Търсене и извличане на информация. Приложение на дълбоко машинно обучение

Стоян Михов

Лекция 11: Рекурентни невронни мрежи. Не-марковски невронен езиков модел.

План на лекцията

1. **Формалности за курса (5 мин)**
2. Марковски невронен езиков (10 мин)
3. Рекурентен езиков модел (10 мин)
4. Пропагиране напред при рекурентна невронна мрежа (15 мин)
5. Обучение на рекурентна невронна мрежа (20 мин)
6. Пропагиране назад при рекурентна невронна мрежа (15 мин)
7. Приложения на езиковите модели (15 мин)

Формалности

- Засега ще провеждаме занятията онлайн всяка сряда от 8:15 до 12:00 часа.
- Засега ще използваме платформата Google meet: meet.google.com/hue-frfx-axb
- Днес ще използваме едновременно слайдове и бяла дъска. Моля следете съответния екран.
- До края на тази седмица в Moodle ще бъде публикувано Домашно задание №2
- Домашното задание следва да бъде предадено до края на деня на 05.01.2021г.
- Лекция 11 се базира на глава 14 от втория учебник.

План на лекцията

1. Формалности за курса (5 мин)
2. **Марковски невронен езиков (10 мин)**
3. Рекурентен езиков модел (10 мин)
4. Пропагиране напред при рекурентна невронна мрежа (15 мин)
5. Обучение на рекурентна невронна мрежа (20 мин)
6. Пропагиране назад при рекурентна невронна мрежа (15 мин)
7. Приложения на езиковите модели (15 мин)

Марковски k-грамен невронен езиков модел

Миналата лекция разгледахме модела на Бенджио и съавтори:

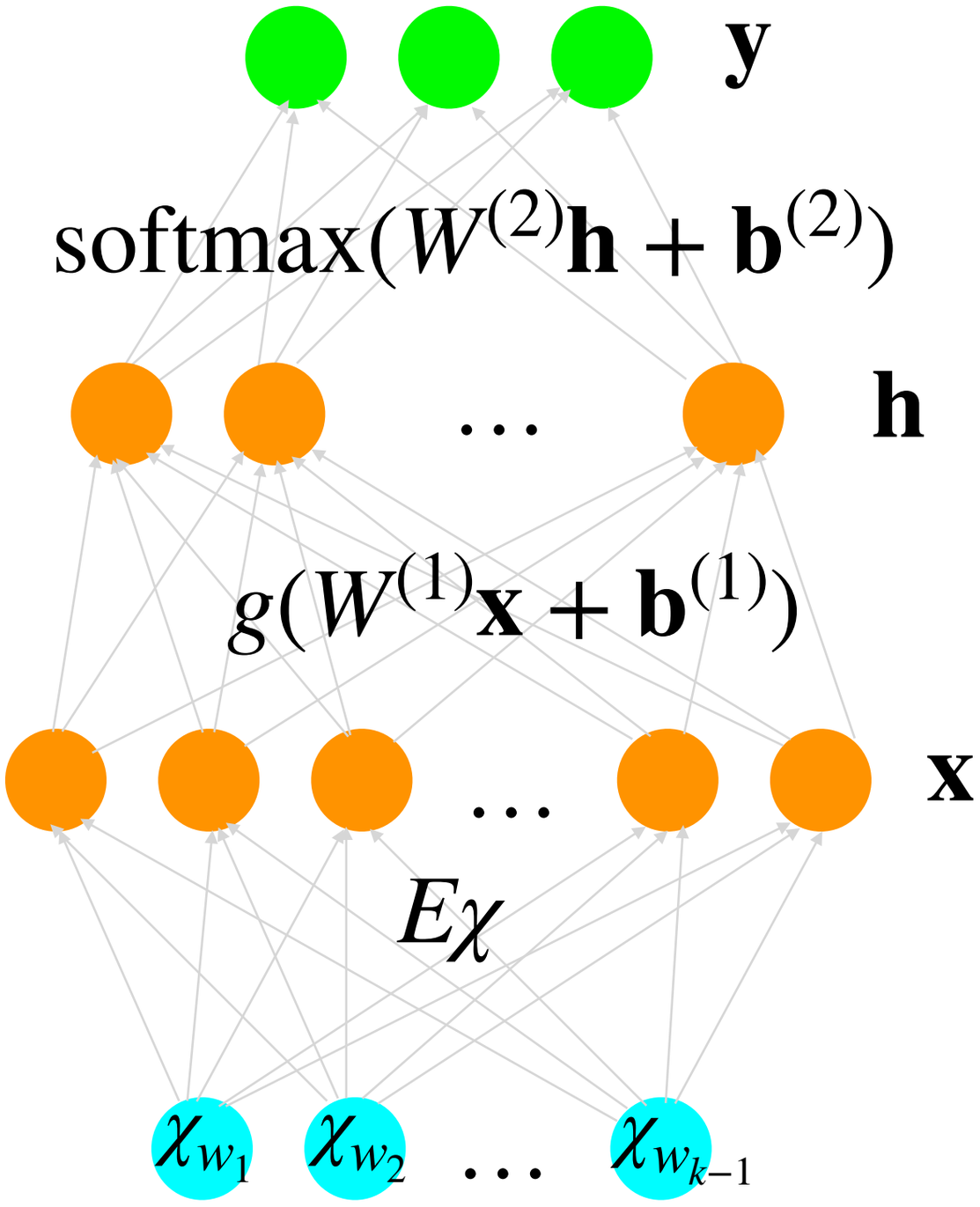
$$\mathbf{y} = \text{softmax}(W^{(2)}\mathbf{h} + \mathbf{b}^{(2)})$$

$$\mathbf{h} = g(W^{(1)}\mathbf{x} + \mathbf{b}^{(1)})$$

$$\mathbf{x} = \begin{bmatrix} E\chi_{w_1} \\ \vdots \\ E\chi_{w_{k-1}} \end{bmatrix}$$

Езиков модел:

$$\Pr[w \mid w_1 w_2 \ldots w_{k-1}] = \mathbf{y}_w$$



Невронни езикови модели с контекст с фиксирана дължина

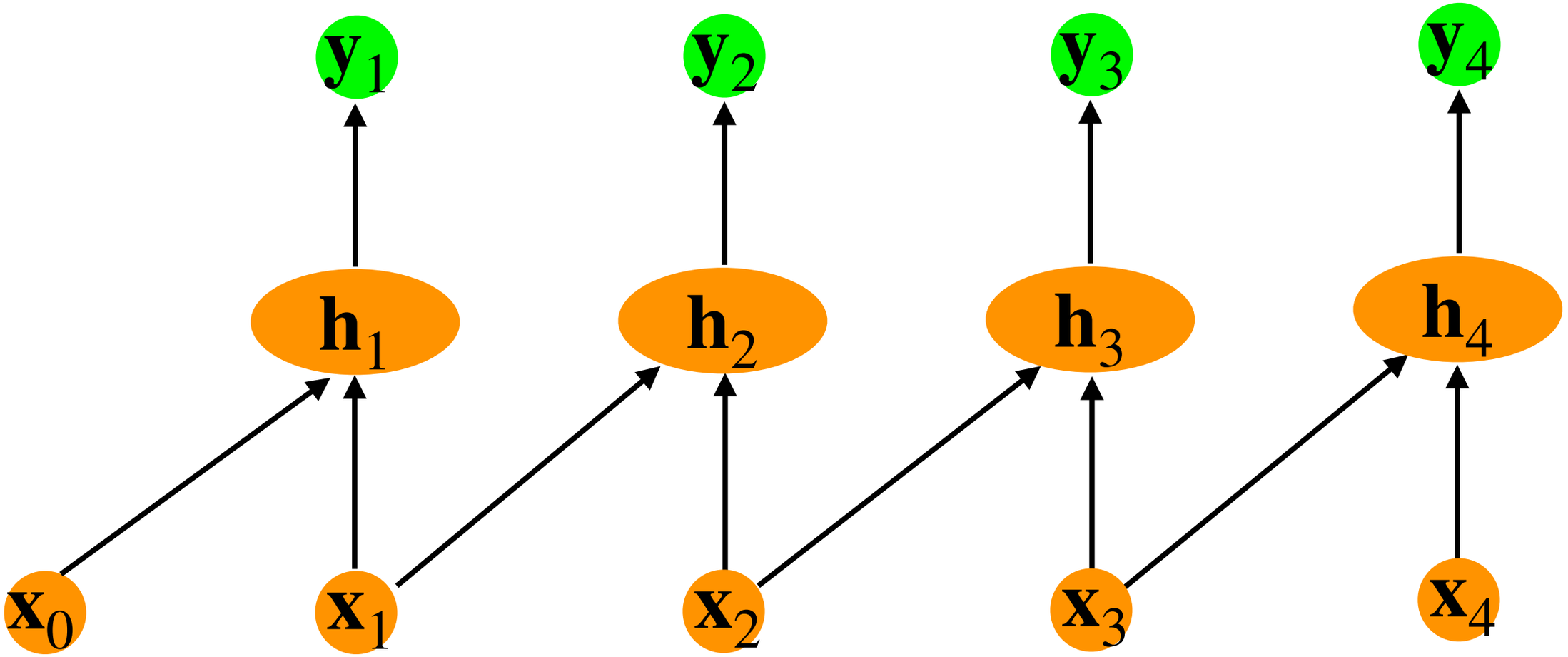
- Преимущества:
  - Реализират естествено изглаждане на ненаблюдавани k-грами, което води до значително подобрение на перплексията.
  - Големината на модела не зависи от големината на корпуса.
- Недостатъци:
  - Контекстът е фиксиран — винаги може да се намери пример, при който контекста да не е достатъчно голям.
  - С увеличаване на контекста расте (линейно) големината на модела.
  - При обучението на матрицата $W^{(1)}$ секциите за отделните думи се тренират независимо.
- Въпрос: Има ли решение, при което да няма ограничение за дължината на контекста?

План на лекцията

1. Формалности за курса (5 мин)
2. Марковски невронен езиков (10 мин)
3. **Рекурентен езиков модел (10 мин)**
4. Пропагиране напред при рекурентна невронна мрежа (15 мин)
5. Обучение на рекурентна невронна мрежа (20 мин)
6. Пропагиране назад при рекурентна невронна мрежа (15 мин)
7. Приложения на езиковите модели (15 мин)

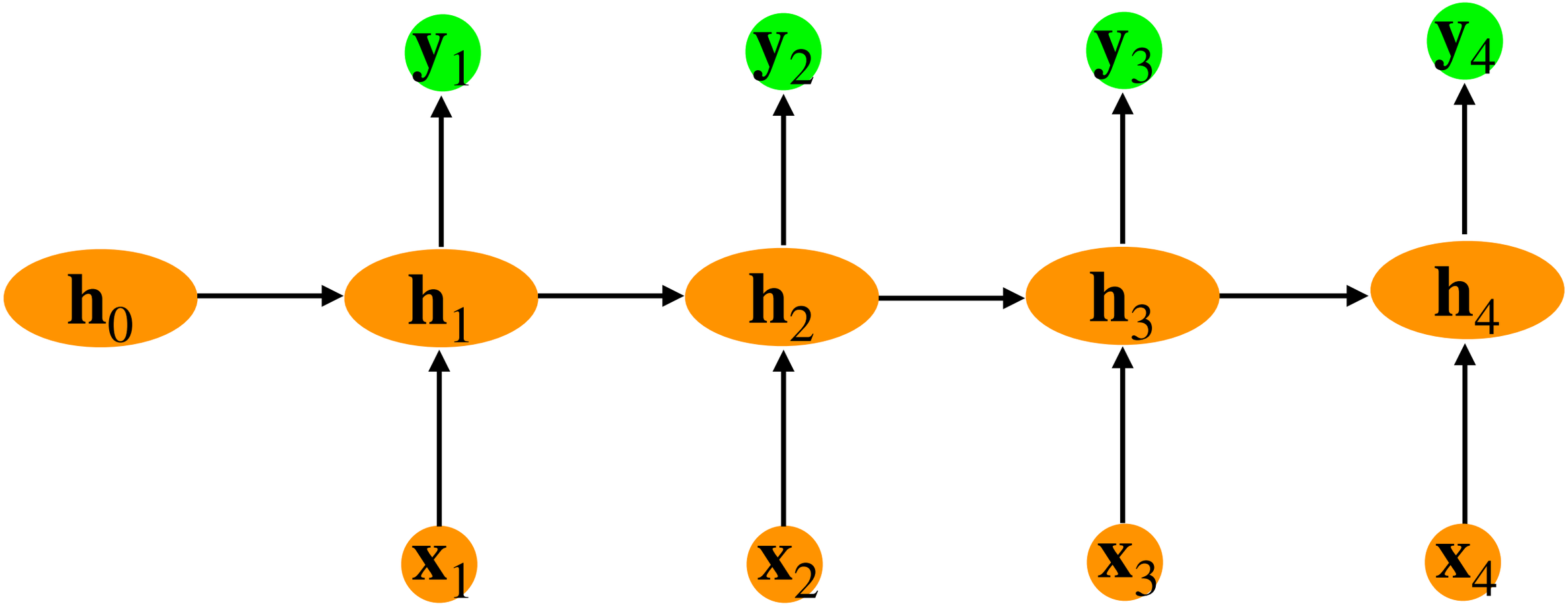
Рекурентни невронни мрежи

- Нека разгледаме триграмен невронен езиков модел.
- Векторите $\mathbf{h}$ представляват влагане на контекста, от което се получава вероятностно разпределение $\mathbf{y}$.
- Може ли да получим новия контекст от предишния, като го допълним със следващата дума?



