Use ROZETKA descriptions from the previous HW and make  
same classifier using:  
- FCNN  
- LSTM  
See how accuracy changed comparing to previous models

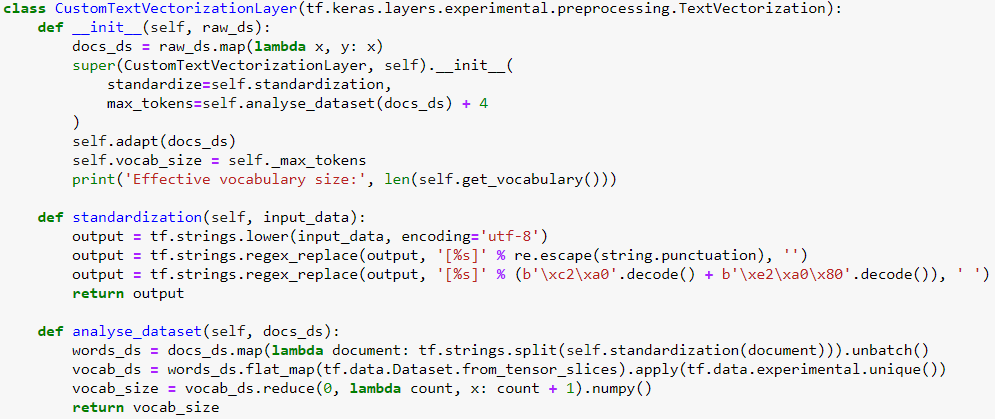
1. **Вибір архітектури моделі:**

Архітектура трансформера зазвичай складається з двох класів компонентів: Encoder і Decoder. Перший віповідає за виділення контексту із вхідної послідовності. Другий же необхідний для генерації вихідної послідовності на основі, зокрема, отриманого контексту. Оскільки для задачі класифікації відсутня необхідність генерації нової послідовності, достатньо використати лише компоненти Encoder.

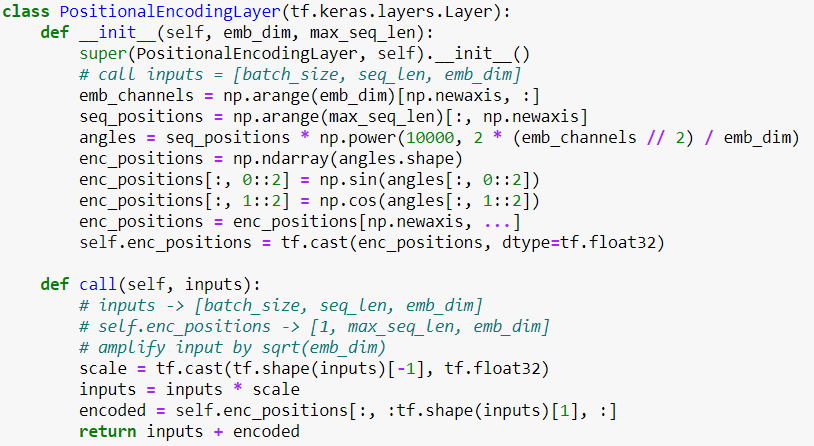
Виходячи з цього, було прийнято розробити наступну архітектуру моделі класифікації:

* Вхідний шар tf.keras.Input, який приймає рядок вхідних даних, тобто сам текст опису товару;
* Шар CustomTextVectorizationLayer, який розділяє рядок на слова та призначає кожному слову унікальне число (аналогічно в попередній роботі);
* Стандартний шар tf.keras.layers.Embedding;
* Шар PositionalEncodingLayer доповнює вектори embedding інформацією про позицію слова відповідно до схеми з синусами/косинусами;
* Один чи декілька шарів EncodingLayer, які фактично є Encoder’ом трансформера;
* Шар tf.keras.layers.GlobalAveragePooling1D, який здійснює перехід із вектору розмірності [batch\_size, seq\_len, emb\_dim] до вектору розмірності [batch\_size, emb\_dim];
* Шар tf.keras.layers.Dense виконує роль прихованого шару стандартної Feed Forward мережі;
* Вихідний шар tf.keras.layers.Dense для трьох класів;

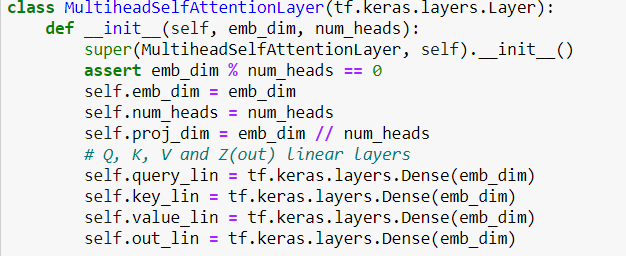
1. **Шар CustomTextVectorizationLayer:**

