Language Modeling

- Goal: Predict the next word
- Statistical methods
- Recent method
  - Recurrent Neural Networks
- Examples and figures copied from
  - Christopher Manning: Language Models and Recurrent Neural Networks

Examples, Applications

- Predictive typing
- Google: Fill the rest of the query
- Spelling/grammar correction
- ChatBot
- Question answering
- Document summarization
- Authorship identification
- Machine translation

- Speech recognition
- Handwriting recognition

Statistically methods

- Probability based model
- Giving a sequence of words, calculate the probability of the next word.

- A sliding window
  - N gram model

n-gram Language Models: Example

the students opened their _____

- **Definition:** A *n*-gram is a chunk of *n* consecutive words.
  - unigrams: “the”, “students”, “opened”, ”their”
  - bigrams: “the students”, “students opened”, “opened their”
  - trigrams: “the students opened”, “students opened their”
  - 4-grams: “the students opened their”
- **Idea:** Collect statistics about how frequent different n-grams are and use these to predict next word.
- How to get the probability of n grams: counting in a big corpus

$$P(\boldsymbol{w}|\text{students opened their}) = \frac{\text{count}(\text{students opened their } \boldsymbol{w})}{\text{count}(\text{students opened their})}$$

Generating text with a n-gram Language Model

You can also use a Language Model to generate text

today the price of gold per ton , while production of shoe lasts and shoe industry , the bank intervened just after it considered and rejected an imf demand to rebuild depleted european stocks , sept 30 end primary 76 cts a share .

Surprisingly grammatical!

…but **incoherent.** We need to consider more than three words at a time if we want to model language well.

But increasing *n* worsens sparsity problem, and increases model size…

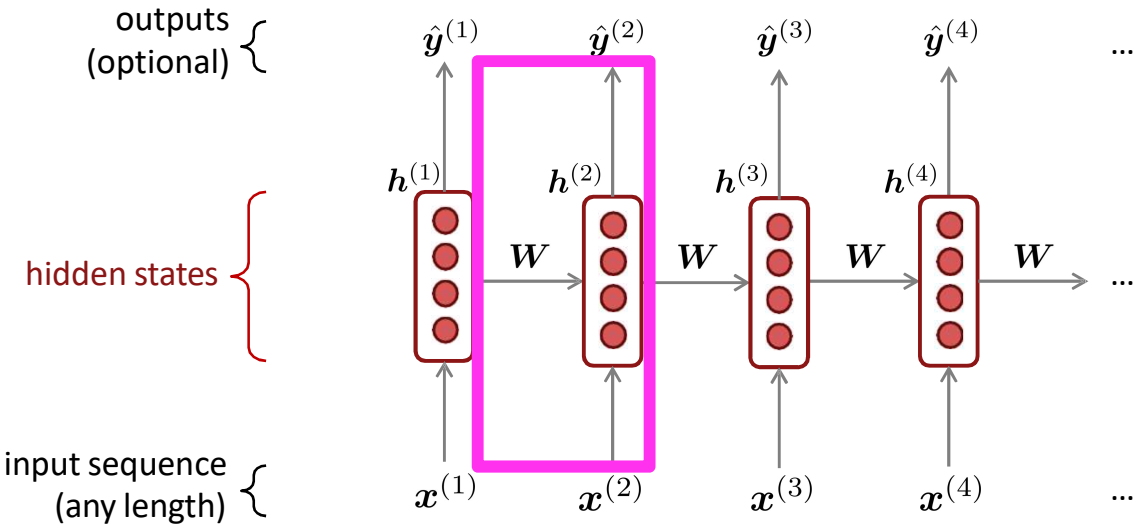
Challenges:

- If the window size is small: not reliable
- If the window size is big, not enough samples, sparsity problem.