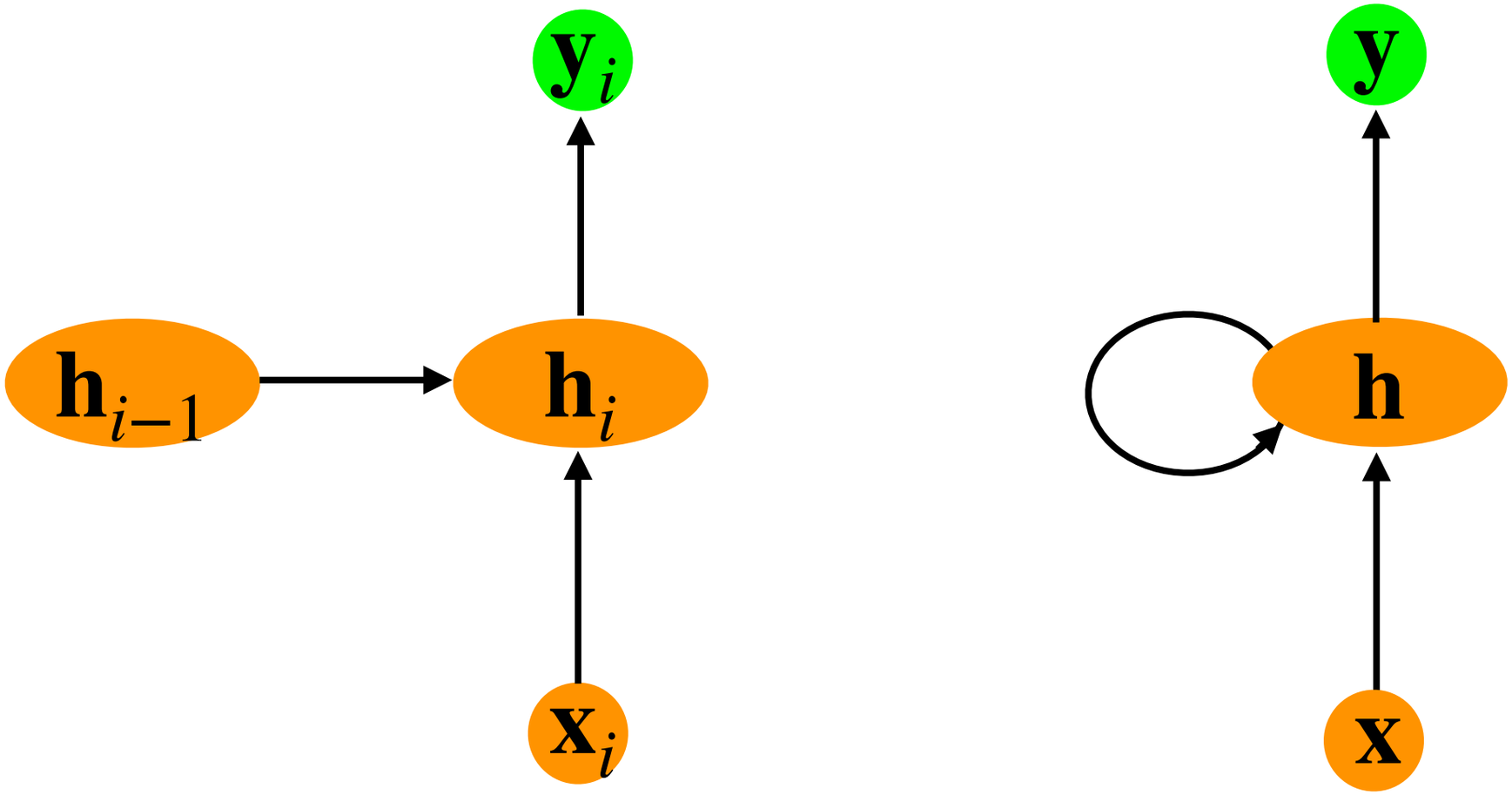
Рекурентни невронни мрежи

- Искаме да натрупваме към контекста до момента новата дума.



Рекурентни невронни мрежи

- Нека да се абстрахираме от поредния номер



Рекурентни невронни мрежи

$$\mathbf{y}_i = \text{softmax}(U\mathbf{h}_i)$$

$$\mathbf{h}_i = g(W\mathbf{h}_{i-1} + V\mathbf{x}_i)$$

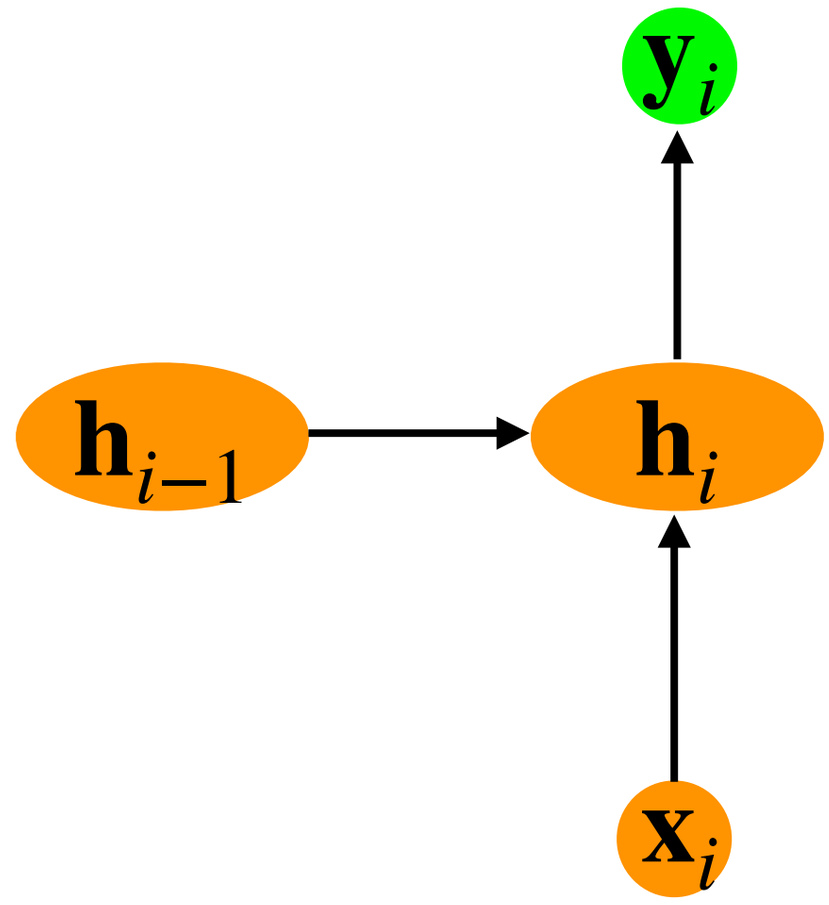
$$\mathbf{x}_i = E\chi_{w_i}$$

$$\chi_{w_i} \in \mathbb{R}^{|L|}, E \in \mathbb{R}^{M \times |L|},$$

$$\mathbf{x}_i \in \mathbb{R}^M, V \in \mathbb{R}^{N \times M},$$

$$\mathbf{h}_i, \mathbf{h}_{i-1} \in \mathbb{R}^N, W \in \mathbb{R}^{N \times N},$$

$$U \in \mathbb{R}^{|L| \times N}, \mathbf{y}_i \in \mathbb{R}^{|L|}$$



Езиков модел с рекурентна невронна мрежа

При входен текст $w_1 w_2 \ldots w_n$ с произволна дължина n, моделираме вероятностното разпределение за следващата дума като:

$$\Pr_{E,V,W,U}[w \mid w_1 w_2 \ldots w_n] = (\mathbf{y}_n)_w, \text{ където}$$

- $\mathbf{y}_i = \text{softmax}(U\mathbf{h}_i)$
- $\mathbf{h}_i = g(W\mathbf{h}_{i-1} + VE\chi_{w_i})$, за $i = 1,2,\ldots,n$, $\mathbf{h}_0$ - фиксирано

Рекурентната невронна мрежа избягва Марковското ограничение

План на лекцията

1. Формалности за курса (5 мин)
2. Марковски невронен езиков (10 мин)
3. Рекурентен езиков модел (10 мин)
4. **Пропагиране напред при рекурентна невронна мрежа (15 мин)**
5. Обучение на рекурентна невронна мрежа (20 мин)
6. Пропагиране назад при рекурентна невронна мрежа (15 мин)
7. Приложения на езиковите модели (15 мин)

Пропагиране напред при рекурентна невронна мрежа

Параметри: $\boldsymbol{E} \in \mathbb{R}^{M \times |L|}, \boldsymbol{V} \in \mathbb{R}^{N \times M}, \boldsymbol{W} \in \mathbb{R}^{N \times N}, \boldsymbol{U} \in \mathbb{R}^{|L| \times N}$

Вход: $\mathbf{h}_0 \in \mathbb{R}^N, \chi_{w_1}, \chi_{w_2}, \chi_{w_3}, \ldots \in \mathbb{R}^{|L|}$

Пропагиране напред при рекурентна невронна мрежа

Параметри: $E \in \mathbb{R}^{M \times |L|}, V \in \mathbb{R}^{N \times M}, W \in \mathbb{R}^{N \times N}, U \in \mathbb{R}^{|L| \times N}$

Вход: $\mathbf{h}_0 \in \mathbb{R}^N, \chi_{w_1}, \chi_{w_2}, \chi_{w_3}, \ldots \in \mathbb{R}^{|L|}$

$$\mathbf{x}_1 = E\chi_{w_1}, \quad \mathbf{h}_1 = g(W\mathbf{h}_0 + V\mathbf{x}_1), \quad \mathbf{y}_1 = \mathrm{softmax}(U\mathbf{h}_1)$$

Пропагиране напред при рекурентна невронна мрежа

Параметри: $E \in \mathbb{R}^{M \times |L|}, V \in \mathbb{R}^{N \times M}, W \in \mathbb{R}^{N \times N}, U \in \mathbb{R}^{|L| \times N}$

Вход: $\mathbf{h}_0 \in \mathbb{R}^N, \chi_{w_1}, \chi_{w_2}, \chi_{w_3}, \ldots \in \mathbb{R}^{|L|}$

$$\mathbf{x}_1 = E\chi_{w_1}, \quad \mathbf{h}_1 = g(W\mathbf{h}_0 + V\mathbf{x}_1), \quad \mathbf{y}_1 = \text{softmax}(U\mathbf{h}_1)$$

$$\mathbf{x}_2 = E\chi_{w_2}, \quad \mathbf{h}_2 = g(W\mathbf{h}_1 + V\mathbf{x}_2), \quad \mathbf{y}_2 = \text{softmax}(U\mathbf{h}_2)$$

Пропагиране напред при рекурентна невронна мрежа

Параметри: $E \in \mathbb{R}^{M \times |L|}, V \in \mathbb{R}^{N \times M}, W \in \mathbb{R}^{N \times N}, U \in \mathbb{R}^{|L| \times N}$

Вход: $\mathbf{h}_0 \in \mathbb{R}^N, \chi_{w_1}, \chi_{w_2}, \chi_{w_3}, \ldots \in \mathbb{R}^{|L|}$

$$\mathbf{x}_1 = E\chi_{w_1}, \quad \mathbf{h}_1 = g(W\mathbf{h}_0 + V\mathbf{x}_1), \quad \mathbf{y}_1 = \text{softmax}(U\mathbf{h}_1)$$

$$\mathbf{x}_2 = E\chi_{w_2}, \quad \mathbf{h}_2 = g(W\mathbf{h}_1 + V\mathbf{x}_2), \quad \mathbf{y}_2 = \text{softmax}(U\mathbf{h}_2)$$

$$\mathbf{x}_3 = E\chi_{w_3}, \quad \mathbf{h}_3 = g(W\mathbf{h}_2 + V\mathbf{x}_3), \quad \mathbf{y}_3 = \text{softmax}(U\mathbf{h}_3)$$

$$\vdots$$

Пропагиране напред при рекурентна невронна мрежа

Параметри: $E \in \mathbb{R}^{M \times |L|}, V \in \mathbb{R}^{N \times M}, W \in \mathbb{R}^{N \times N}, U \in \mathbb{R}^{|L| \times N}$

Вход: $\mathbf{h}_0 \in \mathbb{R}^N, \chi_{w_1}, \chi_{w_2}, \chi_{w_3}, \ldots \in \mathbb{R}^{|L|}$

$$\mathbf{x}_1 = E\chi_{w_1}, \quad \mathbf{h}_1 = g(W\mathbf{h}_0 + V\mathbf{x}_1), \quad \mathbf{y}_1 = \text{softmax}(U\mathbf{h}_1)$$

