1. **What does it mean to perform sentiment analysis on text, and give an example?**

2. **Explain how one can use a bag-of-words model to featurize a text document. Name a pro and con of bag-of-words**

3. **In bag-of-visual-words for image classification, what makes up the vocabulary in this model, how is this similar to text?**

4. **What is the purpose of interest point detection, and how is it different from a regular grid model?**

5. **Why is it sometimes important to reduce the size of a vocabulary, and what are some ways to do this?**

6. **What are stop-words and why are they relevant to bag-of-words?**

7. **What are n-grams and how can they be used to help model a document?**

8. **What is the main goal of a SVM and how does it relate to sentiment analysis/image classification?**

9. **What is the main purpose of bag-of-words?**

    A. Feature Generation
    B. Regularization
    C. Classification
    D. Clustering

10. **Which of the following could be used in conjunction with bag-of-words?**
    A. PCA
    B. K-means
    C. SVM
    D. All of the above

11. **Which of the following makes sense to use instead of the raw count of a word?**
    A. Frequency (n_occurances / n_words_in_document)
    B. log (count)
    C. Normalized feature vectors (divide each bag by its l2 norm)
    D. All of the above

12. How does tf-idf scoring differ from traditional bag-of-words?

A. Puts more weight on words that appear in many documents and also puts more weight on words that occur a lot in a single document
B. Takes away weight from words that appear in many documents and puts less weight on words that occur a lot in a single document
C. Takes away weight from words that appear in many documents and puts more weight on words that occur a lot in a single document
D. Puts more weight on words that appear in many documents and puts less weight on words that are occur a lot in a single document