_{ij} represents the probability of the jth bit δ_{ij} being active, $p_{ij} \in [0,1]$

$$\delta_i = [\delta_{i1}, \ldots, \delta_{ij}, \ldots, \delta_{in}], \delta_{ij} \in \{0,1\}$$

δ_i is a unitary vector of binary elements

$$\sum_j \delta_{ij} = 1$$

Unitary condition

A probabilistic model assumes δ_i as a random variable.

$$\delta_i \in \{e_1, \ldots e_n\}$$

e_i is a unitary vector of n binary elements with the ith element one

$$p_{ij} = \Pr(\delta_i = e_j) \propto \exp(u_{ij})$$

Find the expectation of δ

Find the entropy of δ

$$e_1 = [1,0,\ldots,0]^T$$

$$\vdots$$

$$e_i = [0,0,\ldots,0,1,0\ldots,0]^T$$

$$\vdots$$

$$e_n = [0,0,\ldots,0,1]^T$$

Let $p_{ij} = K \exp(u_{ij})$

Since $\sum_j p_{ij} = 1$,

$$\sum_j p_{ij} = \sum_j K \exp(u_{ij}) = 1$$

Softmax Operation

$$K = \frac{1}{\sum_j \exp(u_{ij})}, \quad p_{ij} = \frac{exp(u_{ij})}{\sum_h \exp(u_{ih})}$$

p_{ij} represents the probability of the jth bit δ_{ij} being active, $p_{ij} \in [0,1]$

Expectation of δ_i

$= [p_{i1}, \ldots, p_{ij}, \ldots, p_{in}]^T$

$$= [\frac{exp(u_{i1})}{\sum_h \exp(u_{ih})}, \ldots, \frac{exp(u_{ij})}{\sum_h \exp(u_{ih})}, \ldots, \frac{exp(u_{in})}{\sum_h \exp(u_{ih})}]^T$$

Entropy of δ

$$p_{ij} = \frac{exp(u_{ij})}{\sum_h \exp(u_{ih})},$$

$$= -\sum_i \sum_j p_{ij} \log p_{ij}$$

$$= -\sum_{ij} (p_{ij} u_{ij} - p_{ij} \log \sum_h exp(u_{ih})) = -\sum_{ij} p_{ij} u_{ij} + \log \sum_h \exp(u_{ih})$$

$$p_{ij} = \frac{exp(u_{ij})}{\sum_h \exp(u_{ih})},$$

Mean Field Approximation: Minimal $\boxed{\text{mean}}$ energy and maximal entropy

Directly substituting mean to energy

Free energy
= mean energy - entropy

T = 1

Mean Energy is approximated by the energy of mean,

$$E(<\delta>) = \sum_{ij} p_i^T C_{ij} p_j \quad approximates \qquad <E(\delta)>$$

$p_i = \langle \delta_i \rangle$

intractable in computation $<E(\delta)> = \sum_{\delta} E(\delta) Pr(\delta)$

Free energy

$$= \sum_{ij} p_i^T C_{ij} p_j - (-\sum_i \sum_j p_{ij} \log p_{ij})$$

$$= \sum_i \sum_j p_i^T C_{ij} p_j + \sum_{ij} (p_{ij} u_{ij} - p_{ij} \log \sum_h exp(u_{ih}))$$

$$= \sum_i \sum_j p_i^T C_{ij} p_j + \sum_{ij} p_{ij} u_{ij} - \sum_i \log \sum_h \exp(u_{ih})$$

Differentiable

$$L(p,u) = \sum_i \sum_j p_i^T C_{ij} p_j + \sum_{ij} p_{ij} u_{ij} - \sum_i \log \sum_h \exp(u_{ih})$$

$$\frac{\partial L(p,u)}{\partial p_i} = 0, \frac{\partial L(p,u)}{\partial u_{ij}} = 0$$

$$L(p,u) = \sum_i \sum_j p_i^T C_{ij} p_j + \sum_{ij} p_{ij} u_{ij} - \sum_i \log \sum_h \exp(u_{ih})$$

$j \neq i$

$$\frac{\partial L(p,u)}{\partial p_i} = 0 \Rightarrow u_i = -\sum_j C_{ij} p_j$$

M step

$j \neq i$

$$L(p,u) = \sum_i \sum_j p_i^T C_{ij} p_j + \sum_i p_i^T u_i - \sum_i \log \sum_h \exp(u_{ih})$$

$$\frac{\partial L(p,u)}{\partial u_{ij}} = 0 \Rightarrow p_{ij} = \frac{exp(u_{ij})}{\sum_h \exp(u_{ih})},$$