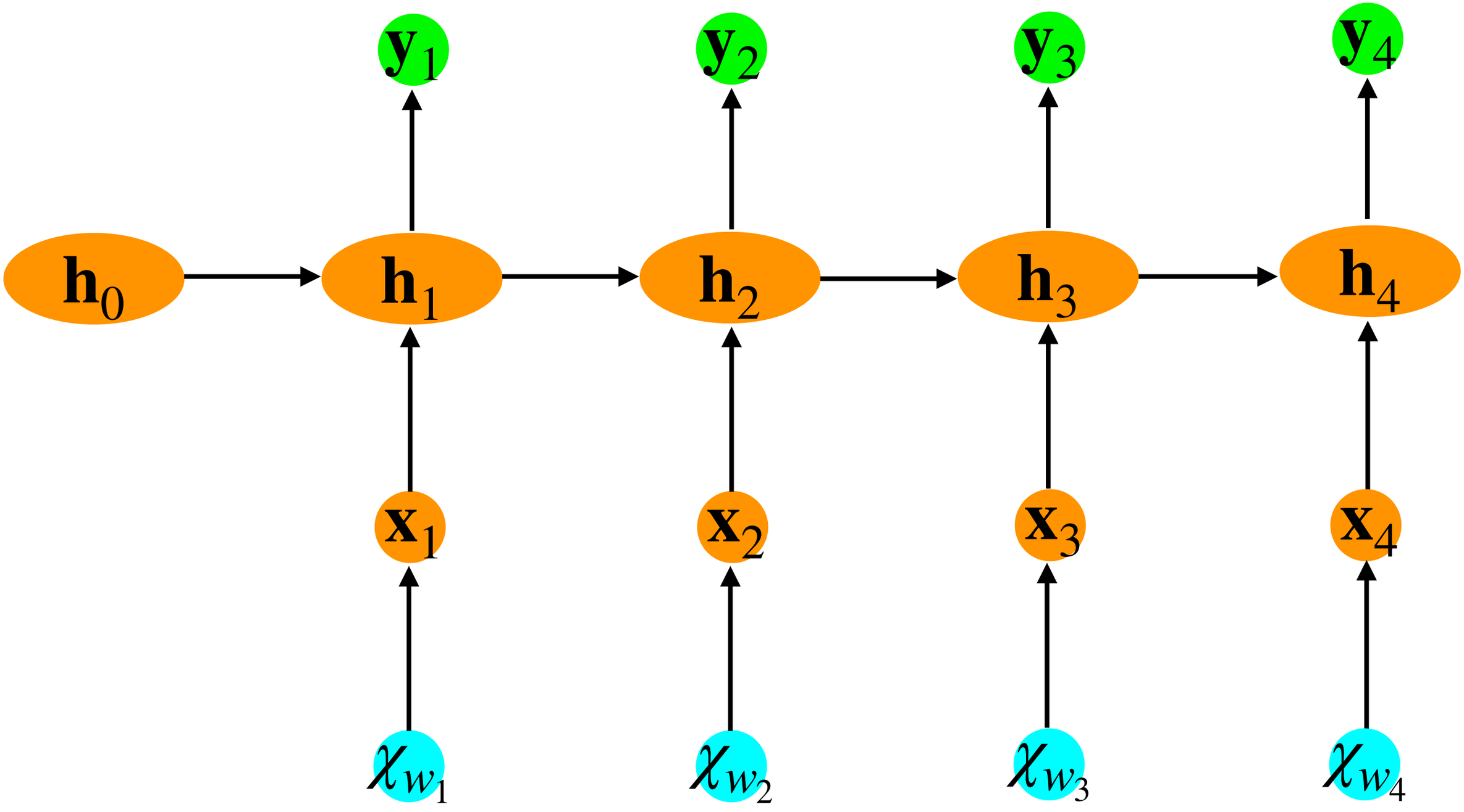
$$\mathbf{x}_2 = E\chi_{w_2}, \quad \mathbf{h}_2 = g(W\mathbf{h}_1 + V\mathbf{x}_2), \quad \mathbf{y}_2 = \text{softmax}(U\mathbf{h}_2)$$

$$\mathbf{x}_3 = E\chi_{w_3}, \quad \mathbf{h}_3 = g(W\mathbf{h}_2 + V\mathbf{x}_3), \quad \mathbf{y}_3 = \text{softmax}(U\mathbf{h}_3)$$

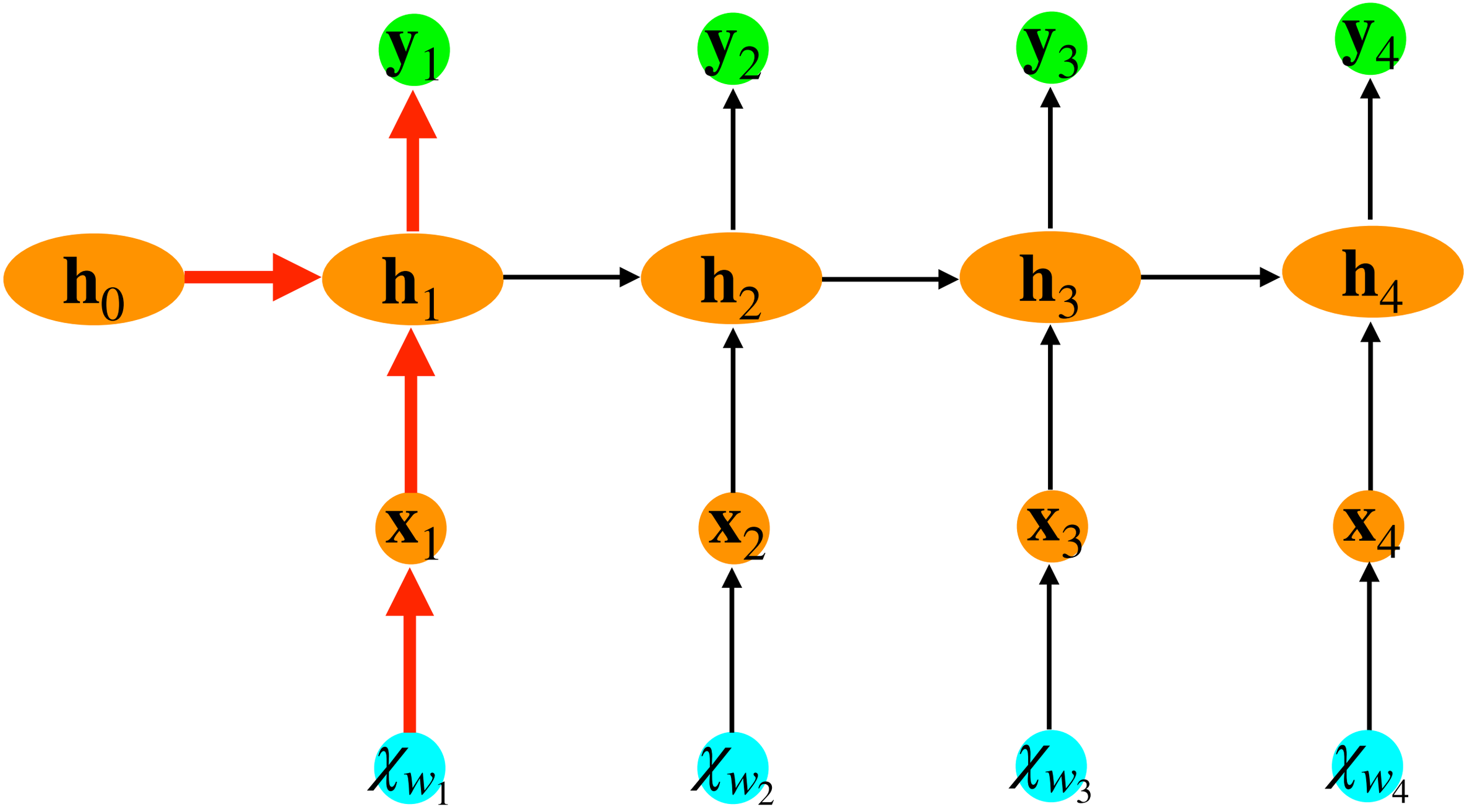
$\vdots$

Изход: $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots \in \mathbb{R}^{|L|}$

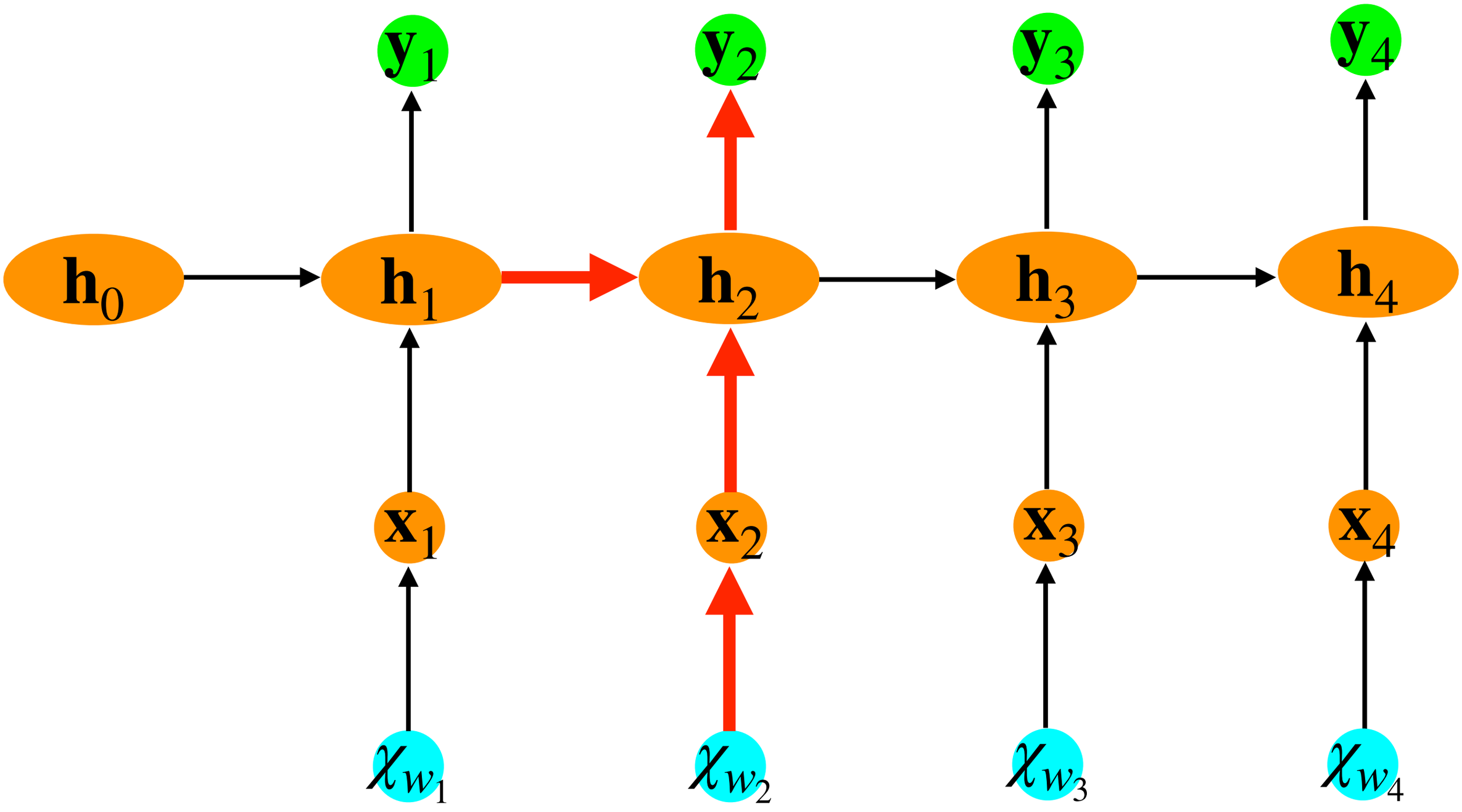
Пропагиране напред при рекурентна невронна мрежа



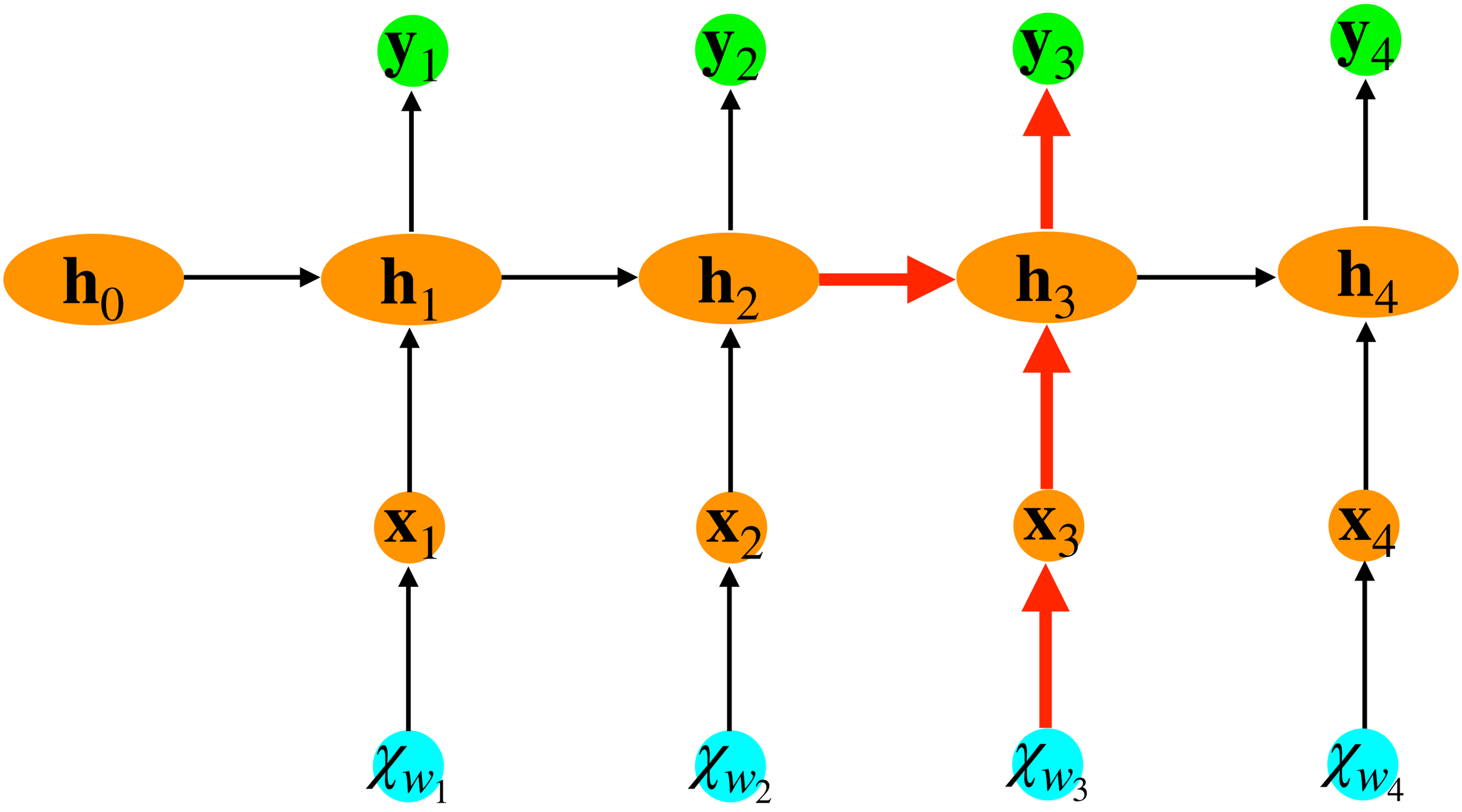
Пропагиране напред при рекурентна невронна мрежа



