

Transformers

Week 11 · November 3 & 5, 2026

Harun Pirim

Table of contents

1	Learning Objectives	1
2	Processing Sequences	2
3	Dot-Product Self-Attention	2
4	Scaled Dot-Product Attention	2
5	Multi-Head Attention	3
6	The Transformer Block	3
7	Positional Encoding	3
8	Encoder, Decoder, and Encoder-Decoder Models	3
9	Self-Attention from Scratch	3
10	Summary	3
11	Exercises	3
12	Further Reading	4

Reading Prince, *Understanding Deep Learning*, Chapter 12 · **Due** Presentation paper selection, Thursday, November 5

 These notes are still being written. The reading above is authoritative until this banner disappears.

1 Learning Objectives

By the end of this week you should be able to:

1. Explain why sequence data breaks the assumptions behind fully connected and convolutional layers.
2. Derive scaled dot-product attention and say why the scaling factor is there.
3. Describe queries, keys, values, and multi-head attention.
4. Assemble a full transformer block: attention, feed-forward, residuals, layer norm.
5. Compare encoder, decoder, and encoder-decoder models and their uses.
6. Implement self-attention from scratch and check it against PyTorch.

```
import numpy as np
import matplotlib.pyplot as plt
import torch

torch.manual_seed(774)
np.random.seed(774)

plt.rcParams.update({
    "figure.figsize": (7, 4),
    "figure.dpi": 130,
    "axes.spines.top": False,
    "axes.spines.right": False,
    "axes.grid": True,
    "grid.alpha": 0.25,
    "font.size": 10,
    "legend.frameon": False,
})

NDSU_GREEN = "#005643"
NDSU_GOLD = "#FFC72C"
ACCENT = "#C1121F"
```

2 Processing Sequences

TODO

3 Dot-Product Self-Attention

TODO

4 Scaled Dot-Product Attention

TODO

5 Multi-Head Attention

TODO

6 The Transformer Block

TODO

7 Positional Encoding

TODO

8 Encoder, Decoder, and Encoder-Decoder Models

TODO

9 Self-Attention from Scratch

TODO

10 Summary

TODO

11 Exercises

Problem numbers refer to Prince, Chapter 12.

1. TODO
 2. TODO
 3. Computational. TODO
-

12 Further Reading

- Prince, *Understanding Deep Learning*, Chapter 12.
- TODO