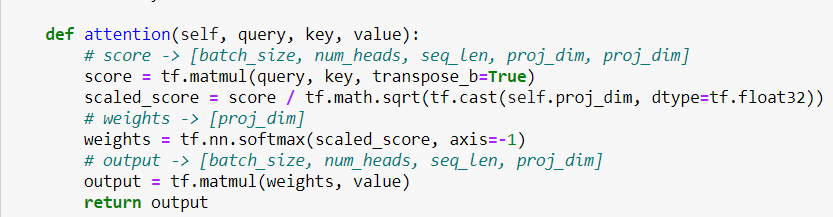
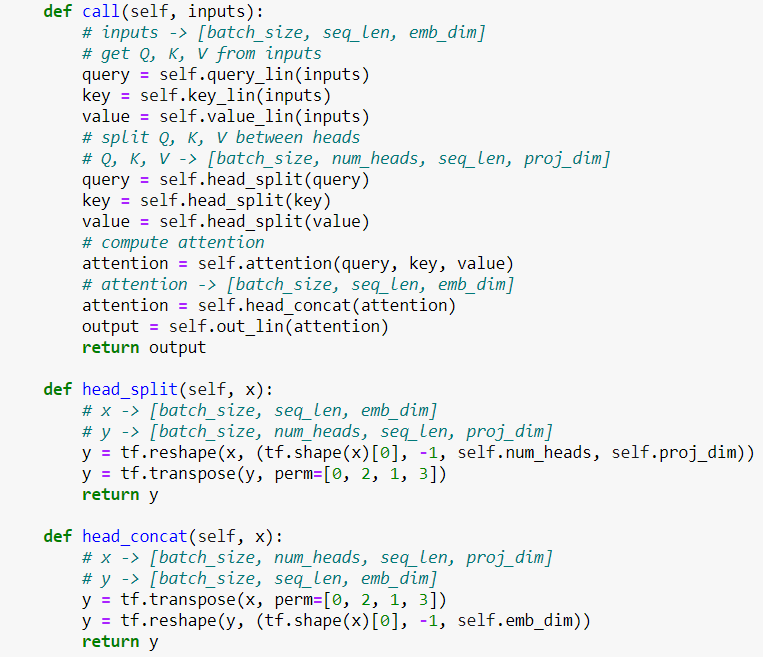
****

1. **Шар PositionalEncodingLayer:**

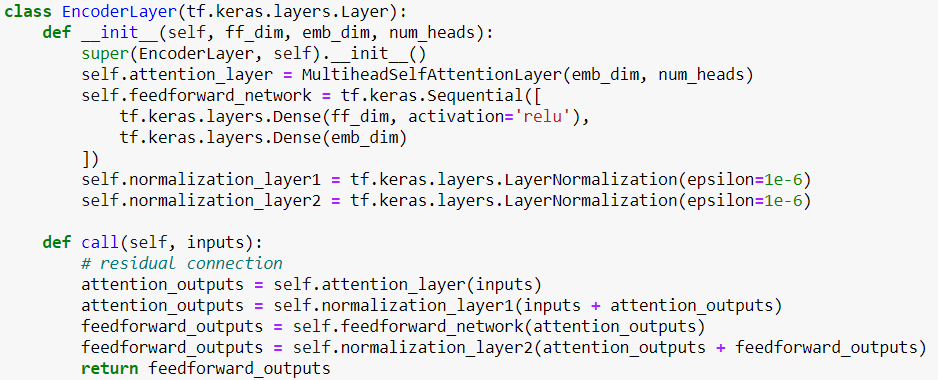


1. **Шар MultiheadSelfAttentionLayer:**

****

****

1. **Шар EncoderLayer:**

****

1. **Створення та результати тестування моделі:**

****

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Розмірність emb\_dim /  num\_heads | Кількість параметрів | Час навчання на епоху | Досягнута точність | Графік навчання |
| 128/4 | 2,770,691 | 43с | 97-98% |  |
| 64/4 | 1,356,675 | 32с | 97% |  |
| 128/4  128/4  (два шари Encoder) | 2,870,275 | 82с | 99% |  |
| 64/4  64/4  (два шари Encoder) | 1,381,891 | 60с | 97-98% |  |

**Висновки**

Для задачі класифікації описів товарів використання компоненту Encoder трансформера дозволило збільшити точність моделі на декілька відсотків, досягнувши 99% на даних для тестування у порівнянні, у порівнянні з FCNN або LSTM. Оскільки процес навчання виконувався CPU, а не GPU, не вдалося досягти задовільного часу тренування (перевагою трансформерів є широка можливість розпаралелювання).