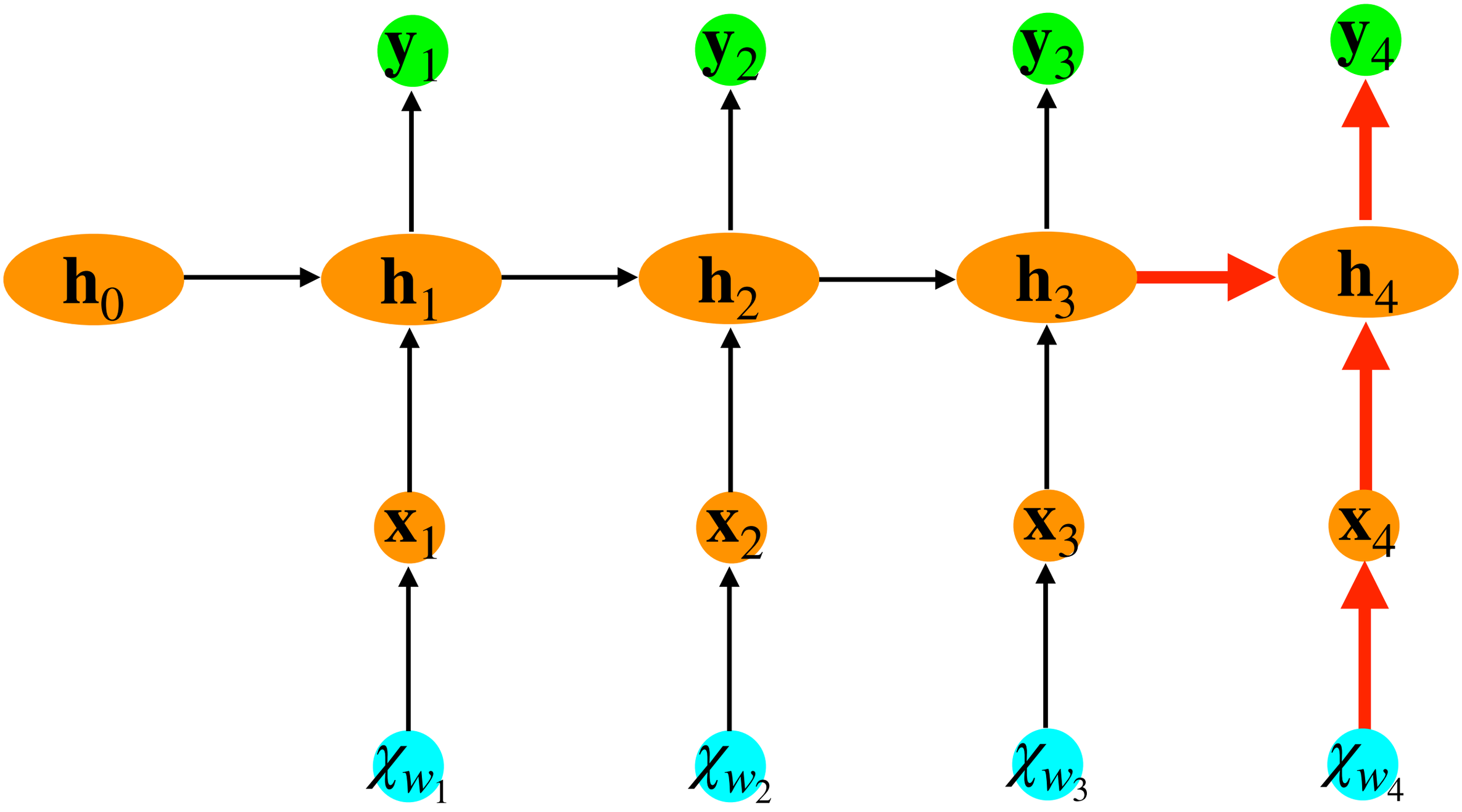
Пропагиране напред при рекурентна невронна мрежа



Пропагиране напред при рекурентна невронна мрежа



Пропагиране напред при рекурентна невронна мрежа



План на лекцията

1. Формалности за курса (5 мин)
2. Марковски невронен езиков (10 мин)
3. Рекурентен езиков модел (10 мин)
4. Пропагиране напред при рекурентна невронна мрежа (15 мин)
5. **Обучение на рекурентна невронна мрежа (20 мин)**
6. Пропагиране назад при рекурентна невронна мрежа (15 мин)
7. Приложения на езиковите модели (15 мин)

Обучение на рекурентна невронна мрежа

- Ще използваме минимизация на кросентропията, което е еквивалентно на максимизация на правдоподобието.

- Ще предполагаме, че ни е даден корпус $\mathbf{X}$. Елементите на корпуса са документи/изречения. С $w \in \mathbf{X}$ означаваме даден документ. С w_i означаваме номера на i-тия терм в документа w.

$$
\begin{aligned}
H_{\mathbf{X}}(E, V, W, U) &= -\frac{1}{\|\mathbf{X}\|} \sum_{w \in \mathbf{X}} \sum_{i=1}^{|w|} \log \Pr_{E,V,W,U}[w_{i+1} \mid w_1 w_2 \ldots w_i] \\
&= -\frac{1}{\|\mathbf{X}\|} \sum_{w \in \mathbf{X}} \sum_{i=1}^{|w|} \log(\mathbf{y}_i)_{w_{i+1}} \\
&= -\frac{1}{\|\mathbf{X}\|} \sum_{w \in \mathbf{X}} \sum_{i=1}^{|w|} \log \operatorname{softmax}(Ug(W\mathbf{h}_{i-1} + VE\chi_{w_i}))_{w_{i+1}}
\end{aligned}
$$

Обучение на рекурентна невронна мрежа

- Нека поточковата кросентропия в точката w_{i+1} означим с $H_{w_{i+1}}(E, V, W, U) = -\log(\mathbf{y}_i)_{w_{i+1}}$. Тогава:

$$
\begin{aligned}
H_{w_4} &= -\log \operatorname{softmax}(U\mathbf{h}_3)_{w_4} = \\
&= -\log \operatorname{softmax}(Ug(W\mathbf{h}_2 + VE\chi_{w_3}))_{w_4} = \\
&= -\log \operatorname{softmax}(Ug(Wg(W\mathbf{h}_1 + VE\chi_{w_2}) + VE\chi_{w_3}))_{w_4} = \\
&= -\log \operatorname{softmax}(Ug(Wg(Wg(W\mathbf{h}_0 + VE\chi_{w_1}) + VE\chi_{w_2}) + VE\chi_{w_3}))_{w_4}
\end{aligned}
$$

Ще трябва да намерим градиентите по параметрите.

- $\frac{\partial H_{w_{i+1}}}{\partial U} = -\frac{\partial}{\partial U} \log \text{softmax}(U\mathbf{h}_i)_{w_{i+1}} = (\bar{\delta}_{w_{i+1}} - \text{softmax}(U\mathbf{h}_i)) \otimes \mathbf{h}_i$
- $\frac{\partial H_{w_{i+1}}}{\partial W} = -\frac{\partial}{\partial W} \log \text{softmax}(U\mathbf{h}_i)_{w_{i+1}}$
- Нека положим $\mathbf{z}_i = W\mathbf{h}_{i-1} + VE\chi_{w_i}$. Тогава $\mathbf{h}_i = g(\mathbf{z}_i)$.
- $\frac{\partial H_{w_{i+1}}}{\partial W} = (\bar{\delta}_{w_{i+1}} - \text{softmax}(U\mathbf{h}_i))^\top U \frac{\partial \mathbf{h}_i}{\partial W}$
- $\frac{\partial \mathbf{h}_i}{\partial W} = \frac{\partial}{\partial W} g(W\mathbf{h}_{i-1} + VE\chi_{w_i}) = g'(\mathbf{z}_i)\left(\mathbf{I}_N \otimes \mathbf{h}_{i-1} + W\frac{\partial \mathbf{h}_{i-1}}{\partial W}\right)$ където:
  - $g'(\mathbf{a})$ е диагонална матрица с диагонал $g'(\mathbf{a}_i)$.
  - Ако $A \in \mathbb{R}^{L\times M}$е матрица и $\mathbf{b} \in \mathbb{R}^N$ е вектор то $A \otimes \mathbf{b} \in \mathbb{R}^{L\times M\times N}$ и $(A \otimes \mathbf{b})_{k,i,j} = A_{k,i}\mathbf{b}_j$.
  - $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ е единичната матрица.

$$
\begin{aligned}
\frac{\partial H_{w_4}}{\partial W} &= (\bar{\delta}_{w_4} - \text{softmax}(U\mathbf{h}_3))^\top U \frac{\partial \mathbf{h}_3}{\partial W} = \\
&= (\bar{\delta}_{w_4} - \text{softmax}(U\mathbf{h}_3))^\top U g'(\mathbf{z}_3)\left(\mathbf{I}_N \otimes \mathbf{h}_2 + W\frac{\partial \mathbf{h}_2}{\partial W}\right) = \\
&= (\bar{\delta}_{w_4} - \text{softmax}(U\mathbf{h}_3))^\top U g'(\mathbf{z}_3)\left(\mathbf{I}_N \otimes \mathbf{h}_2 + W g'(\mathbf{z}_2)\left(\mathbf{I}_N \otimes \mathbf{h}_1 + W\frac{\partial \mathbf{h}_1}{\partial W}\right)\right) = \\
&= (\bar{\delta}_{w_4} - \text{softmax}(U\mathbf{h}_3))^\top U g'(\mathbf{z}_3)\left(\mathbf{I}_N \otimes \mathbf{h}_2 + W g'(\mathbf{z}_2)\left(\mathbf{I}_N \otimes \mathbf{h}_1 + W g'(\mathbf{z}_1)\left(\mathbf{I}_N \otimes \mathbf{h}_0 + W\frac{\partial \mathbf{h}_0}{\partial W}\right)\right)\right) = \\
&= (\bar{\delta}_{w_4} - \text{softmax}(U\mathbf{h}_3))^\top U g'(\mathbf{z}_3)\left(\mathbf{I}_N \otimes \mathbf{h}_2 + W g'(\mathbf{z}_2)\left(\mathbf{I}_N \otimes \mathbf{h}_1 + W g'(\mathbf{z}_1)\mathbf{I}_N \otimes \mathbf{h}_0\right)\right) = \\
&= (\bar{\delta}_{w_4} - \text{softmax}(U\mathbf{h}_3))^\top U \left(g'(\mathbf{z}_3)\mathbf{I}_N \otimes \mathbf{h}_2 + g'(\mathbf{z}_3) W g'(\mathbf{z}_2)\left(\mathbf{I}_N \otimes \mathbf{h}_1 + W g'(\mathbf{z}_1)\mathbf{I}_N \otimes \mathbf{h}_0\right)\right) = \\
&= (\bar{\delta}_{w_4} - \text{softmax}(U\mathbf{h}_3))^\top U \left(g'(\mathbf{z}_3)\mathbf{I}_N \otimes \mathbf{h}_2 + W g'(\mathbf{z}_3) g'(\mathbf{z}_2)\mathbf{I}_N \otimes \mathbf{h}_1 + W^2 g'(\mathbf{z}_3) g'(\mathbf{z}_2) g'(\mathbf{z}_1)\mathbf{I}_N \otimes \mathbf{h}_0\right)
\end{aligned}
$$

$$
\frac{\partial H_{w_{i+1}}}{\partial W} = (\bar{\delta}_{w_{i+1}} - \text{softmax}(U\mathbf{h}_i))^\top U \sum_{j=1}^{i} W^{j-1}\left(\prod_{k=0}^{j-1} g'(\mathbf{z}_{i-k})\right)\mathbf{I}_N \otimes \mathbf{h}_{i-j}
$$

