1. Which architecture is more expressive, keeping dimensions and layers fixed?\*

|  |  |  |
| --- | --- | --- |
| **RNNs (Recurrent Neural Network)** | LSTMs (Long-Short-Term-Memory) - one prevalent gated RNN | GRUs (Gated Recurrent Units) |
| Single layer | Combination of layers |  |
|  |  |  |

1. Which of these language models rely on RNNs?\*

**ELMO**

BERT

GPT-2

1. What's the advantages of fastText over word2vec?\*

|  |  |
| --- | --- |
| FastText | Word2vec |
| Build on feed forward networks | Build on feed forward networks |
| Based on the type |  |
| Operates at a character level | Operates at a word level |
| Character n-grams, words are represented by the sum of the character n-gram vecots | Uses words to predict words |

Computational speed, i.e., it's faster

**Lexical coverage, i.e., it can represent more words**

Interpretability, i.e., it is easier to debug

*Multilinguality, i.e,. it works on more languages*

1. What's the loss function ELMO (Embeddings from Language Models) is trained with?

L2 Loss (Mean squared error) -

Mean Absolute Error

Hinge Loss

**Cross-entropy Loss**

1. What's the difference between CNNs and RNNs, and in which cases would use each one?

|  |  |
| --- | --- |
| RNN | CNN (Convolutional Neural Network) |
| Temporal or sequential data, good at making predictions | Sparse data, generally used to analyze visual images |
| A loop neural network | A feed forward neural network |
| Has some kind of memory that captures information about what has been calculated | Does not have any kind of memory |
| Handle arbitrary input or output lengths | Handle fixed size inputs and generate fixed size outputs according to the input |
| Perform the same task for every element of a sequence, with the output being depended on the previous computations | Only considers the current input |
| Predictive, reusing activation functions from other data points in the sequence to generate next output in a series | Employ filters with convolutional layers to transform data |
| Works primarily on time series information on the past influence of the consumer | Special for video processing and image processing |
| Works primarily on speech analysis and text analysis | Follows interconnectivity pattern between the neurons which is inspired by the animal visual cortex, where the individual neurons are organized in a way that they respond to overlapping areas tilling the visual field |
| Computationally slower | More powerful and has more feature compatibility |

**CNNs is applicable for sparse data and generally is used to analyze visual images or videos, while RNNs is applicable for temporal or sequential and is good at making predictions of the input. Since CNNs is a feed forward neural network, it only considers the current input and does not have any kind of memory, therefore, it handles fixed size inputs and generates fixed size outputs according to the inputs. However, RNNs is a loop neural network, it has internal memory to capture information about what has been calculated, therefore, It handles arbitrary input or output lengths. In addition, CNNs is considered as more powerful and has more feature compatibility while RNNs is considered as computationally slower. Based on their features, CNNs is special for image processing and video processing and RNNs is special for text analysis and speech analysis.**

1. How many dimensions must the inputs of RNN layers have? What do the dimensions represent?

**The inputs of RNN layers must have two dimensions, and they represent weights and biases.**

1. Bonus Question: What's the difference between Recurrent Neural Networks and Recursive Neural Networks?

**Recursive Neural Networks (Recursive NNs) is a generalization of Recurrent Neural Networks (Recurrent NNs), with a different kind of computational graph, which is structured as a deep tree instead of the chain structure in Recurrent NNs (Deep Learning, Chapter 10.6 Recursive Neural Network, 2016). Therefore, in both Recursive NNs and Recurrent NNs the weights are shared, and dimensionality remains constant along the length of the sequence. In Recursive NNs, each element can be connected to more than one element. Where in Recurrent NNs, elements are connected in a sequential way. Hence, Recursive NNs is better to generate a parse tree since it creates hierarchical representations.**