

## Feature extraction from text and images

Total points 6

1. Select true statements about n-grams

2 points

- ☐ N-grams features are typically sparse
- ☐ Levenshtein should always be applied before computing n-grams
- ☐ N-grams always help increase significance of important words
- ☐ N-grams can help utilize local context around each word

2. Select true statements.

1 point

- ☐ Meaning of each value in BOW matrix is unknown.
- ☐ Semantically similar words usually have similar word2vec embeddings.
- ☐ You do not need bag of words features in a competition if you have word2vec features.
- ☐ Bag of words usually produces longer vectors than Word2vec

3. Suppose in a new competition we are given a dataset of 2D medical images. We want to extract image descriptors from a hidden layer of a neural network pretrained on the ImageNet dataset. We will then use extracted descriptors to train a simple logistic regression model to classify images from our dataset.

2 points

We consider to use two networks: ResNet-50 with imagenet accuracy of X and VGG-16 with imageNet accuracy of Y ( $X < Y$ ). Select true statements.

- ☐ Descriptors from ResNet 50 will always be better than the ones from VGG-16 in our pipeline.
- ☐ For any image descriptors from the last hidden layer of ResNet-50 are the same as the descriptors from the last hidden layer of VGG-16.
- ☐ It is not clear what descriptors are better on our dataset. We should evaluate both.
- ☐ With one pretrained CNN model you can get only one vector of descriptors for an image
- ☐ Descriptors from ResNet-50 and from VGG-16 are always very similar in cosine distance.

4. Data augmentation can be used at (1) train time (2) test time

1 point

- ☐ True, True
- ☐ True, False
- ☐ False, True
- ☐ False, False

Coursera Honor Code [Learn more](#)

- ☐ I, **Chinmay kumar Das**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Submit

Save draft