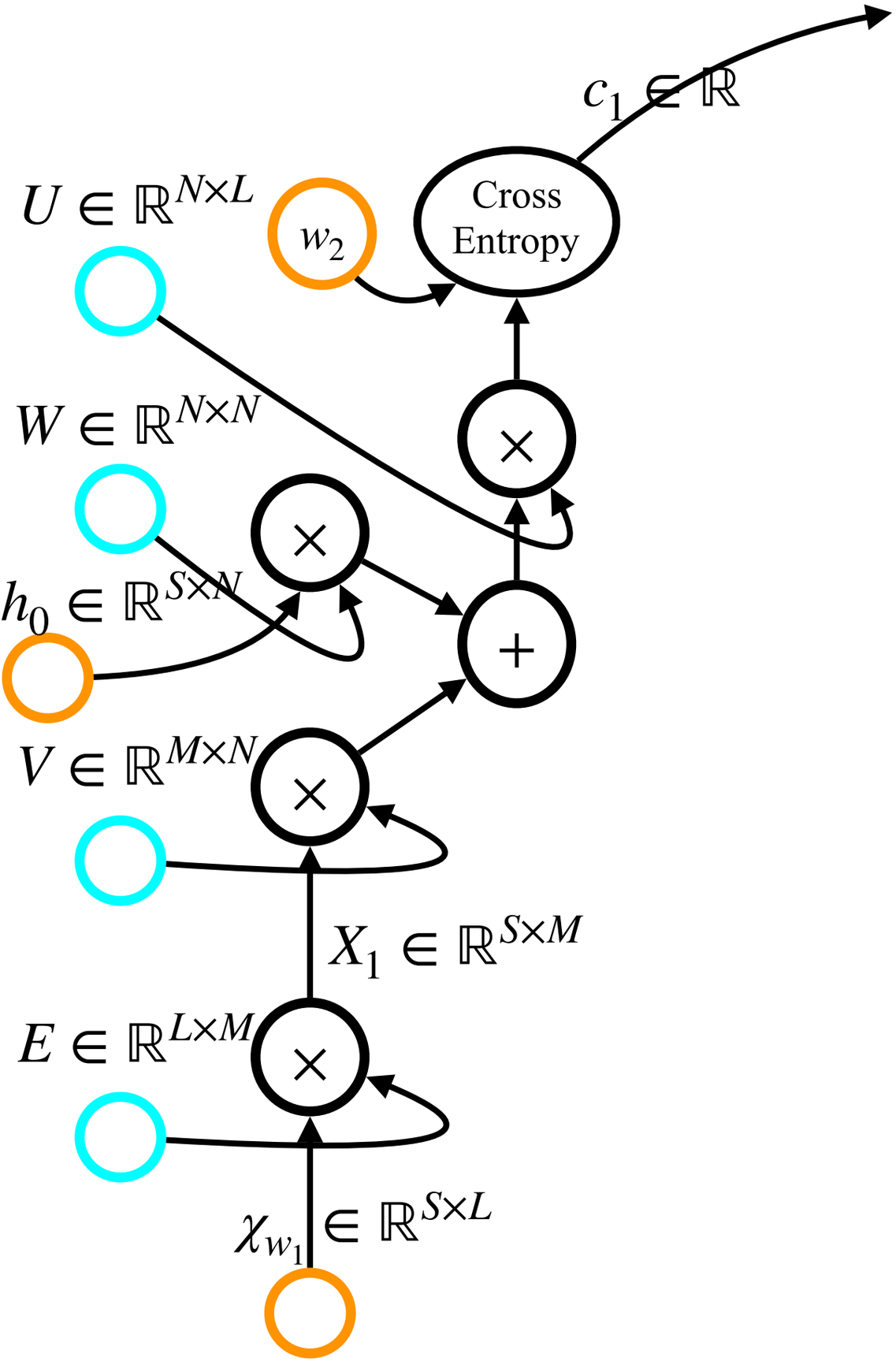
По подобен начин се изразяват $\frac{\partial H_{w_{i+1}}}{\partial V}$ и $\frac{\partial H_{w_{i+1}}}{\partial E}$.

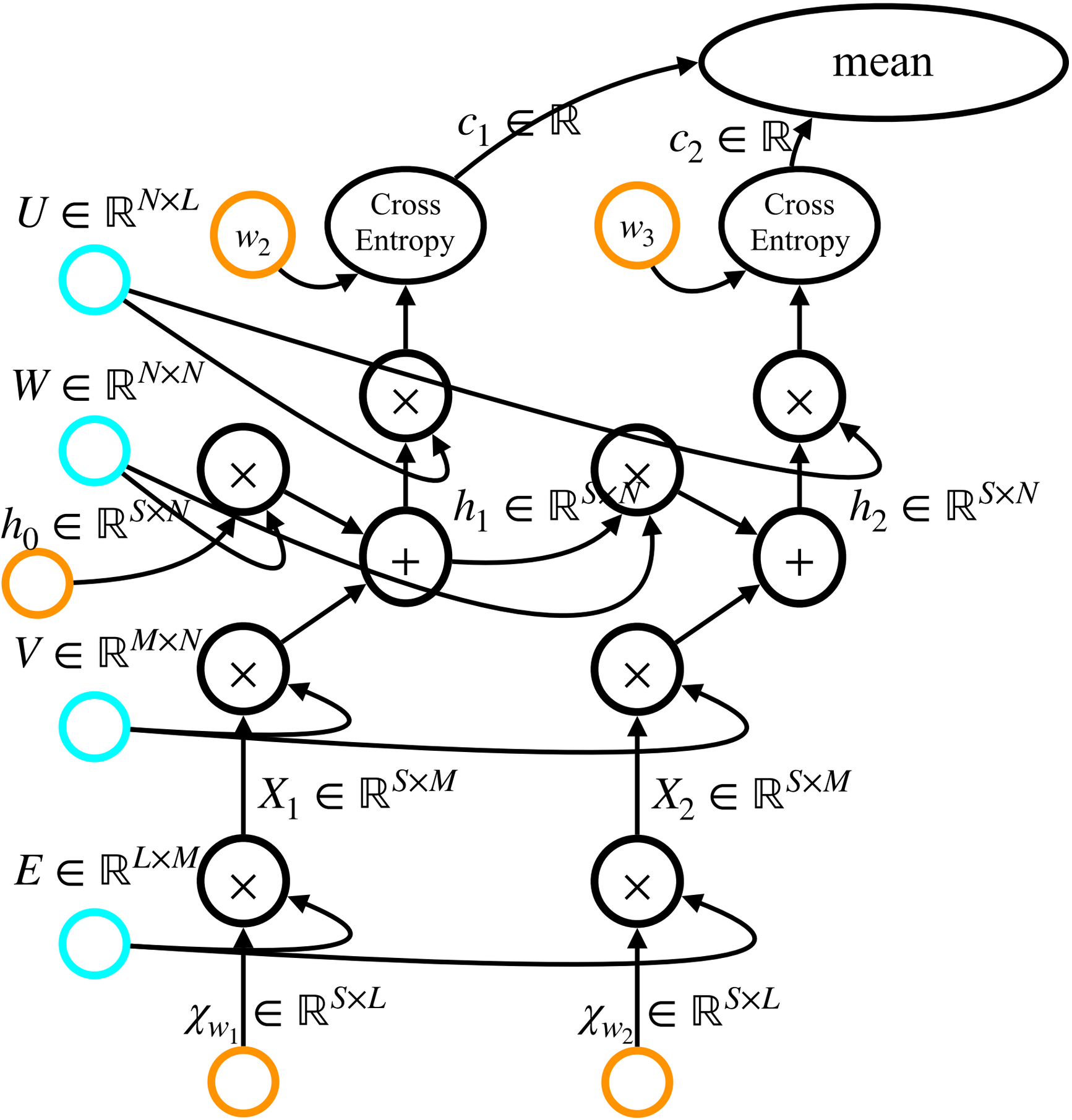
План на лекцията

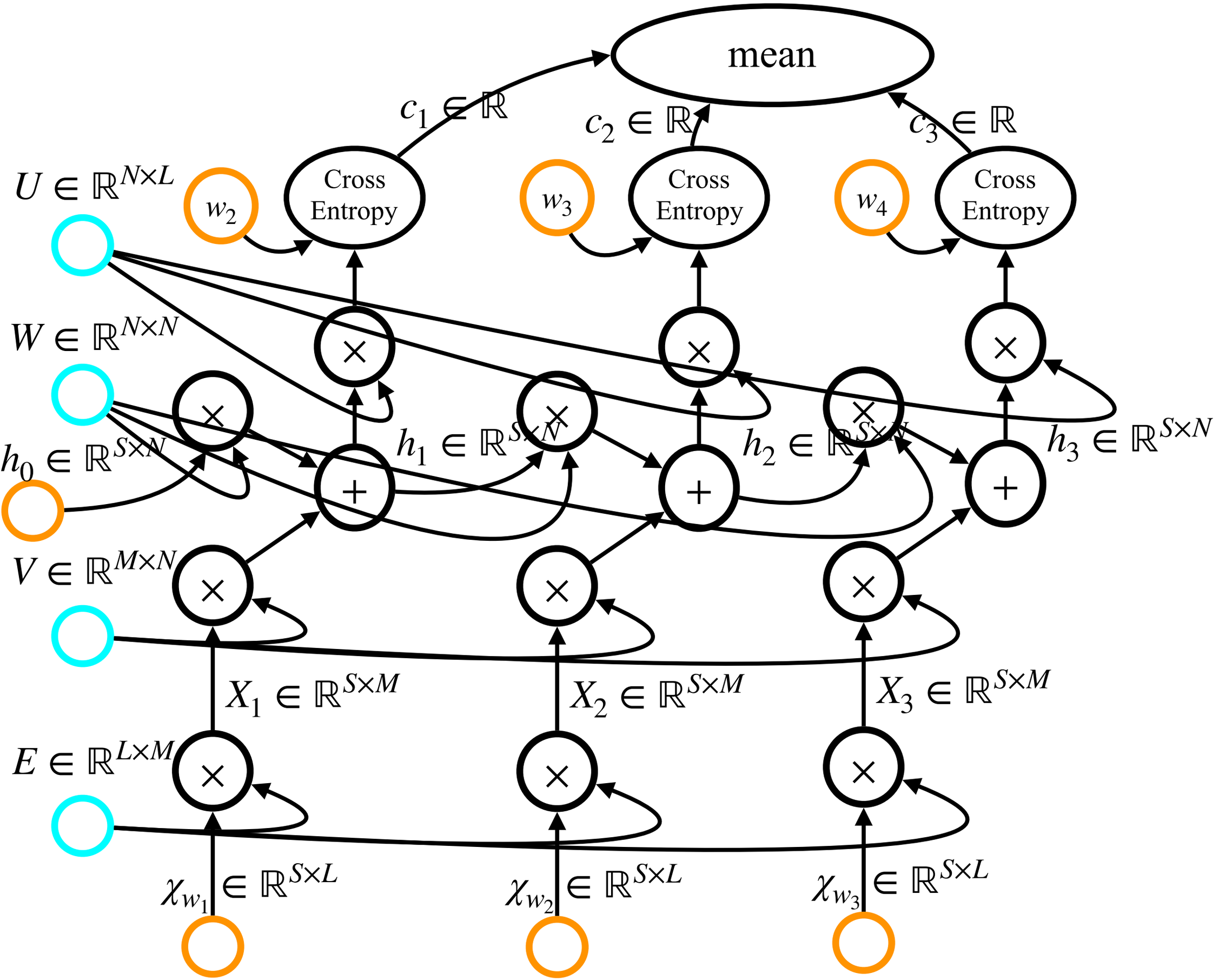
1. Формалности за курса (5 мин)
2. Марковски невронен езиков (10 мин)
3. Рекурентен езиков модел (10 мин)
4. Пропагиране напред при рекурентна невронна мрежа (15 мин)
5. Обучение на рекурентна невронна мрежа (20 мин)
6. **Пропагиране назад при рекурентна невронна мрежа (15 мин)**
7. Приложения на езиковите модели (15 мин)

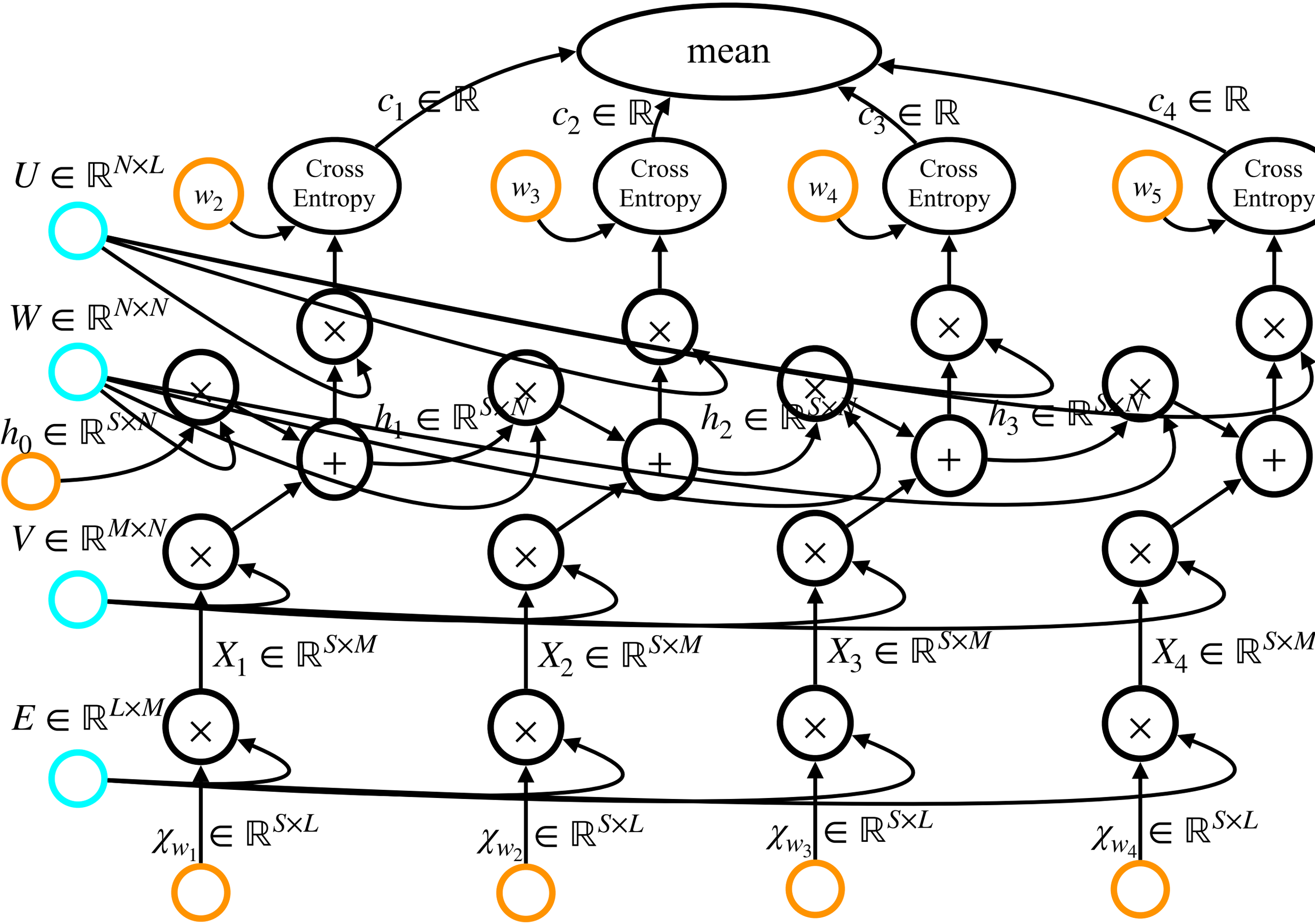
Пресмятане на градиентите с Backpropagation

