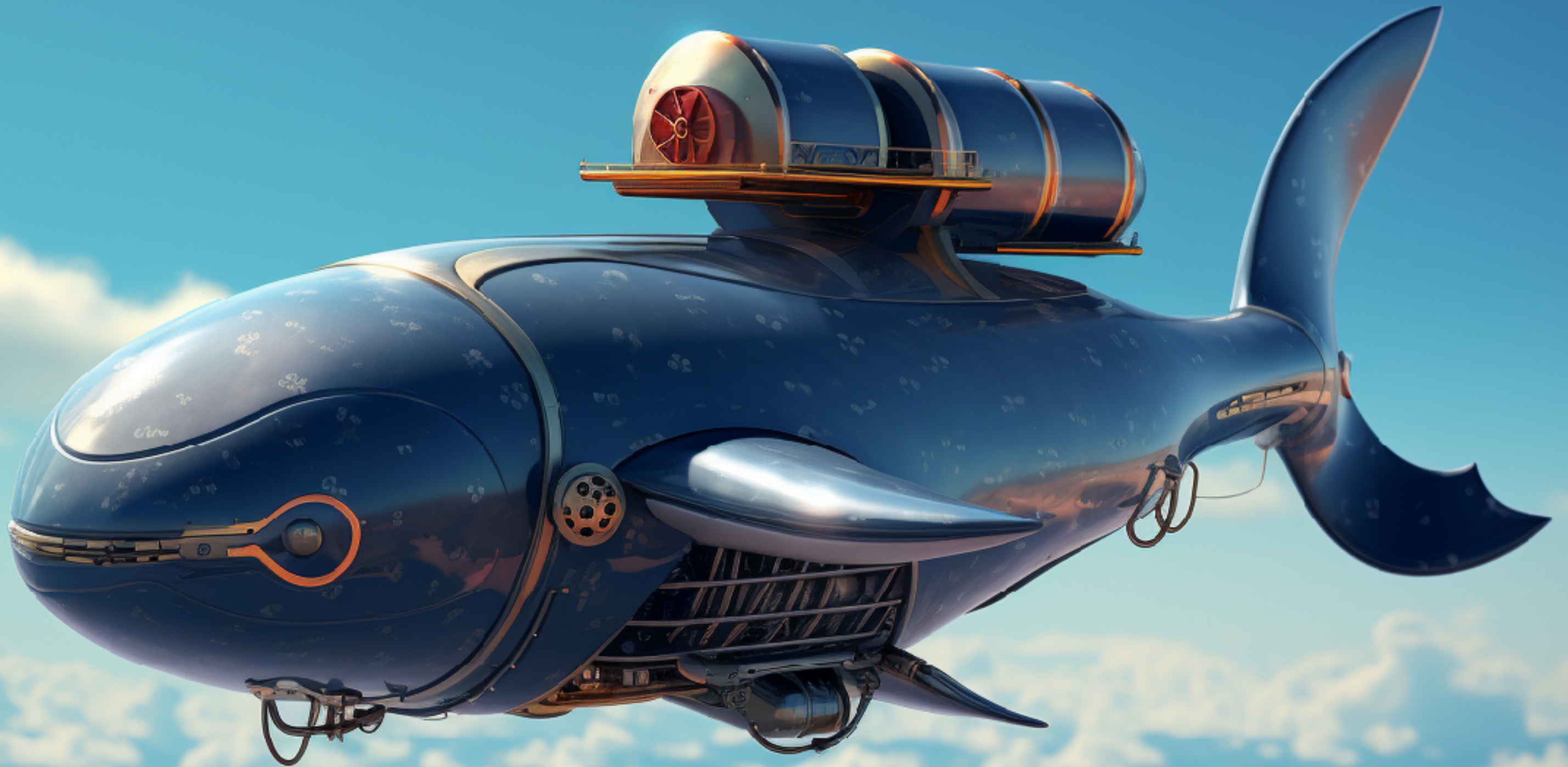


Transformerを詳しく学ぼう！



Section3

Section3の概要



講座の内容

Section1. Transformerの概要

Section2. Attentionの仕組み

 **Section3. Transformerにおける埋め込み**

Section4. Transformerを組み立てる

今回の内容

1. Section3の概要
2. Token Embedding
3. Positional Encoding
4. Position-wise Feed-Forward Networks
5. Layer Normalization
6. 演習

