LECTURER: Nghia Duong-Trung

DEEP LEARNING

MACHINE LEARNING
TOPIC OUTLINE

UNIT 4

Alternative Training Methods

STUDY GOALS

- After completing this unit, you will be able to ...

  … describe how the concept of attention is modeled in deep learning.

  … explain what bidirectional RNNs are and how they are used in language translation.

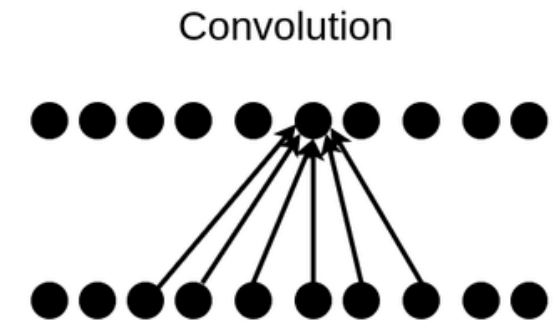
  … understand how feedback alignment addresses issues in the backpropagation algorithm.

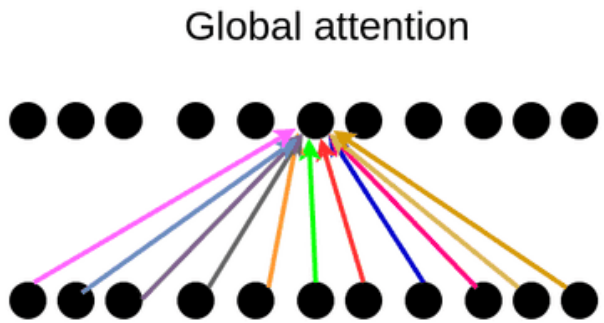
  … use synthetic gradients and decoupled neural interfaces to parallelize computations and speed up training.

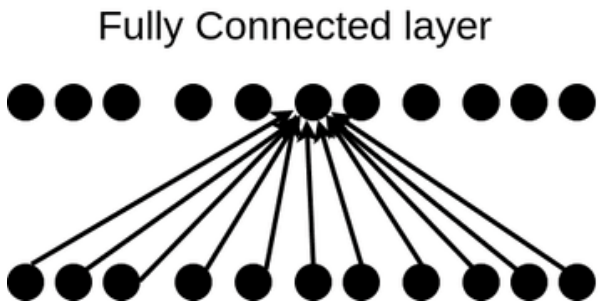
  … recognize the importance of transfer learning in real-world applications of deep learning.

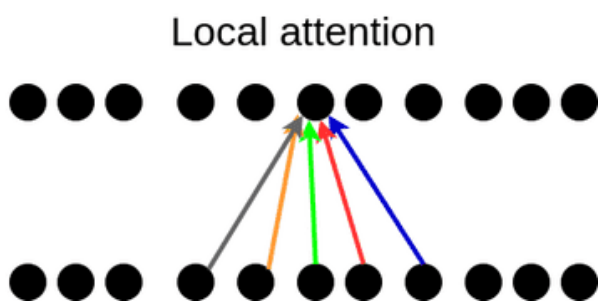
NEURAL MACHINE TRANSLATION AND ATTENTION

- Neural machine translation is the application of deep learning to construct language translation models.
- Attention mechanism focuses on relevant segments of text when translating a very long sentence.
  - allow the model to selectively focus on the parts of the input that are most important for making a prediction, and to ignore the less relevant parts