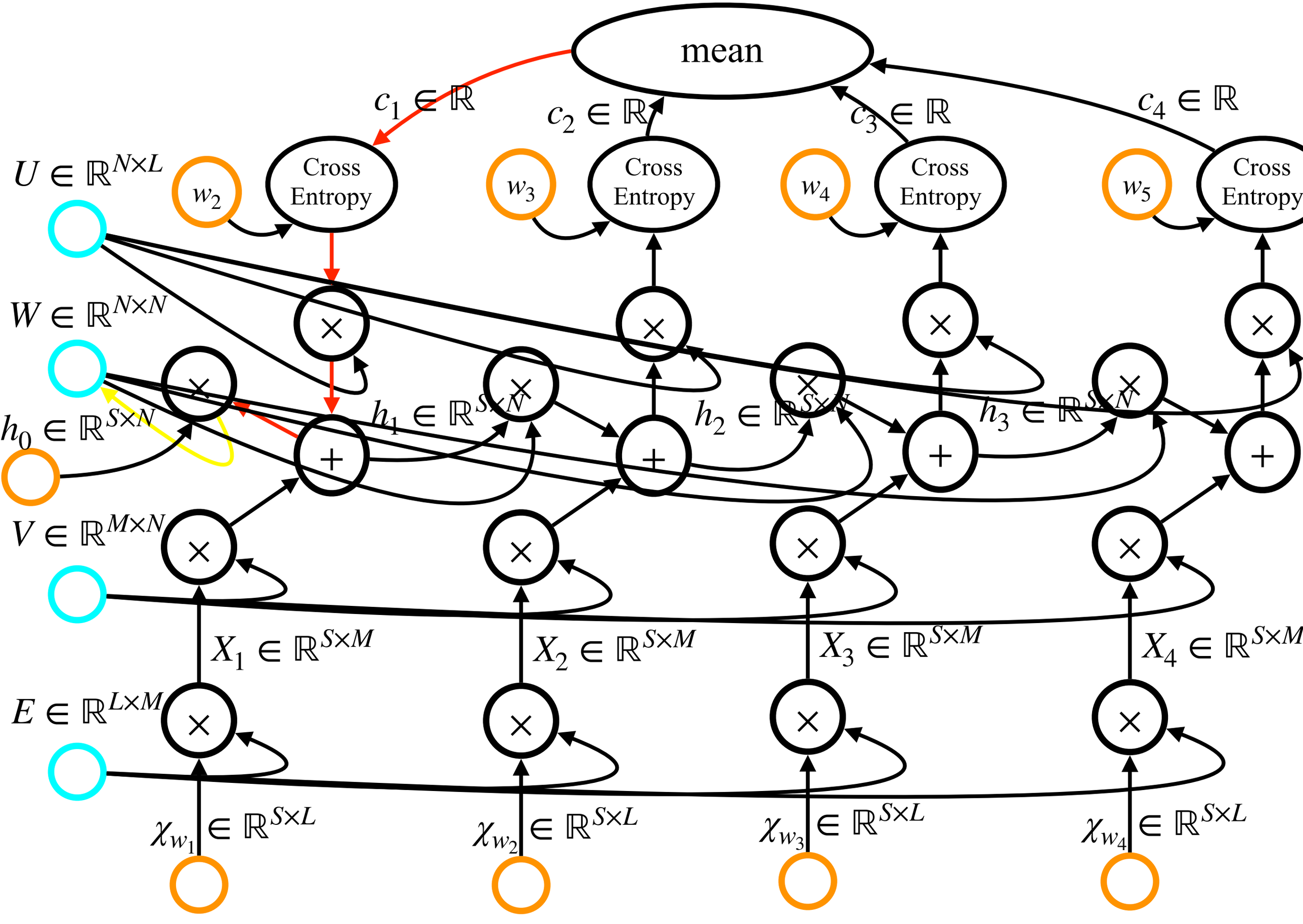
- Трябва да съставим изчислителен граф
- Изчислителния граф следва да бъде разпънат за всички думи в дадено изречение
- Имайки разпънатия изчислителен граф изчисляването на градиентите става автоматично с метода Backpropagation
- Тази техника се нарича “Backpropagation in Time”

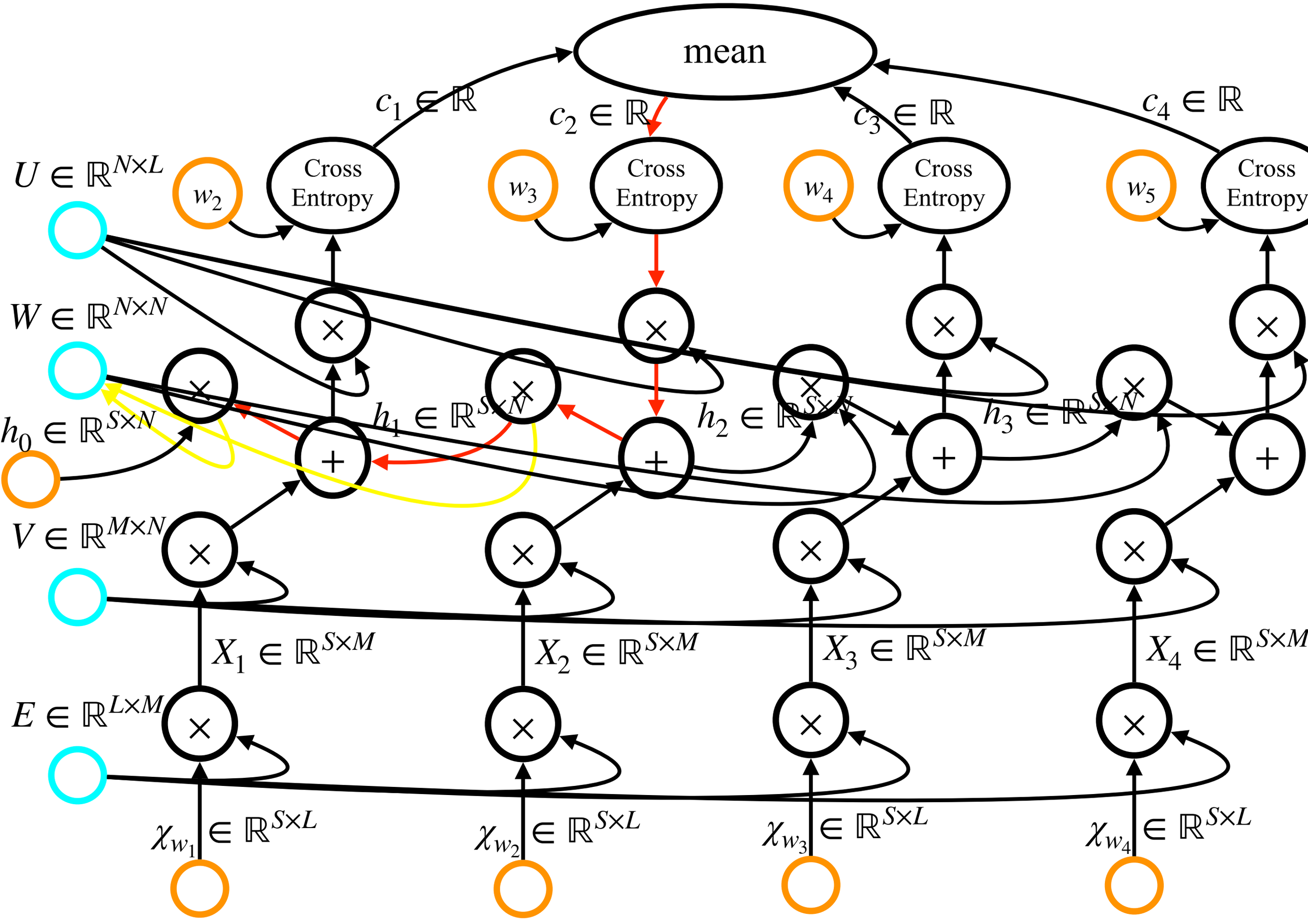

$c_1 \in \mathbb{R}$
Cross Entropy
w_2
$U \in \mathbb{R}^{N \times L}$
$W \in \mathbb{R}^{N \times N}$
$h_0 \in \mathbb{R}^{S \times N}$
$V \in \mathbb{R}^{M \times N}$
$X_1 \in \mathbb{R}^{S \times M}$
$E \in \mathbb{R}^{L \times M}$
$\chi_{w_1} \in \mathbb{R}^{S \times L}$

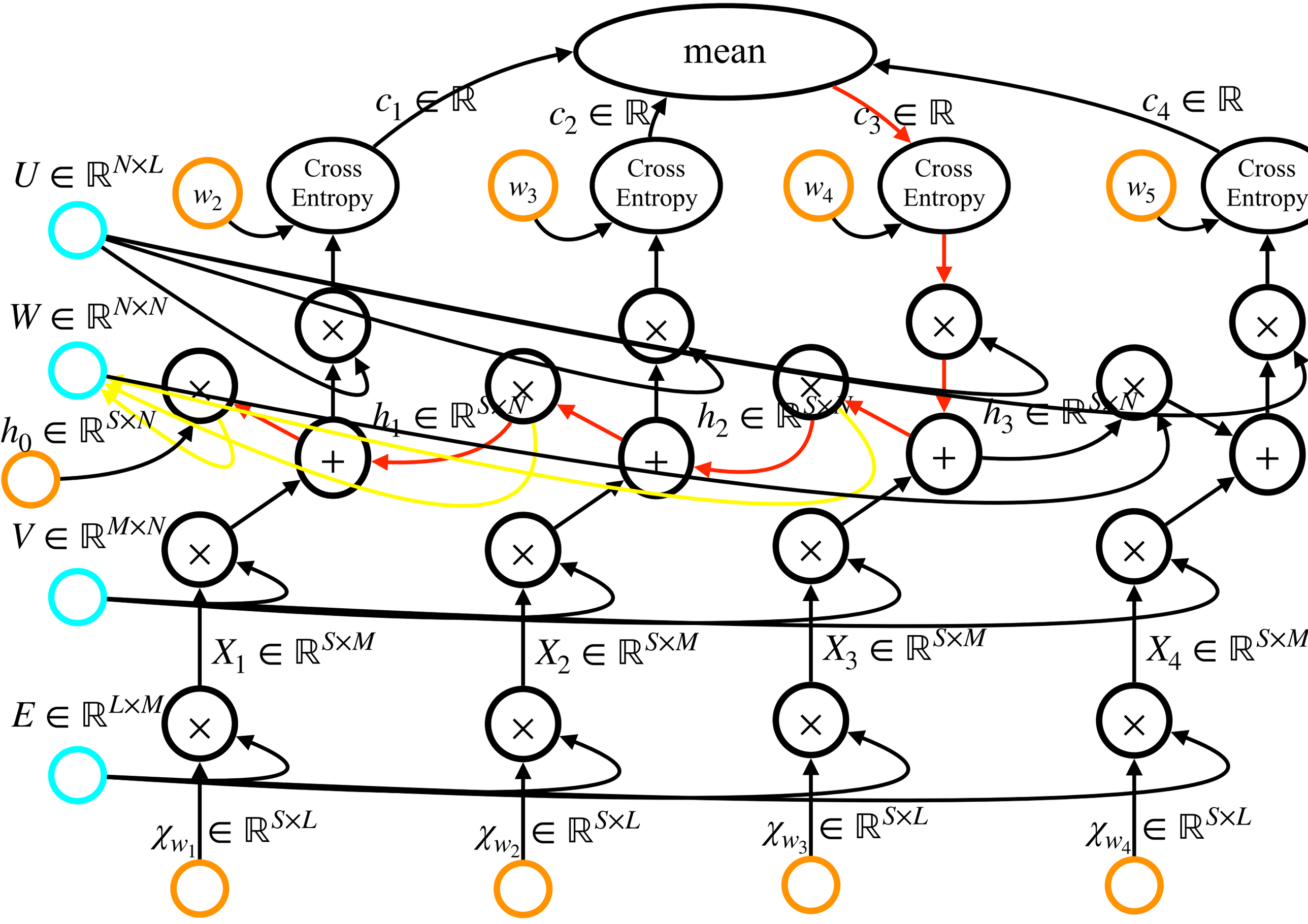

mean
$c_1 \in \mathbb{R}$
$c_2 \in \mathbb{R}$
Cross Entropy
Cross Entropy
w_2
w_3
$U \in \mathbb{R}^{N \times L}$
$W \in \mathbb{R}^{N \times N}$
$V \in \mathbb{R}^{M \times N}$
$E \in \mathbb{R}^{L \times M}$
$h_0 \in \mathbb{R}^{S \times N}$
$h_1 \in \mathbb{R}^{S \times N}$
$h_2 \in \mathbb{R}^{S \times N}$
$X_1 \in \mathbb{R}^{S \times M}$
$X_2 \in \mathbb{R}^{S \times M}$
$\chi_{w_1} \in \mathbb{R}^{S \times L}$
$\chi_{w_2} \in \mathbb{R}^{S \times L}$