教材の紹介

- **Pythonの基礎:**

python_basic

- **Section3の教材**

01_token_embedding.ipynb

02_positional_encoding.ipynb

03_positionwise_feed_forward.ipynb

04_layer_norm.ipynb

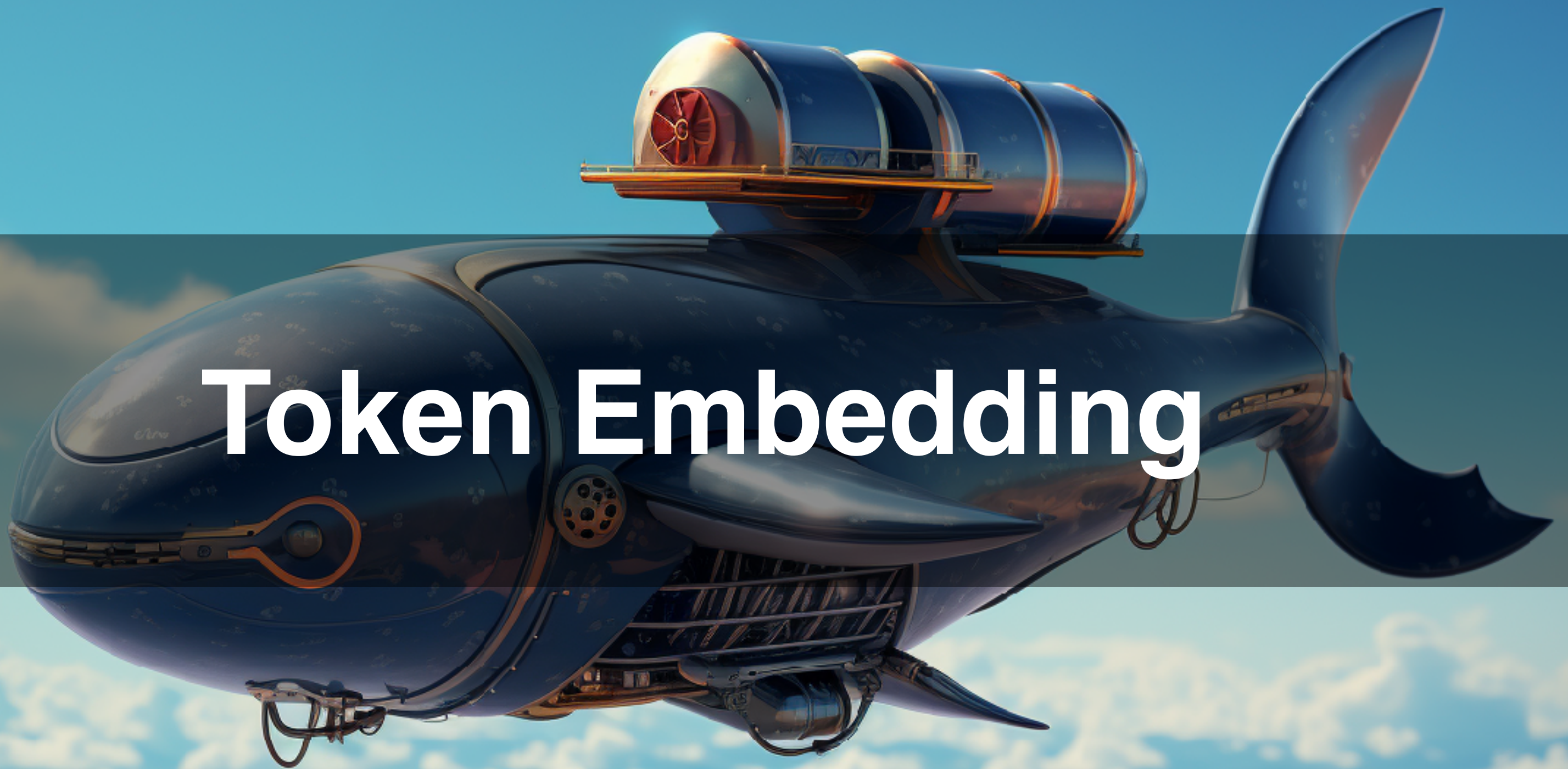