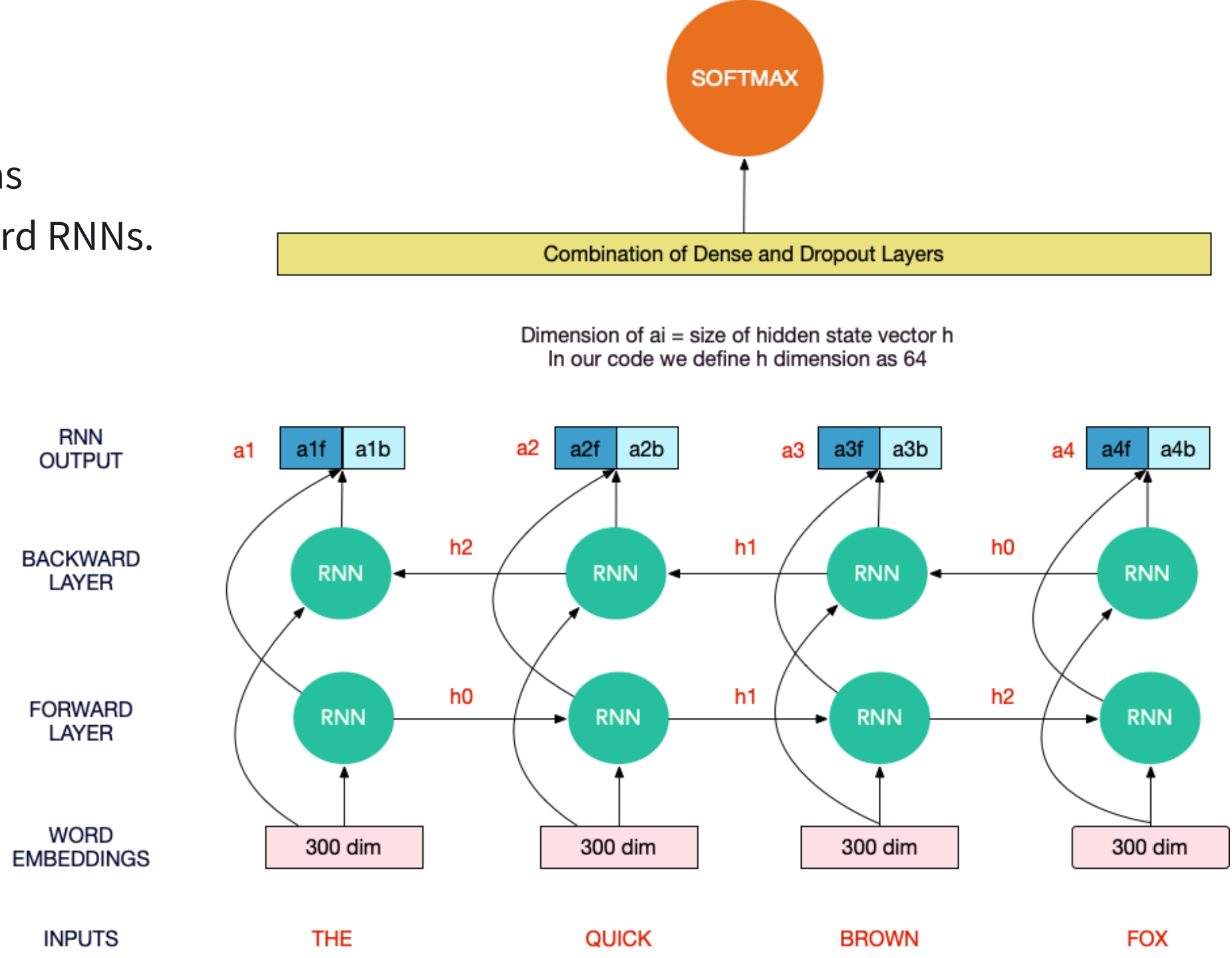


https://theaisummer.com/attention/

BIDIRECTIONAL RNNS

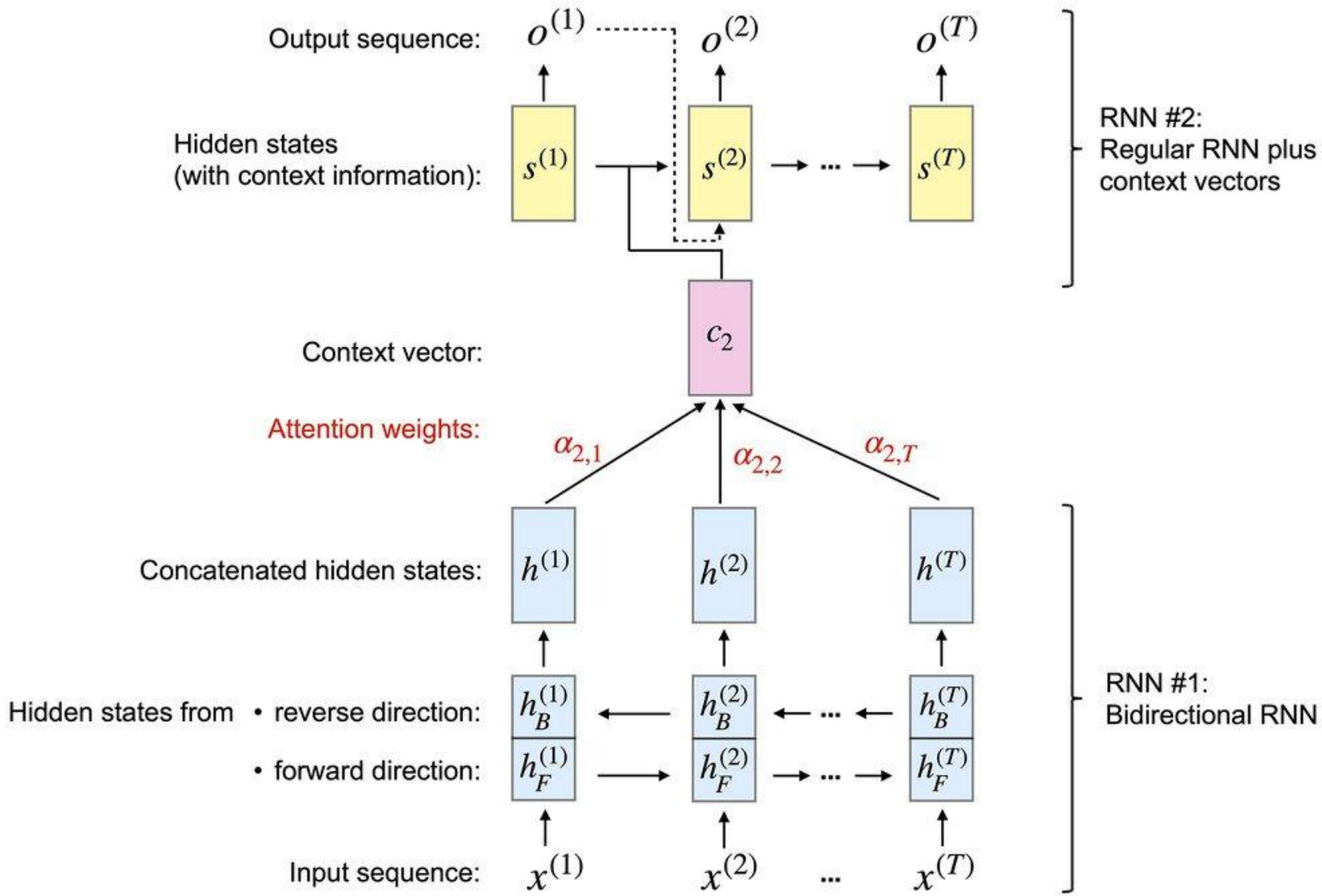
- RNNs allow the processing of time-dependent or **sequential data**. The output at a time step t depends on the previous outputs at steps 1,2,…,t - 1
- In a standard RNN, the output only considers information from words positioned before the current word
- This is problematic, as any given word could be contingent on the words both preceding and following it. **Bidirectional RNNs** solve this problem by traversing each input sequence in both directions and combining the resulting outputs.

- Bidirectional RNN has forward and backward RNNs.