Expectation step

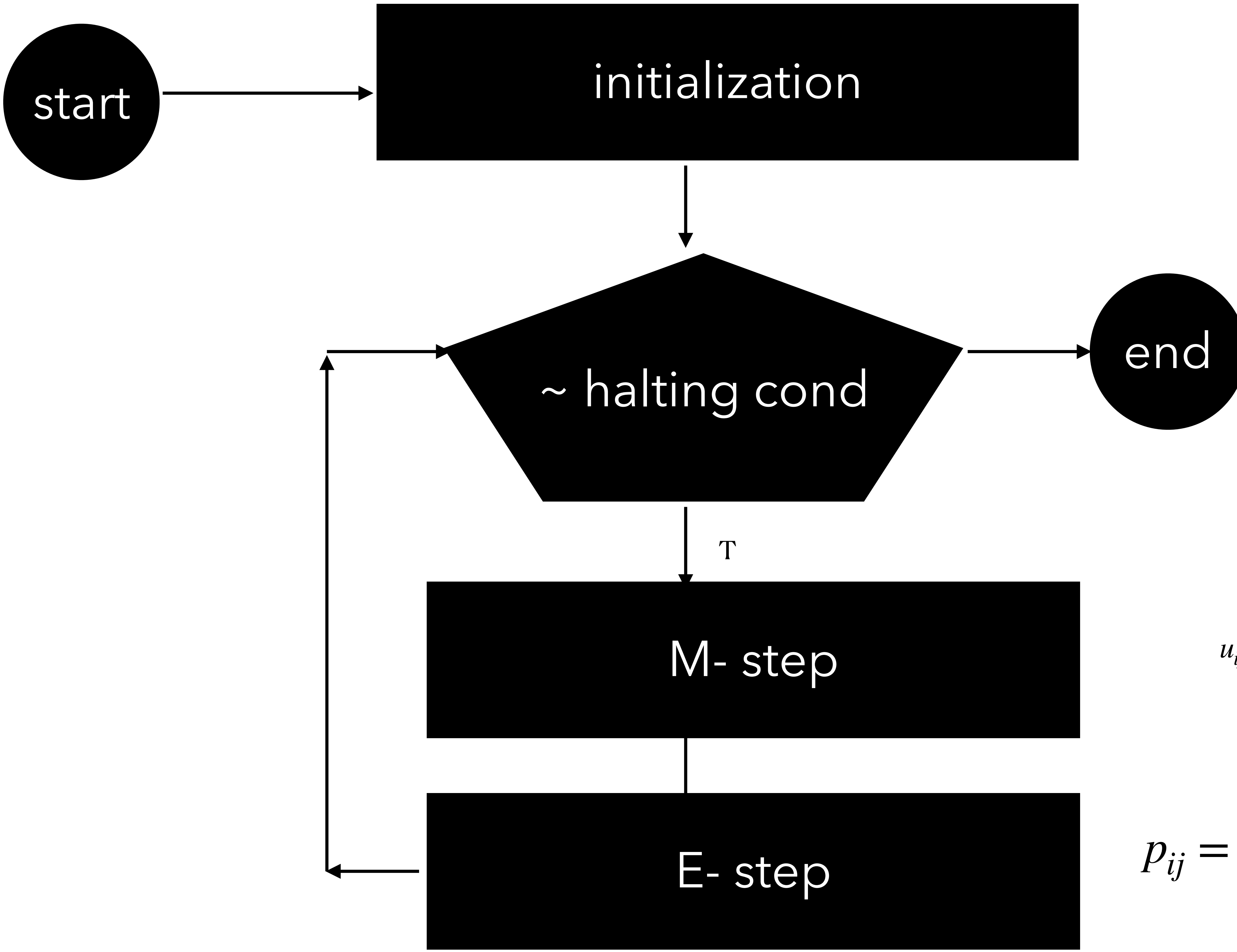
Thermal equilibrium

$$u_{ij} = -\sum_j C_{ij} p_{ij}$$

Maximization step

$$p_{ij} = \frac{exp(u_{ij})}{\sum_h \exp(u_{ih})},$$

Expectation step



$$u_{ij} = -\sum_{j} C_{ij} p_{ij}$$

$$p_{ij} = \frac{exp(u_{ij})}{\sum_h \exp(u_{ih})},$$

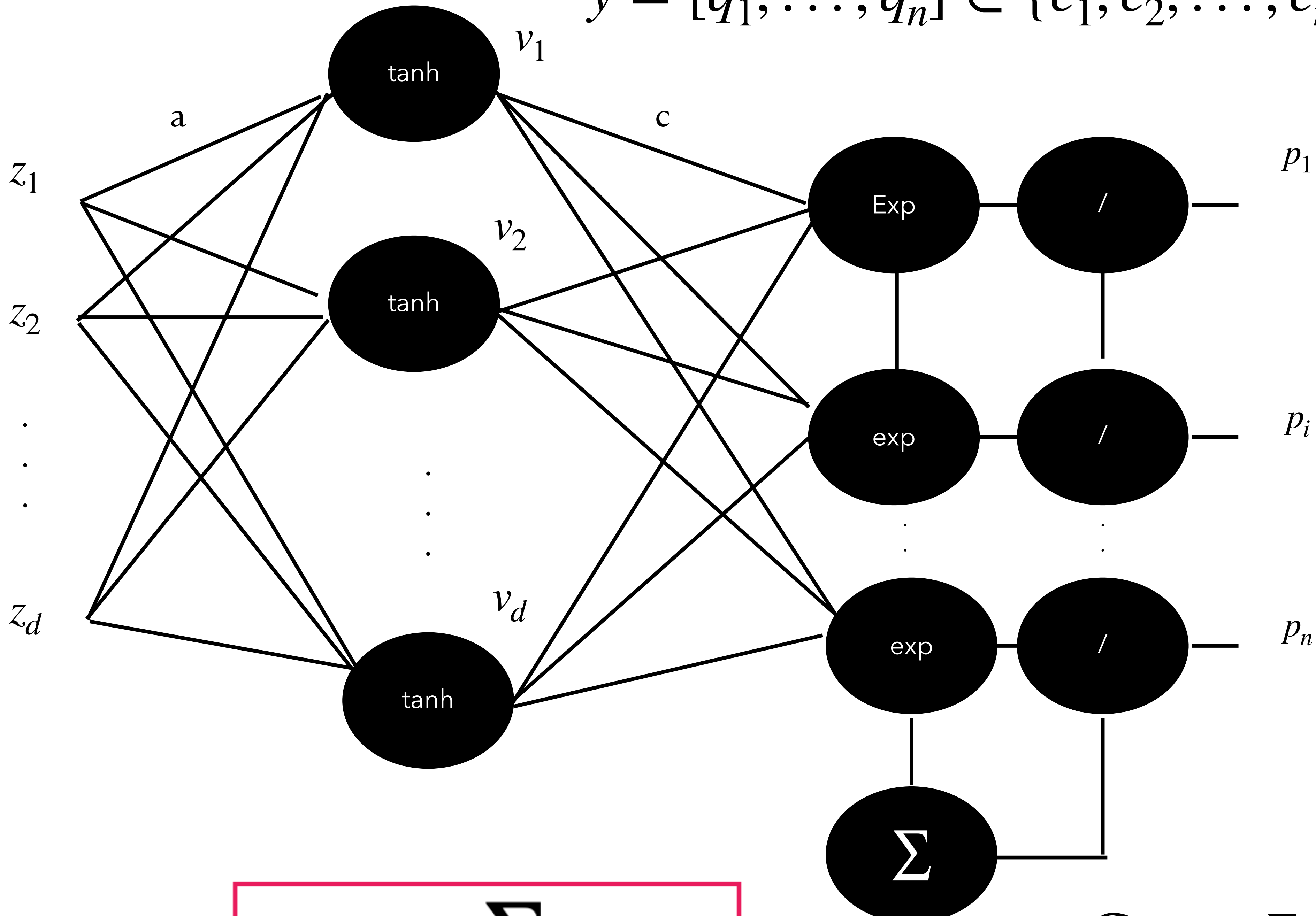
Classification neural network

$$h_j = \sum_{k=1}^{d} a_{jk} v_k \qquad v = [v_1, \ldots, v_d]$$

$$v_j = tanh(h_j)$$

$$\boxed{\begin{aligned} u_i &= -\sum_j C_{ij} v_j \\ p_{ij} &= \frac{exp(u_{ij})}{\sum_h \exp(u_{ih})} \end{aligned}}$$

$$y = [q_1, \ldots, q_n] \in \{e_1, e_2, \ldots, e_n\}$$



$$\boxed{\begin{aligned} u_i &= -\sum_j C_{ij} v_j \\ p_i &= \frac{exp(u_i)}{\sum_h \exp(u_h)} \end{aligned}}$$

Cross Entropy,
softmaxloss

$$Q(q\,||\,p) = \sum_i q_i ln \frac{q_i}{p_i}$$

Let $y = e_b$. Find

1. $$\frac{dQ}{du_i} = ? \quad for \quad i = b$$

2. $$\frac{dQ}{du_i} = ? \quad for \quad i \neq b$$

3. $$\frac{dQ}{da_{jk}} = ?$$