05_exercise.ipynb

https://github.com/yukinaga/learning_transformer/

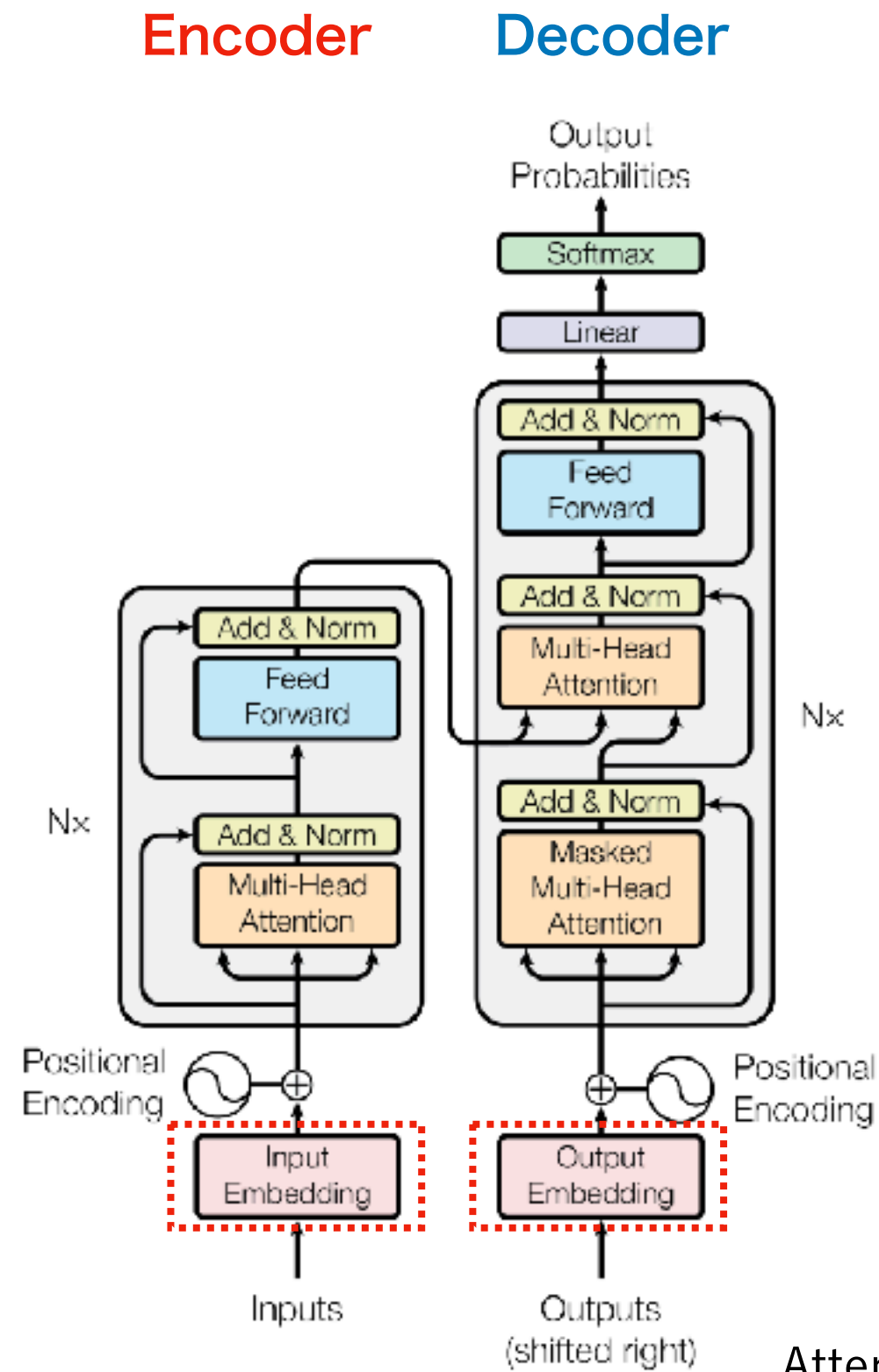
Section2演習の解答例

- 03_exercise.ipynb

Token Embedding



Token Embedding



nn.Embedding()

