Positional Encoding

- What it is ?
- A method to inject sequence order into attention models.

- Goal:
To tell the model where each token appears in the sequence.

- Math:
For position p and dimension i:

$$PE(p, 2i) = \sin\left(\frac{p}{10000^{2i/d}}\right)$$

$$PE(p, 2i+1) = \cos\left(\frac{p}{10000^{2i/d}}\right)$$

~~Findi~~ Final input:

$$X_{input} = X + PE$$

Conclusion:
Positional encoding gives attention models a sense of order without recurrence.