MLE for n-gram LM

$$\max \ell(\theta) = \sum_{i=1}^{N} \sum_{j=1}^{n} \log \underbrace{p(\underset{= w' \in V}{\overset{= w \in V}{x_j} \mid x_{j-1}})}_{\theta_{w|w'}}$$

$$\text{s.t.} \sum_{w \in V} \theta_{w|w'} = 1 \quad \forall w' \in V$$

$$\theta_{w|w'} \geq 0$$

$$L(\theta, \lambda) = \ell(\theta) - \lambda\left(\sum_{w \in V} \theta_{w|w'} - 1\right)$$

$$\frac{\partial L}{\partial \theta_{w|w'}} = 0$$

Idea:

"test"

train	valid

$\text{count}^*(x) \longrightarrow \text{count}(x)$

Leave one out (LOO) validation

$M = (k\text{-fold})$

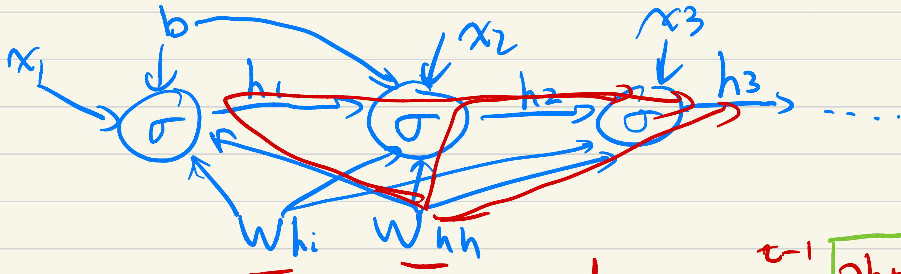
1	2	. . .	k

1 val 1 train M−1

2 val

$$h_t = \sigma(\underline{W}_{hh} h_{t-1} + W_{hi} x_t + b)$$

$$\frac{\partial l_t}{\partial W_{hh}} = \underbrace{\frac{\partial l_t}{\partial o_t} \frac{\partial o_t}{\partial h_t}} \boxed{\frac{\partial h_t}{\partial W_{hh}}}$$



$$\frac{\partial h_t}{\partial W_{hh}} = \sum_{i=1}^{t-1} \boxed{\frac{\partial h_t}{\partial h_i}} \frac{\partial h_i}{\partial W_{hh}}$$

$$\frac{\partial h_t}{\partial h_i} = \underbrace{\frac{\partial h_t}{\partial h_{t-1}}} \frac{\partial h_{t-1}}{\partial h_{t-2}} \cdots \frac{\partial h_{i+1}}{\partial h_i}$$

$$= \prod_{j=t-1}^{i} \sigma' W_{hh}$$

$$W_{hh}^{k} = Q \Lambda^{k} Q$$