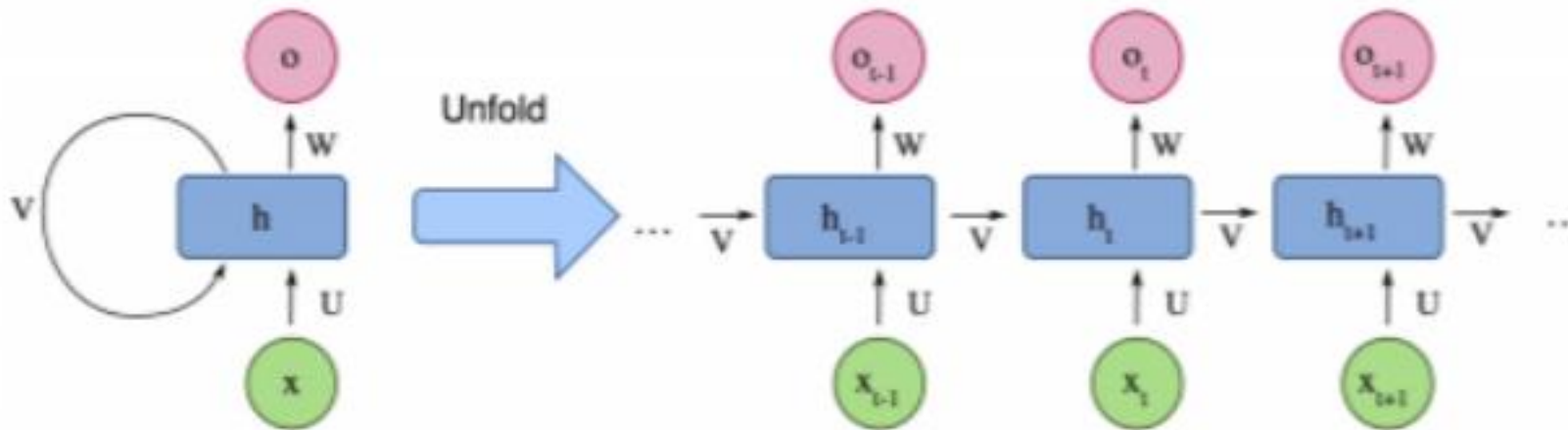


Рекуррентные нейронные сети (RNN)

Что такое RNN

- Рекуррентная нейронная сеть — это сеть, обрабатывающая последовательность данных
- Главная особенность — сеть запоминает контекст из прошлого



Зачем нужны RNN

Применения RNN:

- Обработка текста (перевод, генерация, анализ)
- Предсказание временных рядов
- Распознавание речи и аудио
- Анализ последовательностей в биоинформатике