- Fixed sized window is a problem

Recurrent Neural Networks (RNN)

Core idea: Apply the same weights *W repeatedly*



A Simple RNN Language Model

output distribution

$$\hat{\boldsymbol{y}}^{(t)} = \text{softmax}\left(\boldsymbol{U}\boldsymbol{h}^{(t)} + \boldsymbol{b}_2\right) \in \mathbb{R}^{|V|}$$

hidden states

$$\boldsymbol{h}^{(t)} = \sigma\left(\boldsymbol{W}_h\boldsymbol{h}^{(t-1)} + \boldsymbol{W}_e\boldsymbol{e}^{(t)} + \boldsymbol{b}_1\right)$$

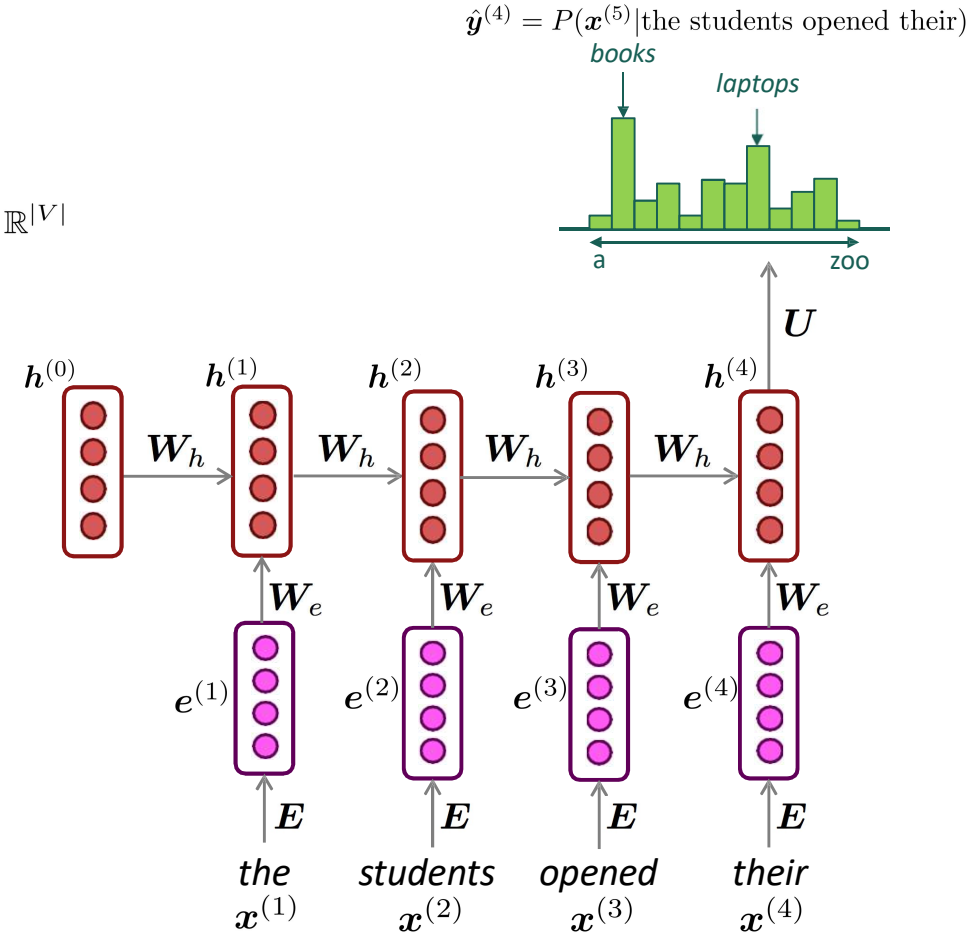
$\boldsymbol{h}^{(0)}$ is the initial hidden state

word embeddings

$$\boldsymbol{e}^{(t)} = \boldsymbol{E}\boldsymbol{x}^{(t)}$$

words / one-hot vectors

$$\boldsymbol{x}^{(t)} \in \mathbb{R}^{|V|}$$



Note: *this input sequence could be much longer now!*

Training an RNN Language Model

- Get a big corpus of text which is a sequence of words $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(T)}$
- Feed into RNN-LM; compute output distribution $\hat{\boldsymbol{y}}^{(t)}$ for *every step t*.
  - i.e. predict probability dist of *every word*, given words so far