- <https://pytorch.org/docs/stable/generated/torch.nn.Embedding>

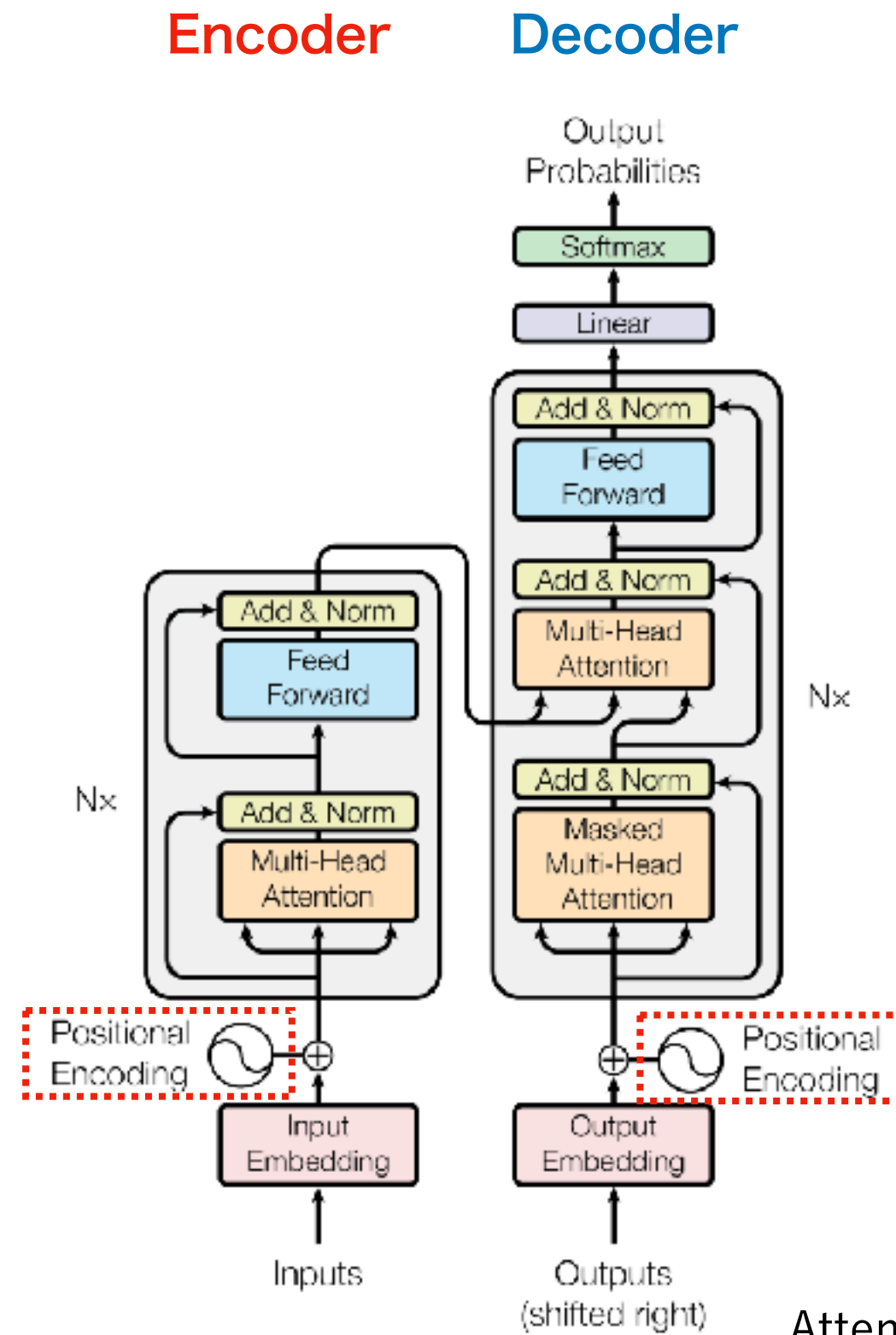
Token Embedding

- 01_token_embedding.ipynb

Positional Encoding



Positional Encoding



Positional Encoding

- Positional Encoding

→ 「単語の位置」の情報を加える

$$PE_{(pos, 2i)} = \sin(pos/10000^{2i/d_{\text{model}}})$$

$$PE_{(pos, 2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}})$$

pos: 単語の位置 2i, 2i+1: Embedding の何番目の次元か d_{model} : 次元数

Positional Encoding

- 02_positional_encoding.ipynb

