Lecture 5 – Sequence Processing

Provide short answers (max. 25 words each) to the following questions.

- Why do classic feedforward neural networks struggle with sequence data?
- Give an example of a task that benefits from sequence processing.
- What is the main idea behind Recurrent Neural Networks (RNNs)?
- How can a Recurrent Neural Network (RNN) be visualized when unrolled?
- Name three applications of Recurrent Neural Networks (RNNs).
- Why do vanilla Recurrent Neural Networks (RNNs) struggle with long-term dependencies?
- What is the key advantage of LSTM over vanilla RNN?
- What is the role of the cell state in LSTM?
- What mechanism allows LSTM to control information flow?
- What activation function is used in LSTM gates? Why?
- What does the forget gate do in LSTM?
- What does the input gate do in LSTM?
- How is the output of an LSTM computed?
- What are peephole connections in LSTM?
- What is the coupled forget and input gate variant of LSTM?
- How does GRU differ from LSTM?
- What is the main advantage of GRU over LSTM?
- What is ConvLSTM and where is it used?
- What is the main idea of NBoFs?
- What are the three main layers in NBoFs?
- What does the quantization layer in NBoFs do?
- What does the accumulation layer in NBoFs do?
- What is the purpose of attention in NBoFs?
- Name three types of attention in NBoFs.
- What function is commonly used to compute attention weights?
- What does Recurrent NBoFs add to standard NBoFs?
- How does a bilinear layer process sequence data?
- What is TABL (Temporal Attention-Augmented Bilinear Network)?

Multiple-choice questions

Why do classic feedforward networks struggle with sequence data?

A) They have too many parameters
B) They do not persist information across time steps
C) They use convolutional layers only
D) They require attention mechanisms

Which of the following is an example of sequence data?

A) Single image classification
B) Static tabular data
C) Financial time-series
D) Image segmentation

What is the main idea behind RNNs?

A) Use pooling layers for temporal data
B) Maintain a hidden state across time steps
C) Replace activations with attention
D) Use convolution for sequences

How can an RNN be visualized when unrolled?

A) As multiple copies of the same network across time steps
B) As a single layer repeated
C) As a feedforward network without loops
D) As a convolutional stack

Which of these is NOT an application of RNNs?

A) Speech recognition
B) Machine translation
C) Image classification
D) Video action recognition

What is the key advantage of LSTM over vanilla RNN?

A) Faster training
B) No need for backpropagation
C) Fewer parameters
D) Ability to learn long-term dependencies

What is the role of the cell state in LSTM?

A) Stores weights
B) Acts as long-term memory
C) Computes gradients
D) Normalizes activations

What mechanism controls information flow in LSTM?

A) Pooling layers
B) Dropout layers
C) Gates (forget, input, output)

D) Batch normalization

Which activation function is used in LSTM gates?

A) ReLU
B) Sigmoid
C) Softmax
D) GELU

What does the forget gate do?

A) Decides which information to discard from cell state
B) Adds new information to cell state
C) Outputs the hidden state
D) Normalizes the input

What does the input gate do?

A) Decides which new information to add to cell state
B) Removes old information
C) Outputs the cell state
D) Computes attention weights

How is the LSTM output computed?

A) Using pooling layers
B) By filtering cell state through sigmoid and tanh
C) By concatenating all gates
D) By applying softmax directly

What are peephole connections in LSTM?

A) Connections that allow gates to access cell state
B) Connections between hidden layers only
C) Dropout applied to gates
D) Attention applied to gates

What is the coupled forget and input gate variant?

A) Combines decisions about forgetting and adding information
B) Removes all gates
C) Uses attention instead of gates
D) Adds convolution to gates

How does GRU differ from LSTM?

A) Uses more gates
B) Merges cell state and hidden state

C) Removes hidden state
D) Adds peephole connections

What is the main advantage of GRU?

A) More complex than LSTM
B) Simpler architecture with fewer parameters
C) Requires attention
D) No gating mechanism

What is the main idea of NBoFs?

A) Treat sequence as a set of items and aggregate features
B) Use convolution for sequences
C) Apply attention only
D) Replace RNNs with pooling

Which layers form NBoFs?

A) Convolution, pooling, softmax
B) Quantization, accumulation, MLP
C) Attention, normalization, dropout
D) GRU, LSTM, ConvLSTM

What does the quantization layer do?

A) Aggregates features
B) Maps inputs to codewords using RBF neurons
C) Applies attention
D) Normalizes sequence length

What does the accumulation layer do?

A) Aggregates quantized features into a single representation
B) Applies convolution
C) Computes attention weights
D) Normalizes activations

What is the purpose of attention in NBoFs?

A) To select or weight important sequence parts
B) To normalize activations
C) To compute gradients
D) To reduce dimensionality

Which function is commonly used for attention weights?

A) Sigmoid

B) Softmax
C) ReLU
D) GELU

Name one type of attention in NBoFs.

A) Forget attention
B) Codeword attention
C) Pooling attention
D) Dropout attention

What does Recurrent NBoFs add?

A) Attention mechanism
B) Convolution layers
C) Recurrent mechanism for sequential quantization
D) Dropout layers

Which gates in Recurrent NBoFs are inspired by GRU?

A) Forget and input gates
B) Reset and update gates
C) Attention gates
D) Peephole gates

How does a bilinear layer process sequences?

A) Applies pooling
B) Transforms input matrix using feature and temporal transformations
C) Uses convolution only
D) Applies Dropout

Give one application of Temporal Attention-Augmented Bilinear networks.

A) Financial time-series forecasting
B) Image segmentation
C) Speech synthesis
D) Object detection