- Loss function on step *t* is cross-entropy between predicted probability distribution $\hat{\boldsymbol{y}}^{(t)}$, and the true next word $\boldsymbol{y}^{(t)}$ (one-hot for $\boldsymbol{x}^{(t+1)}$):

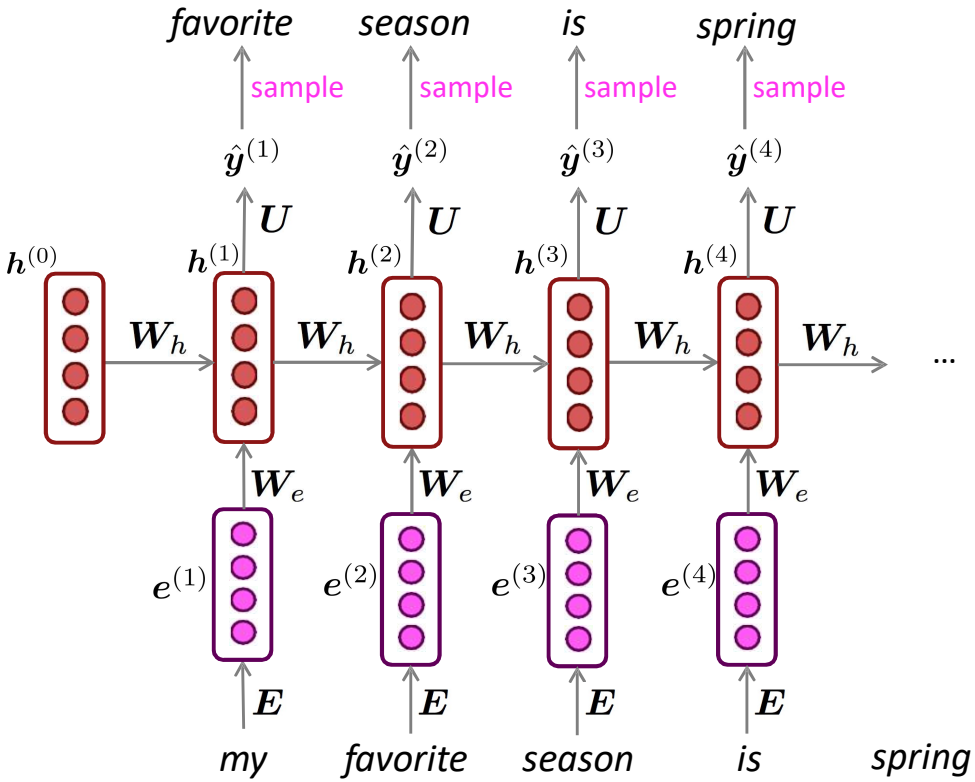
$$J^{(t)}(\theta) = CE(\boldsymbol{y}^{(t)}, \hat{\boldsymbol{y}}^{(t)}) = -\sum_{w \in V} \boldsymbol{y}_w^{(t)} \log \hat{\boldsymbol{y}}_w^{(t)} = -\log \hat{\boldsymbol{y}}_{\boldsymbol{x}_{t+1}}^{(t)}$$

- Average this to get overall loss for entire training set:

$$J(\theta) = \frac{1}{T} \sum_{t=1}^{T} J^{(t)}(\theta) = \frac{1}{T} \sum_{t=1}^{T} -\log \hat{\boldsymbol{y}}_{\boldsymbol{x}_{t+1}}^{(t)}$$

Generating text with a RNN Language Model

Just like a n-gram Language Model, you can use a RNN Language Model to generate text by repeated sampling. Sampled output becomes next step's input.



Generating text with an RNN Language Model

Let's have some fun!

- You can train an RNN-LM on any kind of text, then generate text in that style.
- RNN-LM trained on Obama speeches:

The United States will step up to the cost of a new challenges of the American people that will share the fact that we created the problem. They were attacked and so that they have to say that all the task of the final days of war that I will not be able to get this done.