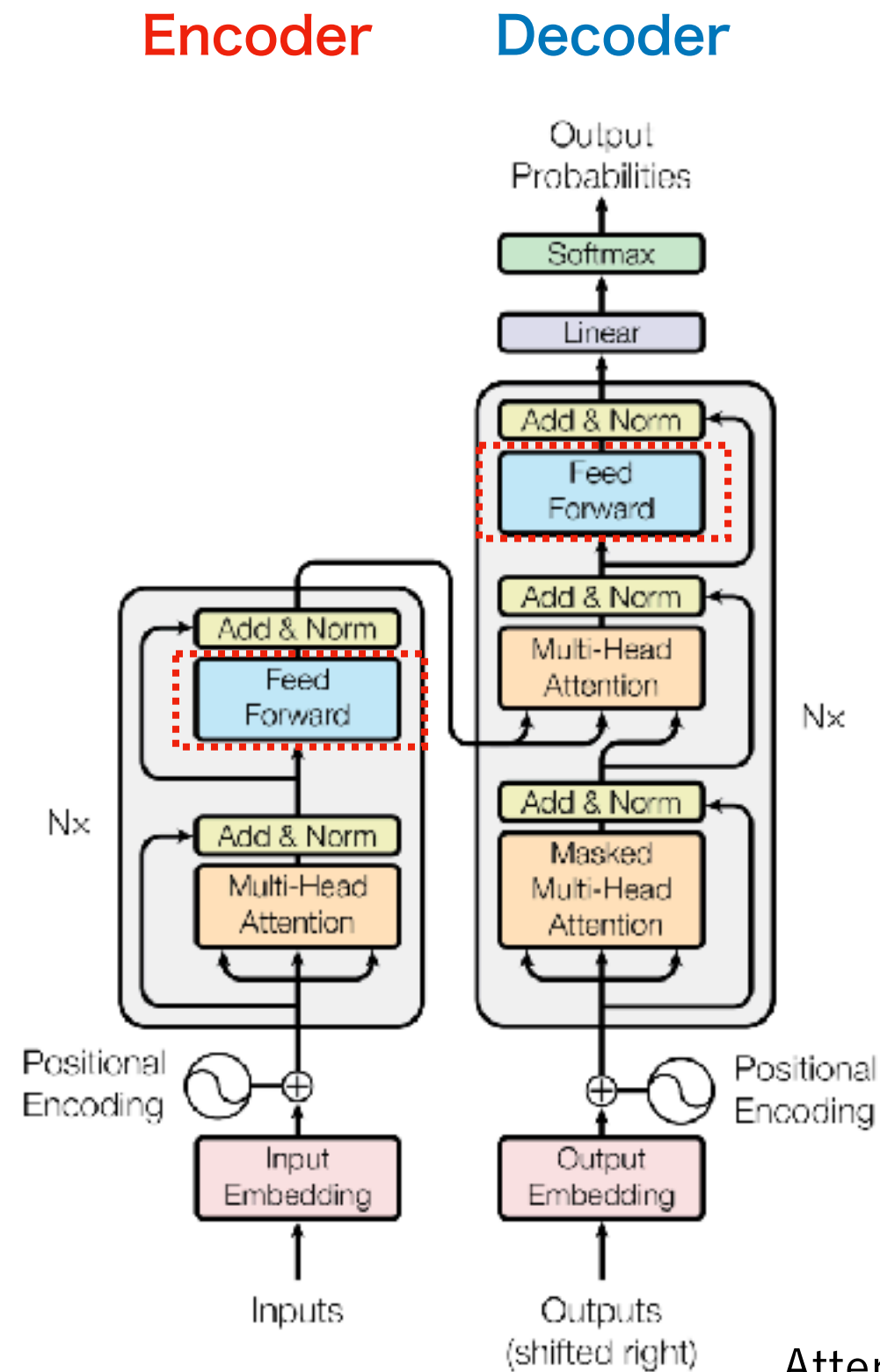


休憩...

Position-wise Feed-Forward Networks

A detailed illustration of a steampunk-style airship, primarily blue with white and gold accents, flying through a bright blue sky filled with soft white clouds. The airship features a large propeller at the rear, a complex engine structure on top, and various mechanical details like pipes and rivets. The overall aesthetic is reminiscent of early 20th-century aviation with a fantastical, mechanical twist.

