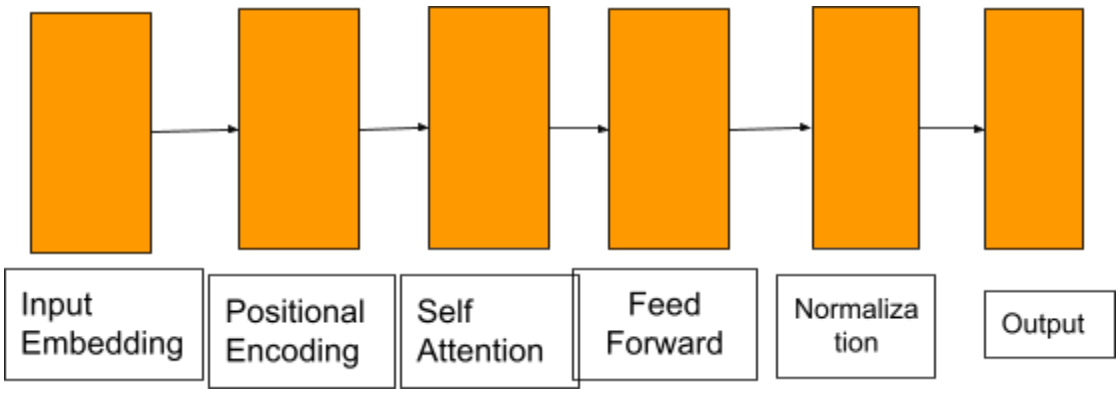
ADD ON COURSE(GENERATIVE AI)

Day 1(24-06-2024)

Self attention-finding important points in a sentence .Important parameters are QUERY,KEY,VALUE(QKV) ,Self attention=softmax(QK^T/$\sqrt{dk}$)V

ARCHITECTURE



Transformers

Layers

How transformers work-Each are blocks/layers these are combination of transformers.Feature extraction takes place in these layers.

- **Input Embedding**
  Converting words to no.s.Tokenisation is taking place (every single characters are converted to no.s .This is important)
  4 dimensional vector representation
- **Positional Encoding**
  Gives extra no.s to the no.s from the i/p embedding.

$$PE(1) = [\sin(\frac{1}{10000^{2\times0/4}}), \cos(\frac{1}{10000^{2\times0/4}}), \sin(\frac{1}{10000^{2\times1/4}}), \cos(\frac{1}{10000^{2\times1/4}})]$$
$$PE(2) = [\sin(\frac{2}{10000^{2\times0/4}}), \cos(\frac{2}{10000^{2\times0/4}}), \sin(\frac{2}{10000^{2\times1/4}}), \cos(\frac{2}{10000^{2\times1/4}})]$$
$$PE(3) = [\sin(\frac{3}{10000^{2\times0/4}}), \cos(\frac{3}{10000^{2\times0/4}}), \sin(\frac{3}{10000^{2\times1/4}}), \cos(\frac{3}{10000^{2\times1/4}})]$$
$$PE(4) = [\sin(\frac{4}{10000^{2\times0/4}}), \cos(\frac{4}{10000^{2\times0/4}}), \sin(\frac{4}{10000^{2\times1/4}}), \cos(\frac{4}{10000^{2\times1/4}})]$$
$$PE(5) = [\sin(\frac{5}{10000^{2\times0/4}}), \cos(\frac{5}{10000^{2\times0/4}}), \sin(\frac{5}{10000^{2\times1/4}}), \cos(\frac{5}{10000^{2\times1/4}})]$$
$$PE(6) = [\sin(\frac{6}{10000^{2\times0/4}}), \cos(\frac{6}{10000^{2\times0/4}}), \sin(\frac{6}{10000^{2\times1/4}}), \cos(\frac{6}{10000^{2\times1/4}})]$$

$$PE = Sin\left(\frac{pos}{1000^{\left(\frac{2i}{dmodel}\right)}}\right)$$

$$PE = Cos\left(\frac{pos}{1000^{\left(\frac{2i}{dmodel}\right)}}\right)$$

Where model=4 size of matrix(vector)

- **<u>Self Attention</u>**(imp)

  softmax(QK^T/$\sqrt{dk}$)V

  $Q = XW_Q$, $K = XW_K$, V = XW_V

  From softmax we will get probability
  Q=Quetstion
  K= Features
  V=Vectors
  T= Tangentiability

- **<u>Feed Forward</u>**(imp)

  Indepth feature extraction ,add new no.s.

  $FFN(X) = max(0, xw_1 + b_2)W_2 + b_2$

  Relu Activation fn,max(0,x)
  Output will be h= [0.927 , 0.5913 ,0.755 , 0.919]

- **<u>Normalization</u>**

  Normalizing output,errors in the model are resolved

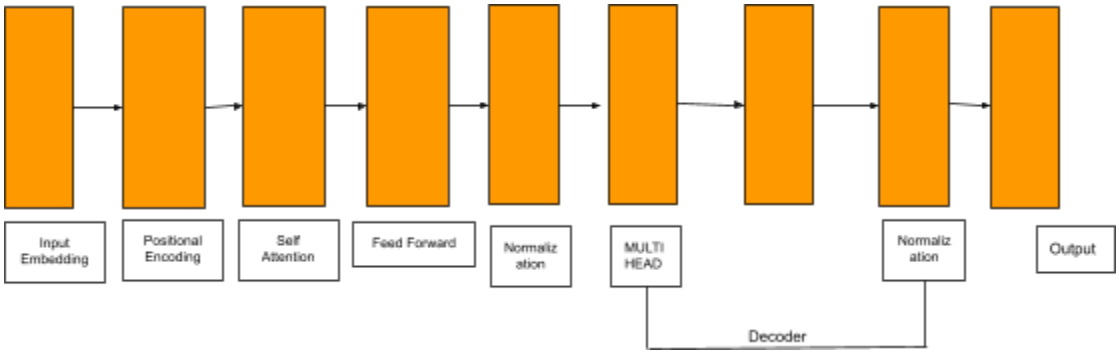
  LayerNorm(x)=$\left(\frac{x-\mu}{\sigma}\right)*\gamma + \beta$

  x=n

  $\mu = mean$

  $\sigma = sd$

<u>DECODER LAYER</u>



multihead(Q,K,V)=cncat(head1,head2 ,....headn)

- Q projection,K projection,V projection,gain projection, up projection,down projection

- **<u>Output</u>**
  May be functional to get probability

Example:

The quick brown fox jump over a lazy dog.

(9 words so e^1 *to* e^9 *assigned in order.*
The set values (refer note))

Day 2(25-06-2024)