https://devopedia.org/bidirectional-rnn

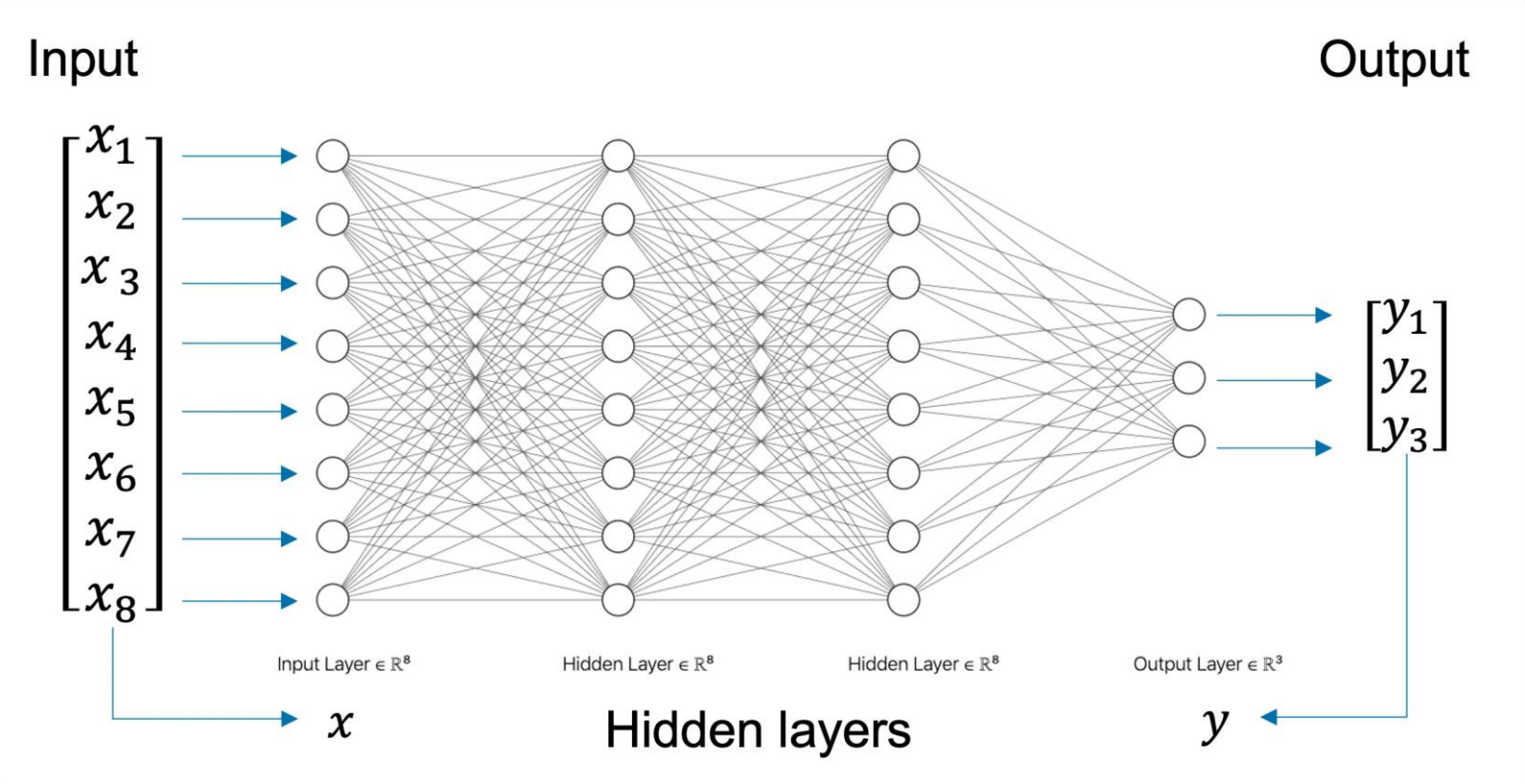
ATTENTION MECHANISMS



https://www.kdnuggets.com/2022/03/packt-adding-attention-mechanism-rnns.html
https://www.youtube.com/watch?v=mDZil99CtSU

FEEDBACK ALIGNMENT

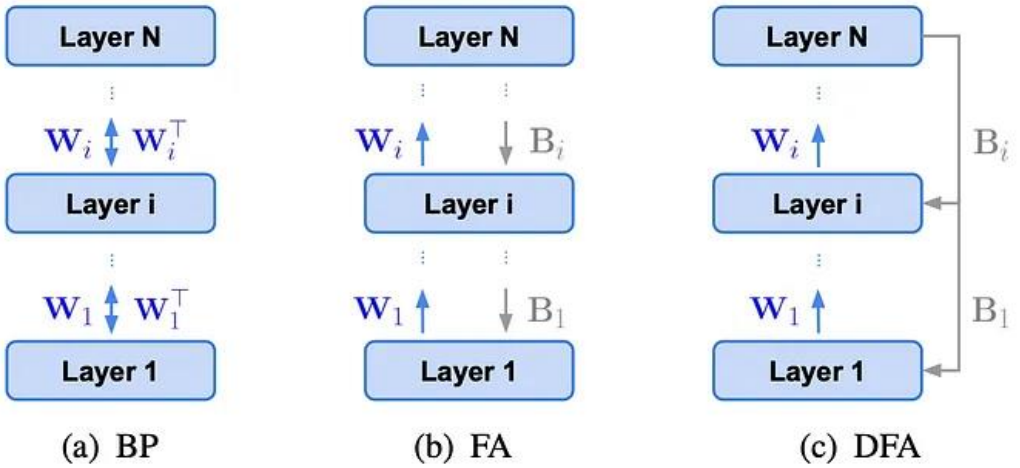
- Neural networks are composed of neuron layers connected to one another. This structure allows the input signal to be processed into an output signal using a series of non-linear transformations, which in the end creates a global complex transformation. The transformation performed by a given neural network is defined by its parameters (weights & biases) and activation functions.