Position-wise Feed-Forward Networks



Position-wise Feed-Forward Networks

- **Positionwise fully connected feed-forward network**
 - 2 層の全結合ニューラルネットワーク
 - 単語の位置ごとに個別の順伝播ネットワーク
 - 他単語との影響関係を排除
 - パラメータは全てのネットワークで共通

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2$$

Attention Is All You Need, Ashish, V. et al. (2017) より引用

Position-wise Feed-Forward Networks

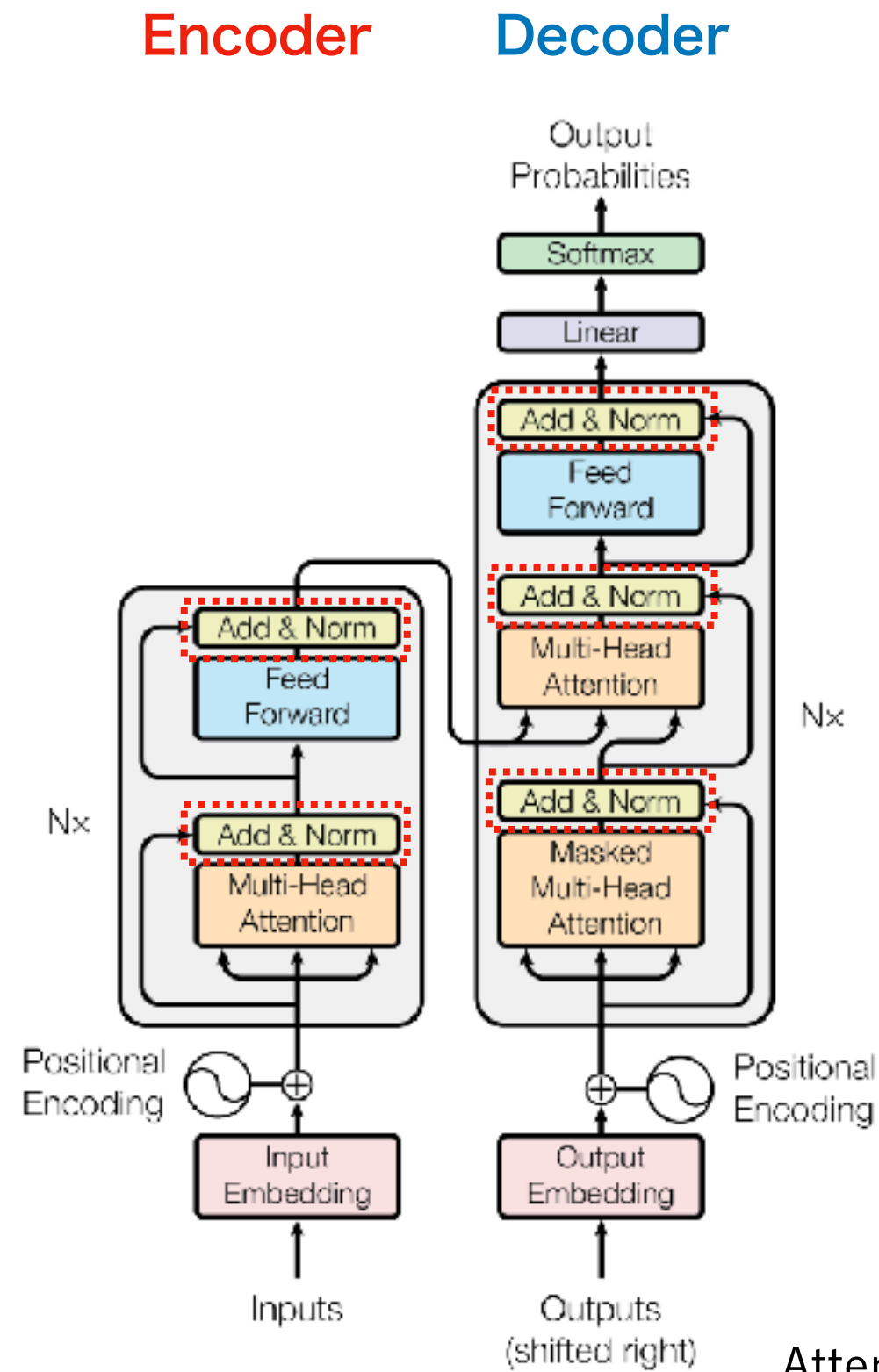


- 03_positionwise_feed_forward.ipynb

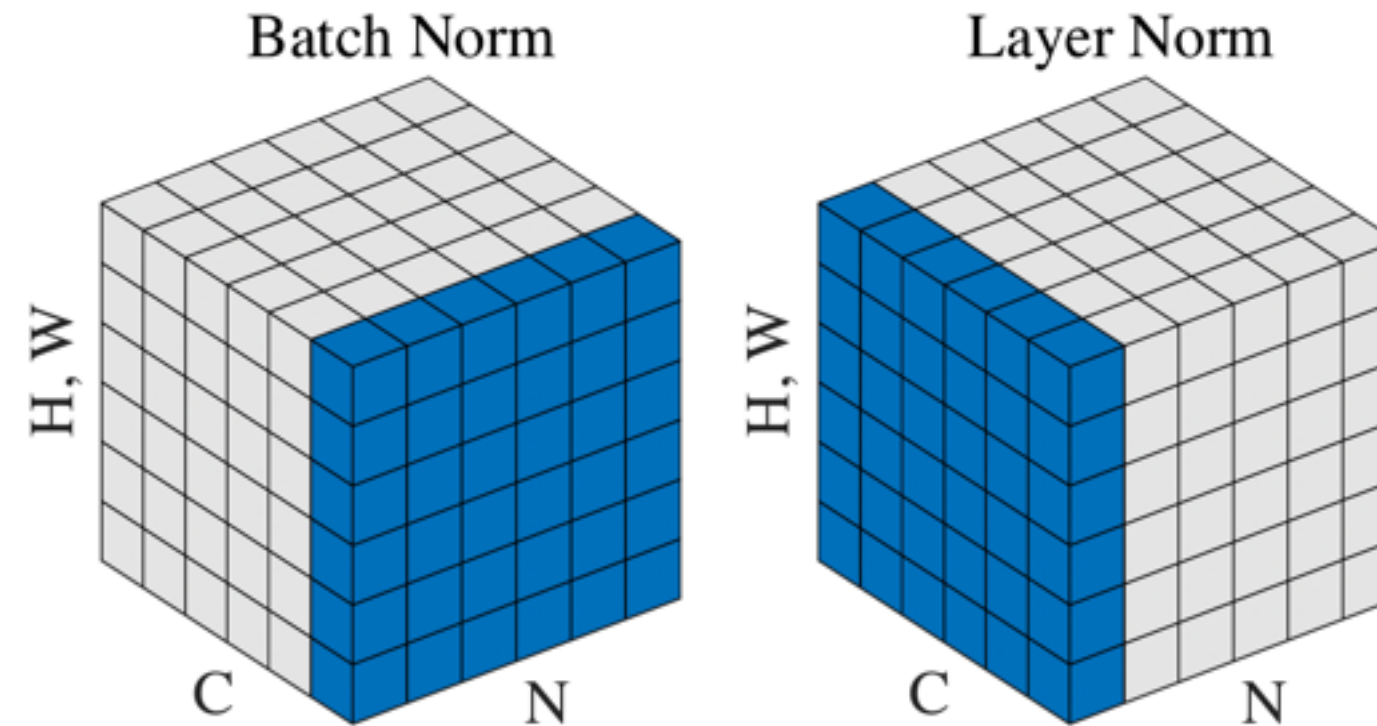
Layer Normalization



Layer Normalization



Layer Normalization



Group Normalization, Yuxin Wu, et al. (2018) より引用

$$\mu^l = \frac{1}{H} \sum_{i=1}^H a_i^l \quad \sigma^l = \sqrt{\frac{1}{H} \sum_{i=1}^H (a_i^l - \mu^l)^2}$$

Layer Normalization, Jimmy Lei Ba, et al. (2016) より引用

Layer Normalization



- 04_layer_norm.ipynb

演習



演習

- 05_exercise.ipynb

次回の内容

Section1. Transformerの概要

Section2. Attentionの仕組み

Section3. Transformerにおける埋め込み

 **Section4. Transformerを組み立てる**