Как работает RNN

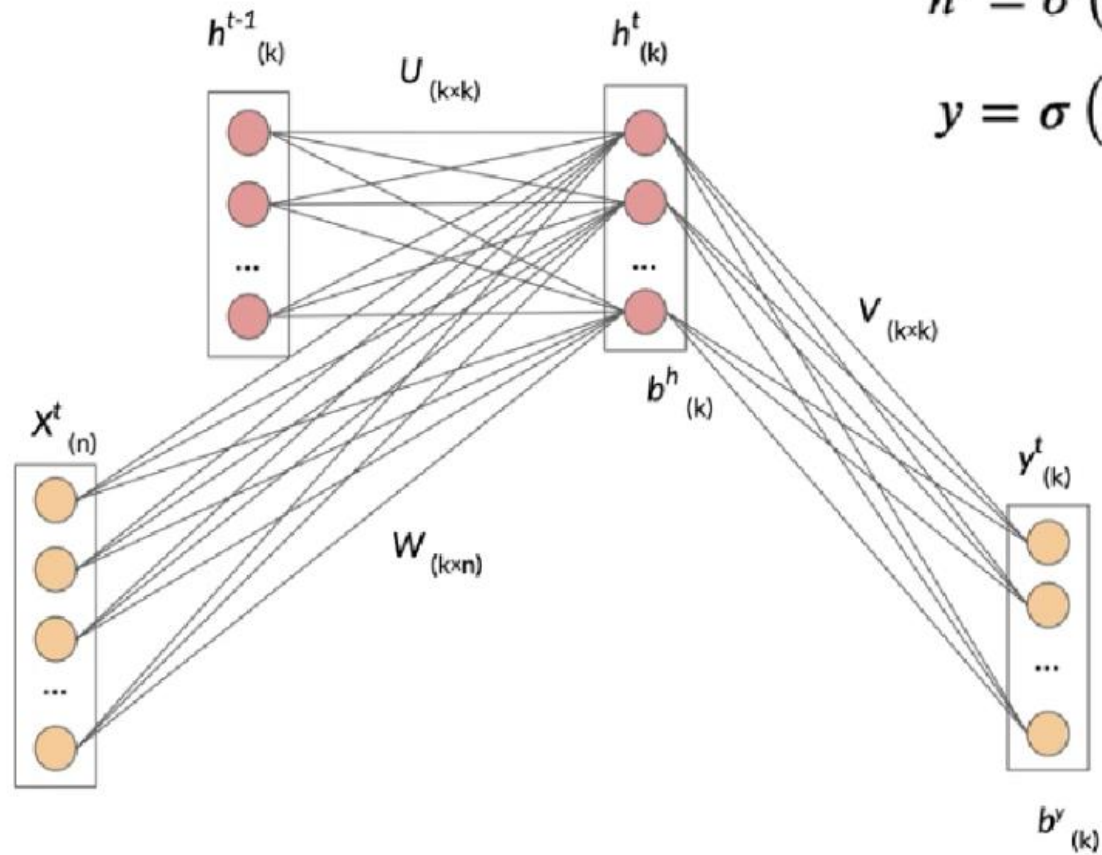


Обновление вектора
скрытого состояния и
вычисление выхода слоя:

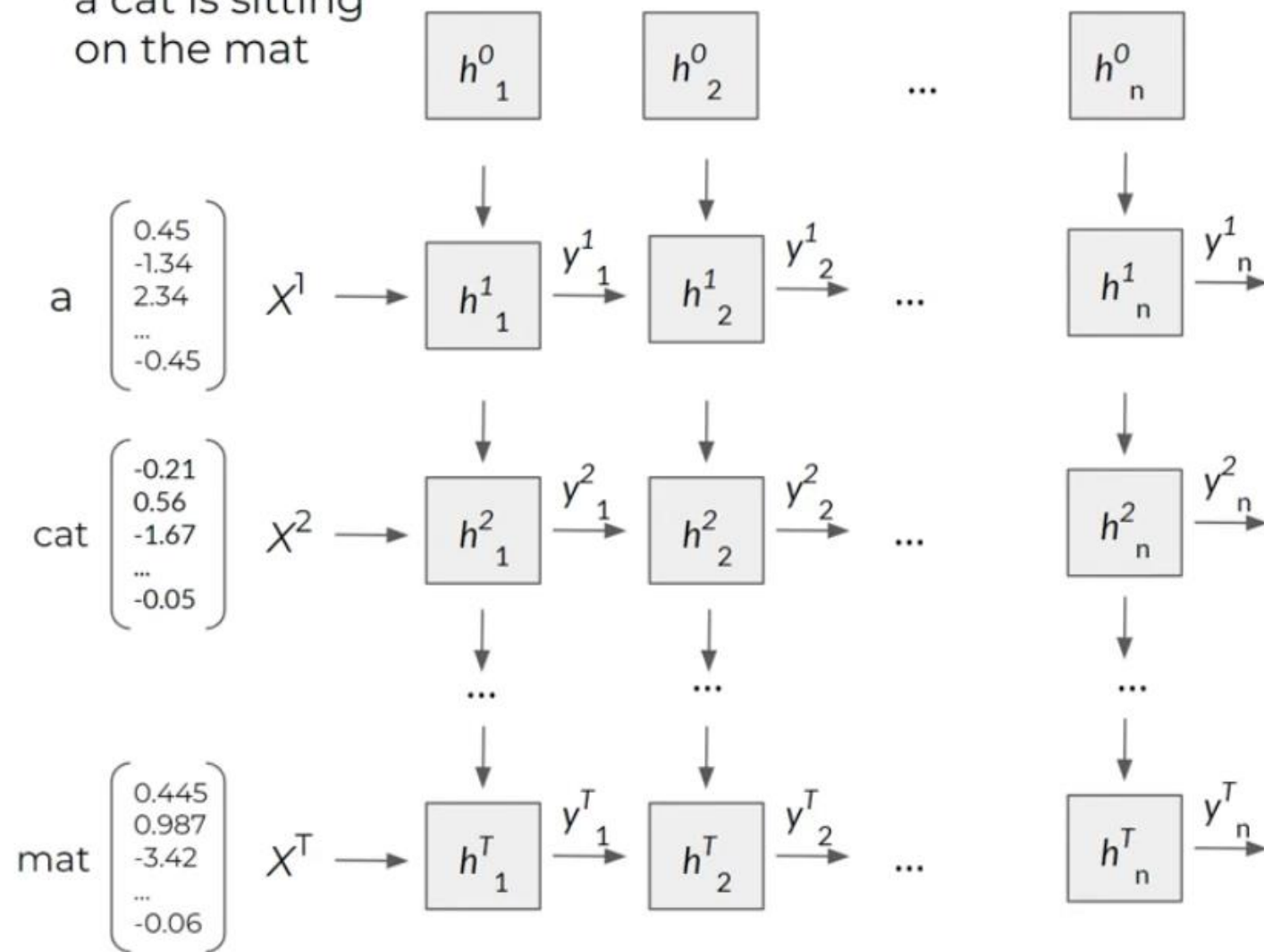
сначала вычисляется h^t , а затем y^t

$$h^t = \sigma(WX^t + Uh^{t-1} + b_h)$$

$$y = \sigma(Vh^t + b_y)$$



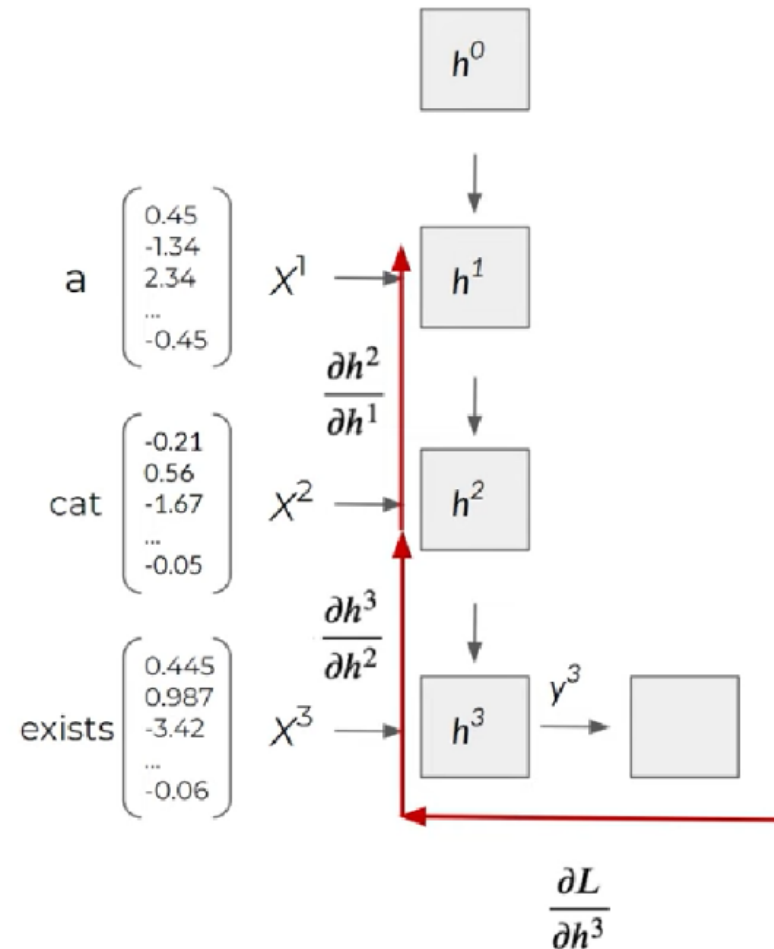
a cat is sitting
on the mat



Разные типы входов и выходов

- One-to-One: обычная сеть
- One-to-Many: генерация текста
- Many-to-One: анализ текста, прогноз
- Many-to-Many: перевод, теггинг

Обратное распространение ошибки через время



$$h^t = \sigma(WX^t + Uh^{t-1} + b_h)$$

$$\begin{aligned} \frac{dL}{dW} &= \frac{\partial L}{\partial h^3} \frac{dh^3}{dW} = \\ &= \frac{\partial L}{\partial h^3} \frac{\partial h^3}{\partial W} + \frac{\partial L}{\partial h^3} \frac{\partial h^3}{\partial h^2} \frac{dh^2}{dW} = \\ &= \frac{\partial L}{\partial h^3} \frac{\partial h^3}{\partial W} + \frac{\partial L}{\partial h^3} \frac{\partial h^3}{\partial h^2} \frac{\partial h^2}{\partial W} + \\ &+ \frac{\partial L}{\partial h^3} \frac{\partial h^3}{\partial h^2} \frac{\partial h^2}{\partial h^1} \frac{\partial h^1}{\partial W} \end{aligned}$$

Проблемы RNN

- Исчезающие градиенты
- Взрывающиеся градиенты
- Сложность в запоминании долгосрочных зависимостей

Модификации RNN

- LSTM (Long Short-Term Memory)
- GRU (Gated Recurrent Unit)
- Bidirectional RNN
- Stacked RNN