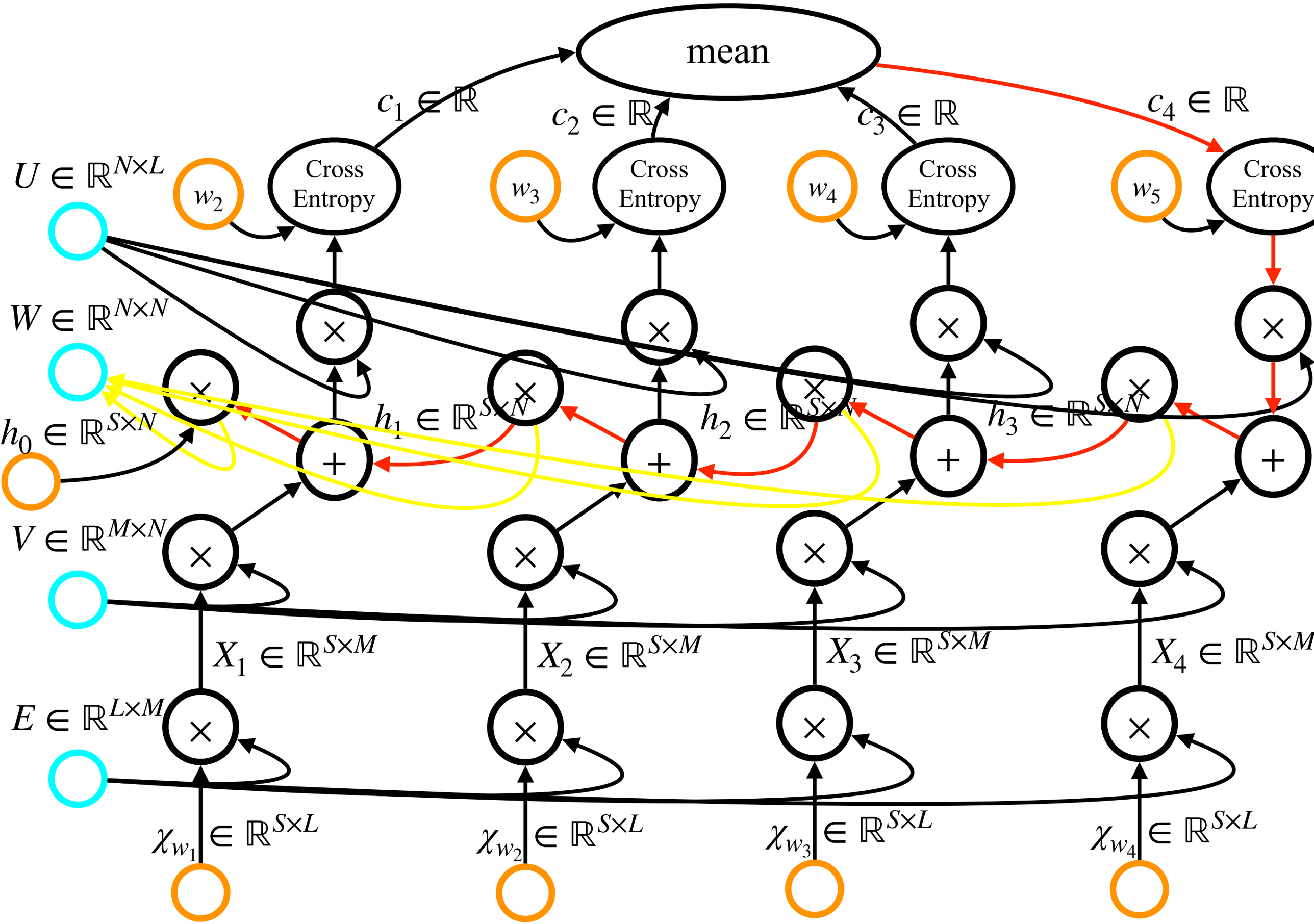

mean
$c_1 \in \mathbb{R}$
$c_2 \in \mathbb{R}$
$c_3 \in \mathbb{R}$
Cross Entropy
Cross Entropy
Cross Entropy
w_2
w_3
w_4
$U \in \mathbb{R}^{N\times L}$
$W \in \mathbb{R}^{N\times N}$
$V \in \mathbb{R}^{M\times N}$
$E \in \mathbb{R}^{L\times M}$
$h_0 \in \mathbb{R}^{S\times N}$
$h_1 \in \mathbb{R}^{S\times N}$
$h_2 \in \mathbb{R}^{S\times N}$
$h_3 \in \mathbb{R}^{S\times N}$
$X_1 \in \mathbb{R}^{S\times M}$
$X_2 \in \mathbb{R}^{S\times M}$
$X_3 \in \mathbb{R}^{S\times M}$
$\chi_{w_1} \in \mathbb{R}^{S\times L}$
$\chi_{w_2} \in \mathbb{R}^{S\times L}$
$\chi_{w_3} \in \mathbb{R}^{S\times L}$

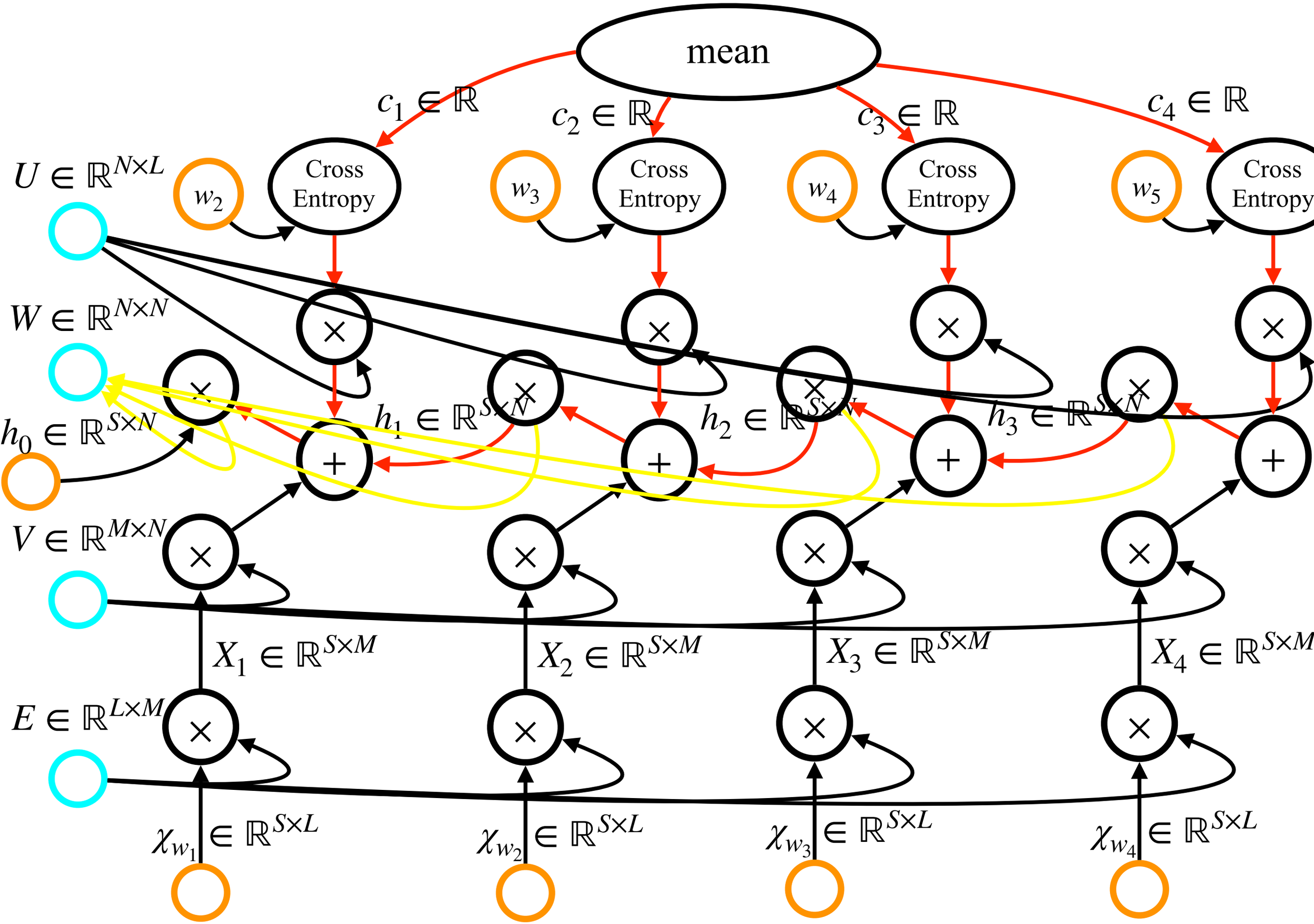

mean
$c_1 \in \mathbb{R}$
$c_2 \in \mathbb{R}$
$c_3 \in \mathbb{R}$
$c_4 \in \mathbb{R}$
Cross Entropy
Cross Entropy
Cross Entropy
Cross Entropy
w_2
w_3
w_4
w_5
$U \in \mathbb{R}^{N \times L}$
$W \in \mathbb{R}^{N \times N}$
$V \in \mathbb{R}^{M \times N}$
$E \in \mathbb{R}^{L \times M}$
$h_0 \in \mathbb{R}^{S \times N}$
$h_1 \in \mathbb{R}^{S \times N}$
$h_2 \in \mathbb{R}^{S \times N}$
$h_3 \in \mathbb{R}^{S \times N}$
$X_1 \in \mathbb{R}^{S \times M}$
$X_2 \in \mathbb{R}^{S \times M}$
$X_3 \in \mathbb{R}^{S \times M}$
$X_4 \in \mathbb{R}^{S \times M}$
$\chi_{w_1} \in \mathbb{R}^{S \times L}$
$\chi_{w_2} \in \mathbb{R}^{S \times L}$
$\chi_{w_3} \in \mathbb{R}^{S \times L}$
$\chi_{w_4} \in \mathbb{R}^{S \times L}$


mean
$c_1 \in \mathbb{R}$
$c_2 \in \mathbb{R}$
$c_3 \in \mathbb{R}$
$c_4 \in \mathbb{R}$
Cross Entropy
Cross Entropy
Cross Entropy
Cross Entropy
w_2
w_3
w_4
w_5
$U \in \mathbb{R}^{N \times L}$
$W \in \mathbb{R}^{N \times N}$
$V \in \mathbb{R}^{M \times N}$
$E \in \mathbb{R}^{L \times M}$
$h_0 \in \mathbb{R}^{S \times N}$
$h_1 \in \mathbb{R}^{S \times N}$
$h_2 \in \mathbb{R}^{S \times N}$
$h_3 \in \mathbb{R}^{S \times N}$
$X_1 \in \mathbb{R}^{S \times M}$
$X_2 \in \mathbb{R}^{S \times M}$
$X_3 \in \mathbb{R}^{S \times M}$
$X_4 \in \mathbb{R}^{S \times M}$
$\chi_{w_1} \in \mathbb{R}^{S \times L}$
$\chi_{w_2} \in \mathbb{R}^{S \times L}$
$\chi_{w_3} \in \mathbb{R}^{S \times L}$
$\chi_{w_4} \in \mathbb{R}^{S \times L}$


mean
$c_1 \in \mathbb{R}$
$c_2 \in \mathbb{R}$
$c_3 \in \mathbb{R}$
$c_4 \in \mathbb{R}$
Cross Entropy
Cross Entropy
Cross Entropy
Cross Entropy
w_2
w_3
w_4
w_5
$U \in \mathbb{R}^{N \times L}$
$W \in \mathbb{R}^{N \times N}$
$h_0 \in \mathbb{R}^{S \times N}$
$h_1 \in \mathbb{R}^{S \times N}$
$h_2 \in \mathbb{R}^{S \times N}$
$h_3 \in \mathbb{R}^{S \times N}$
$V \in \mathbb{R}^{M \times N}$
$X_1 \in \mathbb{R}^{S \times M}$
$X_2 \in \mathbb{R}^{S \times M}$
$X_3 \in \mathbb{R}^{S \times M}$
$X_4 \in \mathbb{R}^{S \times M}$
$E \in \mathbb{R}^{L \times M}$
$\chi_{w_1} \in \mathbb{R}^{S \times L}$
$\chi_{w_2} \in \mathbb{R}^{S \times L}$
$\chi_{w_3} \in \mathbb{R}^{S \times L}$
$\chi_{w_4} \in \mathbb{R}^{S \times L}$