Source: https://medium.com/@samim/obama-rnn-machine-generated-political-speeches-c8abd18a2ea0

Generating text with an RNN Language Model

Let's have some fun!

- You can train an RNN-LM on any kind of text, then generate text in that style.
- RNN-LM trained on *Harry Potter*:

"Sorry," Harry shouted, panicking—"I'll leave those brooms in London, are they?"

"No idea," said Nearly Headless Nick, casting low close by Cedric, carrying the last bit of treacle Charms, from Harry's shoulder, and to answer him the common room perched upon it, four arms held a shining knob from when the spider hadn't felt it seemed. He reached the teams too.

Source: https://medium.com/deep-writing/harry-potter-written-by-artificial-intelligence-8a9431803da6

Generating text with an RNN Language Model

Let's have some fun!

- You can train an RNN-LM on any kind of text, then generate text in that style.
- RNN-LM trained on recipes:

```
      Title: CHOCOLATE RANCH BARBECUE
Categories: Game, Casseroles, Cookies, Cookies
      Yield: 6 Servings

      2 tb Parmesan cheese -- chopped
      1 c  Coconut milk
      3    Eggs, beaten

Place each pasta over layers of lumps. Shape mixture into the moderate oven and simmer
until firm. Serve hot in bodied fresh, mustard, orange and cheese.

                                                                                5bcc

Combine the cheese and salt together the dough in a large skillet; add the ingredients
and stir in the chocolate and pepper.
```

Generating text with a RNN Language Model

Let's have some fun!

- You can train a RNN-LM on any kind of text, then generate text in that style.
- RNN-LM trained on paint color names:

This is an example of a character-level RNN-LM (predicts what character comes next)

Source: http://aiweirdness.com/post/160776374467/new-paint-colors-invented-by-neural-network

Challenges/Limitations

- Vocabulary is good
- Style is good
- Grammar is reasonably good

- Context,
- idea,
- fluent,
- Coherence

Transformers

- Attention
- Self-attention
- Multi-head self attention

- Training data

- Large amount of training data

GPT

- Generative Pre-trained Transformer
- GPT, 2018,117 million parameters.
- GPT-2, 2019, 1.5 billion parameters.
- GPT-3, 2020, 175 billion parameters. As of early 2021, GPT-3 is the largest neural network ever produced.
- GPT-3 was trained on 570 GB plaintext from several data sets, including Common Crawl, WebText2, books1, books2, and Wikipedia.
- GPT-4, March 14, 2023
- GPT-4o, May 13, 2024

Language Models are Few-Shot Learners
https://arxiv.org/abs/2005.14165

ChatGPT

- ChatGPT is a chatbot
- Launched by OpenAI in November 2022.
- Built on top of OpenAI's GPT-3 family of large language models
- Fine-tuned with both supervised and reinforcement learning

- Your experience of using ChatGPT

GPT-3 Limitations

- **Pre-training.** ChatGPT trained up to 2021
- **Limited input size.** GPT-3 has a prompt limit of about 2,048 tokens.
- **Slow inference time. Expensive and inconvenient.**
- **Lack of explainability.**
- Lack of common sense
- Lack of semantic coherence
- Difficulty in natural language reasoning, reading comprehension tasks.
- Some risks
  - Accuracy
  - Bias
  - Mimicry

NLP applications in New Zealand

- ChatBot
  - UneeQ is an Auckland company who has developed a digital human platform
    - Southern Cross, Vodafone, Noel Leeming, ASB, Kiwibank
    - Ambit (another smart tech company)
  - Beca, Auckland-based engineering firm, ChatBot for Samoan language speakers
  - SpaceTime, IBM Watson, “Fine wine delivery”, advice and chat
- Aider, a mobile app, digital assistant for small business
- Government:
  - Data company Orbica, Visualization of public data, consultation process
  - NLP tool: Explain rates

Philosophy

- Does GPT understand what it is writing?
- Will machine understand natural language?
- Will machine develop consciousness?