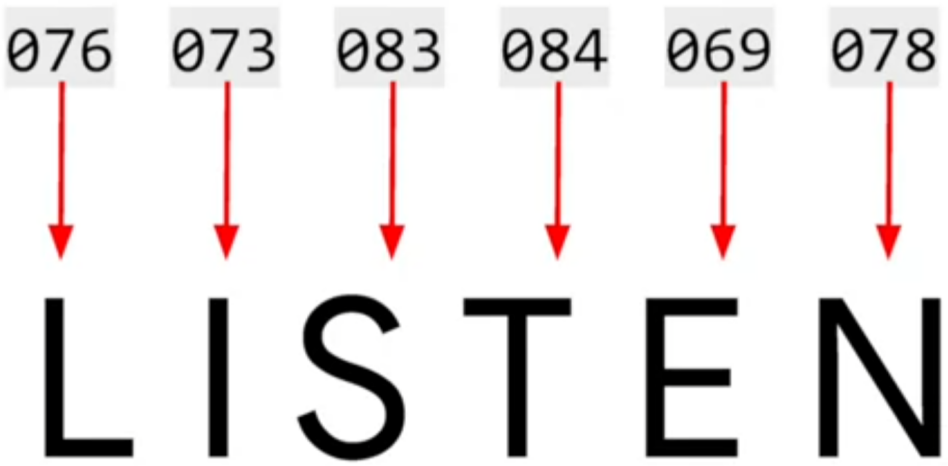
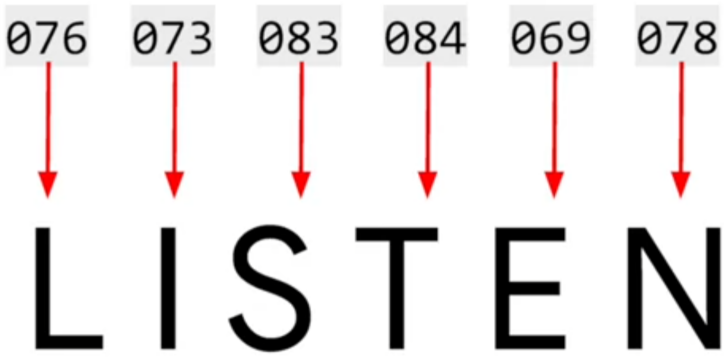
Week 1

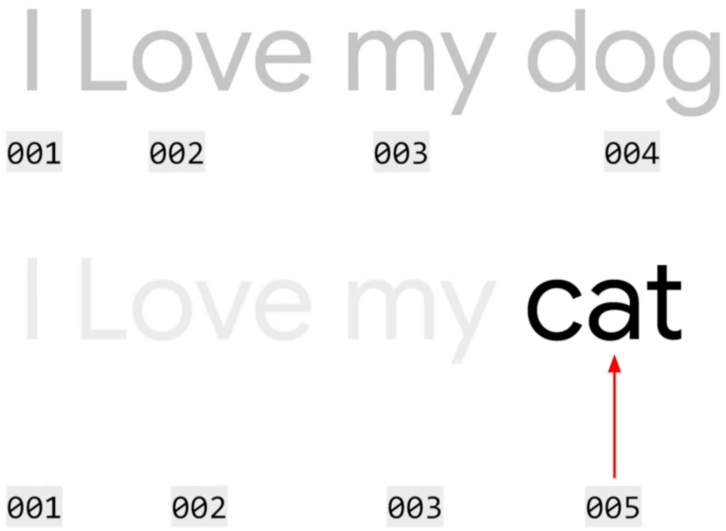
- In convolutional network, the images are numbers already
- But in words, they are characters not numbers
- A possibility could be using ACCII encoding and turn it to numbers like below:



- This does not work because as the example below, silent means a completely different thing but have similar numbers



- A actual way to do it could be giving a number to each word
- An example is listed below



- `keras` tokenizer takes care of that

- `tf.keras.preprocessing.text.Tokenizer` also removes punctuation when processing words
- The code below processes a few simple sentences and encode them

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/65a8dff4-83cc-492a-a686-8daafbaa6c24/simple_text.py

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/a9711377-5f7d-4e76-a101-c7bd7ac079d3/simple_text.ipynb

Resizing Text length (Padding)

- It's required for all texts to be at the same length to process just like image processing
- In order to use padding function, it's required to import `pad_sequences` from `tensorflow.keras.preprocessing.sequence`
- Can change padding location by adding `padding='post'`
- Can override maximum length of the sentence by setting `maxlen=10`
- set truncating location, default is in front, change this by editing `truncating='post'`

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/b39e7763-8bba-4cc1-ade4-10af13090861/padding_text.ipynb

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/24a2b195-d813-42fe-a66a-707c1819bf64/padding_text.py

Out of Vocabulary (OOV)

- It's better to replace unseen words with a place holder

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/89d03a7b-82d6-4489-9aae-3e45e5c31ee9/OOV.ipynb

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/e6c5f69a-f4a8-4e0c-97cc-06d5549ef59e/oov.py

Sarcasm example

- This example uses data from Sarcasm in News Headlines Dataset by_ Rishabh Misra
- Code:

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/75bbd699-56f8-4b47-83db-f7c778d38d7d/sarcasm_example.ipynb

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/7fd5873d-0ccc-435d-8401-da0aa5df6543/sarcasm_example.py

Exercises

- Code for exercise 1

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/b900ba75-5287-42a5-98f1-656a694590e8/Exercise_1.ipynb

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/8f8ab56c-5af0-4de2-bcb2-f2a2c3dfea84/exercise_1.py