

Assignment Report

--> Anubhav Sharma-2018114007

❖ Both Skip Gram and CBOW models were trained on full softmax, with following parameters:

- Embedding length : 100
- Window Size : 3
- No. of epochs: 5
- Dataset: 50k reviews

Also the embeddings it learnt on smaller datasets(say for 1000 lines) were moving in the correct directions as that of gensim. IT was possible only with a gradual learning rate , where I reduced the learning rate as the model progressed.

10 closest words for camera(CBOW):

In gensim:

- Cameras
- Canon
- Rebel
- Camcorder
- Scope
- Lens
- Slr
- Nikon
- Elph
- Western

In my model (on a dataset of 1000 sentences and 5 epoch)

- Stand
- within
- Wilderness
- excellent
- Inflexible
- Dinner
- Monday
- Vinyl
- Urge

In my model (on a dataset of 50000 sentences and 2 epoch)

- Lens
- Camcorder
- Picture
- Frame
- Feature
- Bag
- Scope
- Light
- Body
- Speed

10 closest words for camera(Skip-Gram)

In gensim:

- Cameras
- Cam
- Rebel
- Camcorder
- Digicam

3

- Lens
- Lense
- Canon
- Elph
- Minolta

In my model (on a dataset of 1000 sentences and 3 epoch with offset):

- picture
- Features
- Mode
- Mention
- Machine
- Light
- Bugs
- Quickly

In my model (on a dataset of 50000 sentences and 2 epoch)

- Camcorder
- dslr
- Lens
- Cannon
- Tripod
- Set
- Features
- Picture
- Resolution
- Frame