mean
$c_1 \in \mathbb{R}$
$c_2 \in \mathbb{R}$
$c_3 \in \mathbb{R}$
$c_4 \in \mathbb{R}$
Cross Entropy
Cross Entropy
Cross Entropy
Cross Entropy
w_2
w_3
w_4
w_5
$U \in \mathbb{R}^{N \times L}$
$W \in \mathbb{R}^{N \times N}$
$V \in \mathbb{R}^{M \times N}$
$E \in \mathbb{R}^{L \times M}$
$h_0 \in \mathbb{R}^{S \times N}$
$h_1 \in \mathbb{R}^{S \times N}$
$h_2 \in \mathbb{R}^{S \times N}$
$h_3 \in \mathbb{R}^{S \times N}$
$X_1 \in \mathbb{R}^{S \times M}$
$X_2 \in \mathbb{R}^{S \times M}$
$X_3 \in \mathbb{R}^{S \times M}$
$X_4 \in \mathbb{R}^{S \times M}$
$\chi_{w_1} \in \mathbb{R}^{S \times L}$
$\chi_{w_2} \in \mathbb{R}^{S \times L}$
$\chi_{w_3} \in \mathbb{R}^{S \times L}$
$\chi_{w_4} \in \mathbb{R}^{S \times L}$


mean
$c_1 \in \mathbb{R}$
$c_2 \in \mathbb{R}$
$c_3 \in \mathbb{R}$
$c_4 \in \mathbb{R}$
Cross Entropy
Cross Entropy
Cross Entropy
Cross Entropy
w_2
w_3
w_4
w_5
$U \in \mathbb{R}^{N \times L}$
$W \in \mathbb{R}^{N \times N}$
$h_0 \in \mathbb{R}^{S \times N}$
$h_1 \in \mathbb{R}^{S \times N}$
$h_2 \in \mathbb{R}^{S \times N}$
$h_3 \in \mathbb{R}^{S \times N}$
$V \in \mathbb{R}^{M \times N}$
$X_1 \in \mathbb{R}^{S \times M}$
$X_2 \in \mathbb{R}^{S \times M}$
$X_3 \in \mathbb{R}^{S \times M}$
$X_4 \in \mathbb{R}^{S \times M}$
$E \in \mathbb{R}^{L \times M}$
$\chi_{w_1} \in \mathbb{R}^{S \times L}$
$\chi_{w_2} \in \mathbb{R}^{S \times L}$
$\chi_{w_3} \in \mathbb{R}^{S \times L}$
$\chi_{w_4} \in \mathbb{R}^{S \times L}$


mean
$c_1 \in \mathbb{R}$
$c_2 \in \mathbb{R}$
$c_3 \in \mathbb{R}$
$c_4 \in \mathbb{R}$
$U \in \mathbb{R}^{N\times L}$
$W \in \mathbb{R}^{N\times N}$
$V \in \mathbb{R}^{M\times N}$
$E \in \mathbb{R}^{L\times M}$
w_2
w_3
w_4
w_5
Cross Entropy
Cross Entropy
Cross Entropy
Cross Entropy
$h_0 \in \mathbb{R}^{S\times N}$
$h_1 \in \mathbb{R}^{S\times N}$
$h_2 \in \mathbb{R}^{S\times N}$
$h_3 \in \mathbb{R}^{S\times N}$
$X_1 \in \mathbb{R}^{S\times M}$
$X_2 \in \mathbb{R}^{S\times M}$
$X_3 \in \mathbb{R}^{S\times M}$
$X_4 \in \mathbb{R}^{S\times M}$
$\chi_{w_1} \in \mathbb{R}^{S\times L}$
$\chi_{w_2} \in \mathbb{R}^{S\times L}$
$\chi_{w_3} \in \mathbb{R}^{S\times L}$
$\chi_{w_4} \in \mathbb{R}^{S\times L}$

Партидно изчисляване на градиентите

- **Проблем**: Изреченията са с различна дължина.
- Решения:
  - Изравняваме всички изречения в дадена партида, като ги допълваме с нов специален символ, при който не пропагираме градиента назад.
  - Сортираме изреченията по дължина и ги групираме в партиди с еднаква дължина.
  - Конкатенираме изреченията и след това ги нарязваме на фиксирана дължина.

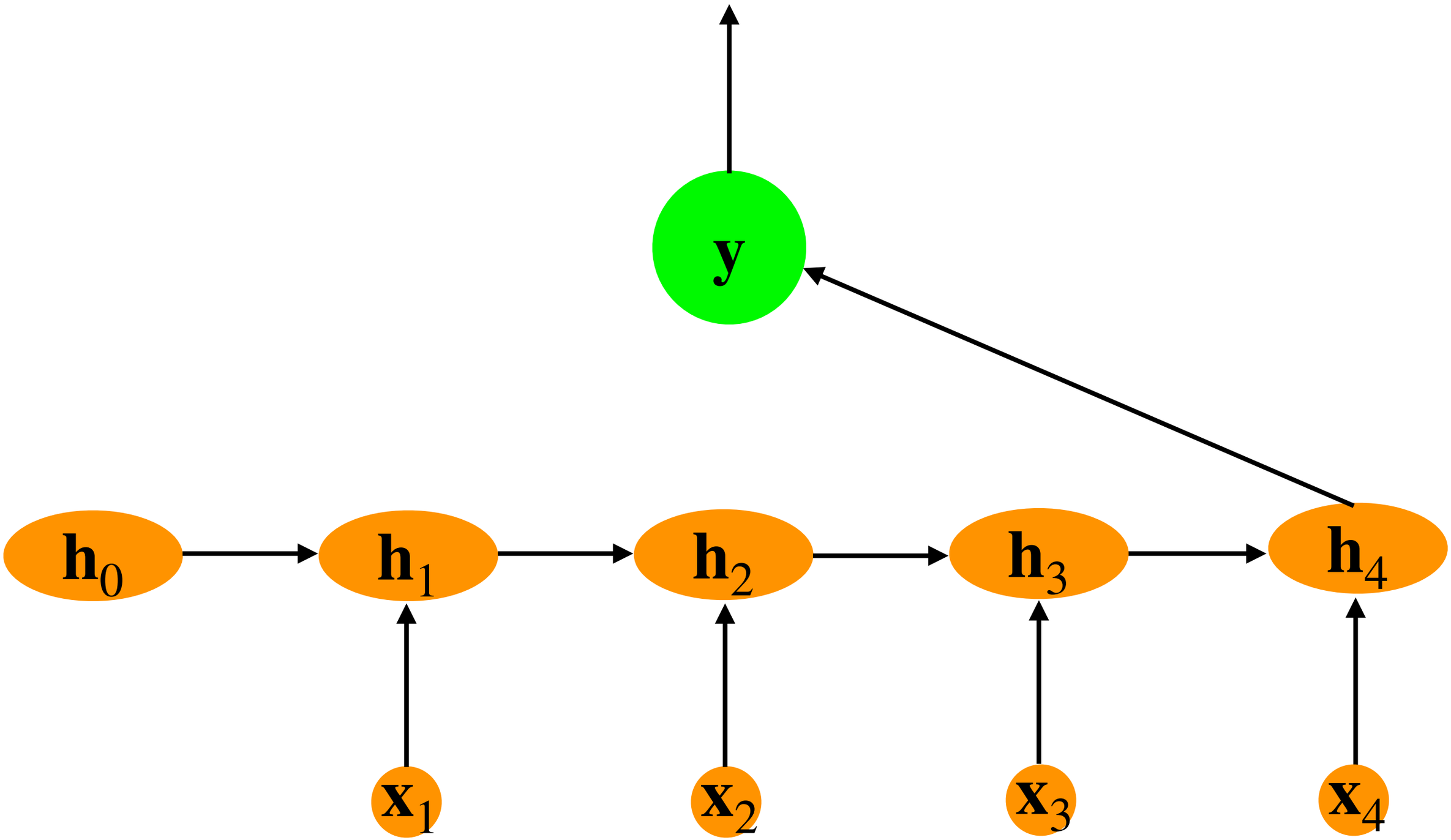
План на лекцията

1. Формалности за курса (5 мин)
2. Марковски невронен езиков (10 мин)
3. Рекурентен езиков модел (10 мин)
4. Пропагиране напред при рекурентна невронна мрежа (15 мин)
5. Обучение на рекурентна невронна мрежа (20 мин)
6. Пропагиране назад при рекурентна невронна мрежа (15 мин)
7. **Приложения на езиковите модели (15 мин)**

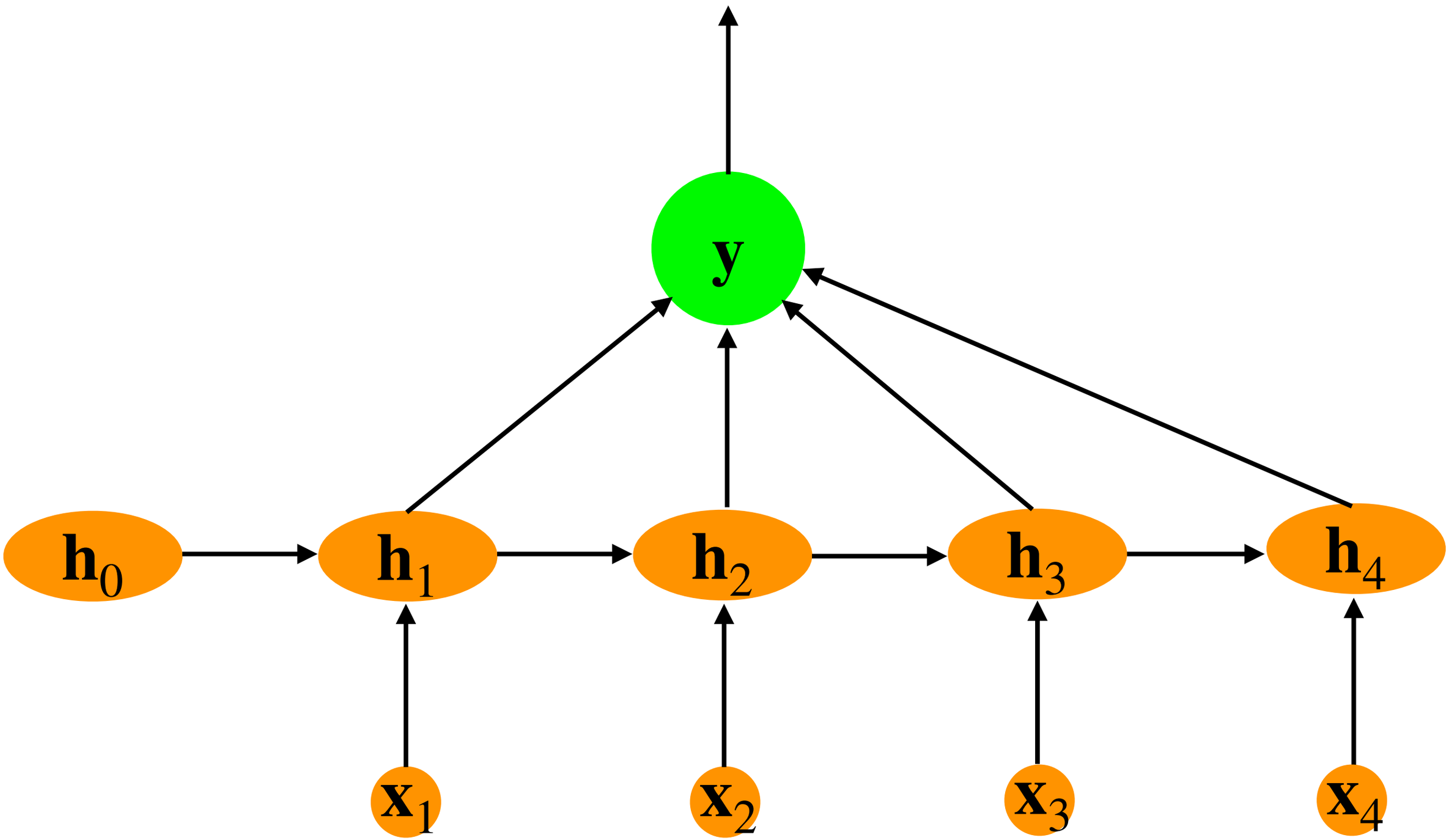
Приложения на езиковите модели

- Предсказване по време на изписване на заявка
- Корекция на текст
- Разпознаване на авторство
- Класификация на документи
- Резюмиране на документи
- Машинен превод
- Разпознаване на реч
- Отговаряне на въпроси
- и много други …

Приложение на езиков модел за класификация



Приложение на езиков модел за класификация



Приложения при които се налага да се генерира текст

- В някои приложения е необходимо да се генерира текст.
  - Генериране на резюме за документ
  - Генериране на превод за дадено изречение.
  - Генериране на текст съответстващ на речта представена като аудио сигнал
- Бихме могли да използваме езиков модел за генерирането на съответен текст.
- Езиковият модел следва да отразява желания семантичен контекст.

Генериране на текст с езиков модел

- Започваме с текст съдържащ единствено символа за начало
- На всяка стъпка избираме следващата дума случайно, като използваме разпределението за следващата дума при контекста досега, получено от езиковия модел.
- Когато изберем символа за край спираме процедурата.

Генериран текст	**Вероятностно разпределение**
<START>	
<START> днес	днес ⬅ Pr[X | <START>]
<START> днес е	е ⬅ Pr[X | <START> днес]
<START> днес е коледа	коледа ⬅ Pr[X | <START> днес е]
<START> днес е коледа <STOP>	<STOP> ⬅ Pr[X | <START> днес е коледа]

Базов алгоритъм за генериране на текст чрез езиков модел

```
generateText(P)
    t <- “<START>”
    w <- [t]
    while not t = “<END>” do
        t <- sample(P(t|w))
        w <- concatenate(w,[t])
return w
```