https://towardsdatascience.com/harder-better-faster-lighter-deep-learning-with-direct-feedback-alignment-9e837b076baa

FEEDBACK ALIGNMENT

- The speed of the sequence-related algorithm is limited by the fact that backpropagation is inherently sequential, thereby prohibiting parallel computation.
- Many believe that backpropagation is not biologically feasible, which places an interesting theoretical obstruction in the analogies between deep learning and the human brain.
- Apply for fully connected networks.
- Feedback Alignment (FA): The weight updates are computed in the same fashion as in backpropagation, but the backward weight matrix is a random matrix.
  - Remove transposition
- Direct Feedback Alignment (DFA): While the weight update in FA is computed recursively across layers, it is possible to project the error propagation by directly backwarding the derivative of the loss at the last layer to all
  - Remove transposition and sequentiality



https://towardsdatascience.com/feedback-alignment-methods-7e6c41446e36

TRANSFER LEARNING

- Transfer learning is a training method based on the idea that knowledge learned for one task is often useful in many other contexts.
- The main approach here is to use a **pre-trained** model as the starting point for building a model performing a similar task.
  - It is often difficult to obtain a large and varied enough training data set to train a model that generalizes well. In this case, it makes sense to train on a larger, similar data set and then fine-tune the model on the more specialized data.
  - Training a model from scratch demands a large amount of computational resources. We can reduce this amount drastically by starting with a pre-trained model.
  - It is difficult to decide on an appropriate network architecture for a given task. Transfer learning allows us to adapt models created by deep learning specialists.

TRANSFER LEARNING METHODS

- The typical approach to transfer learning is to copy the first n layers (along with the weights learned during training) from a pre-trained model to the first n layers of a target network.
- The choice of n depends on the similarity of the tasks. We then add one or more new layers with randomly initialized weights, which are then optimized for the new task during training.
  - **Fine-tuning**
- mini-course-dl-for-biotech.pdf

TRANSFER TASK

- https://github.com/duongtrung/minicourse-dl-bio

REVIEW STUDY GOALS

- After completing this unit, you will be able to ...

  … describe how the concept of attention is modeled in deep learning.

  … explain what bidirectional RNNs are and how they are used in language translation.

  … understand how feedback alignment addresses issues in the backpropagation algorithm.

  … use synthetic gradients and decoupled neural interfaces to parallelize computations and speed up training.

  … recognize the importance of transfer learning in real-world applications of deep learning.

SESSION 4

Alternative Training Methods

1. What is the basic concept of RNN?

A. Use recurrent features from dataset to find the best answers.

B. Use previous inputs to find the next output according to the training set.

C. Use a loop between inputs and outputs in order to achieve the better prediction.

D. Use loops between the most important features to predict next output.

2. For what RNN is used and achieve the best results?

A. Handwriting and speech recognition.

B. Speech and images recognition.

C. Financial predictions.

D. Handwriting and images recognition.

3. The other RNN´s issue is called 'Vanishing Gradients'. What is that?

A. When the values of a gradient are too small and the model joins in a loop because of that.

B. When the values of a gradient are too big and the model joins in a loop because of that.

C. When the values of a gradient are too big and the model stops learning or takes way too long because of that.

D. When the values of a gradient are too small and the model stops learning or takes way too long because of that.

4. What is LSTM?

A. LSTM networks are an extension for recurrent neural networks, which basically extends their memory. Therefore it is not recommended to use it, unless you are using a small Dataset.

B. LSTM networks are an extension for recurrent neural networks, which basically extends their memory. Therefore it is well suited to learn from important experiences that have very low time lags in between.

C. LSTM networks are an extension for recurrent neural networks, which basically extends their memory. Therefore it is well suited to learn from important experiences that have very long time lags in between.

D. LSTM networks are an extension for recurrent neural networks, which basically shorten their memory. Therefore it is well suited to learn from important experiences that have very low time lags in between

LEARNING CONTROL QUESTIONS

1. B
2. A
3. D
4. C