Homework 4 Written Questions

Document Instructions

- 5 questions **[8 + 10 + 8 + 4 + 10 = 40 + 2 bonus points]**.
- Fill all your answers within the answer boxes, and **please do NOT remove the answer box outlines**.
- Questions are highlighted in the **orange boxes**, bonus questions are highlighted in **blue boxes**, answers should be recorded in the **green boxes**.
- Include code, images, and equations where appropriate.
- To identify all places where your responses are expected, search for 'TODO'.
- The answer box sizes have been set by the staff beforehand and **your responses should not exceed the green borders**. Any overfull text may be truncated, so make sure your responses fit. **Extra pages are not permitted unless otherwise specified.**
- Make sure your submission has the right number of pages to validate page alignment sanity (check the footer).
- Please make this document anonymous.

Gradescope Instructions

- When you are finished, compile this document to a PDF and submit it directly to Gradescope.
- The pages will be automatically assigned to the right questions on Gradescope *assuming you do not add any unnecessary pages*. **Inconsistently assigned pages will lead to a deduction of 2 points per misaligned page (capped at a maximum 6 point deduction).**

Q1: [8 points]

(a) **[4 points]**

Describe the main difference between these two types of metric errors:

(i) **[1 point]** Bias **[1-3 sentences]**

TODO: Your answer for (a) (i) here

(ii) **[1 point]** Variance **[1-3 sentences]**

TODO: Your answer for (a) (ii) here

(iii) **[2 points]**

Suppose you find that a model you've created is showing significant metric error. How would you mitigate an error stemming from bias compared to how you would mitigate an error stemming from variance? **[1-3 sentences]**

TODO: Your answer for (a) (iii) here

(b) **[2 points]**

Define these terms in the context of evaluating a classifier:

Only A1 (a) should be on this page

(i) **[1 point]** Overfitting **[1-3 sentences]**

TODO: Your answer for (b) (i) here

(ii) **[1 point]** Underfitting **[1-3 sentences]**

TODO: Your answer for (b) (ii) here

(c) **[2 points]**

How do overfitting and underfitting relate to bias and variance? **[2-5 sentences]**

TODO: Your answer for (c) here

Q2: [10 points] Suppose we create a visual word dictionary using SIFT and k-means clustering for a scene recognition algorithm.

(a) **[4 points]**

What characteristics would be desirable in the dataset that you choose to use to build your visual dictionary? Consider data bias trends in popular machine learning datasets, the groups that these biases can harm, and how your choice in dataset could help. **[5-7 sentences]**

TODO: Your answer for (a) here

(b) **[4 points]** After choosing an appropriate dataset, examination of the SIFT features generated from our training database tells us that many features are almost equidistant from two or more visual words.

(i) **[2 points]**

Why might this affect classification accuracy? **[2-4 sentences]**

TODO: Your answer for (b) (i) here

(ii) **[2 points]**

Why shouldn't we pick out a new dataset that better matches our model? **[4-6 sentences]**

TODO: Your answer for (b) (ii) here

(c) **[2 points]**

Given the situation, describe *two* methods to improve classification accuracy, and explain why they would help. (*These can be for k-means, or otherwise.*) **[4-6 sentences]**

Only A2 (b) should be on this page

TODO: Your answer for (c) here

Q3: **[8 points]** The bag of words representation handles the spatial layout of information in a way that can be an advantage or a disadvantage in different cases.

(a) **[5 points]**

How might we determine whether bag of words is a good model? Make sure to discuss both technical and nontechnical benchmarks for a ”good” model. **[5-6 sentences]**

TODO: Your answer for (a) here

(b) **[2 points]**

Describe two example scenarios when dealing with visual information, one that illustrates an advantage of the bag of words representation, and another that shows a disadvantage of the bag of words representation. **[5-6 sentences]**

TODO: Your answer for (b) here

(c) **[1 point]**

Describe a modification or additional algorithm which might overcome the disadvantage. **[2-4 sentences]**

TODO: Your answer for (c) here

Q4: **[4 points]** Given a linear classifier such as SVM which separates two classes (binary decision), how might we use multiple linear classifiers to create a new classifier which separates k classes?

Below, we provide pseudocode for a linear classifier. It trains a model on a training set, and then classifies a new test example into one of two classes. Please edit the pseudo-code to convert this into a multi-class classifier.

Hints: See slides in supervised learning crash course deck, plus your own research. You can take either the one vs. all (or one vs. others) approach or the one vs. one approach in the slides; please declare which approach you take.

More hints: Be aware that:

1. The input labels in the multi-class case are different, and you will need to match the expected label input for the `train_linear_classifier` function
2. You need to make a new decision on how to aggregate or decide on the most confident prediction

Note: A more efficient software application would separate the classifier training and testing into two different functions so that the model could be reused without retraining. Feel free to ignore this for now.

TODO: Select the implementation you chose.

One vs One	☐
One vs Many	☐

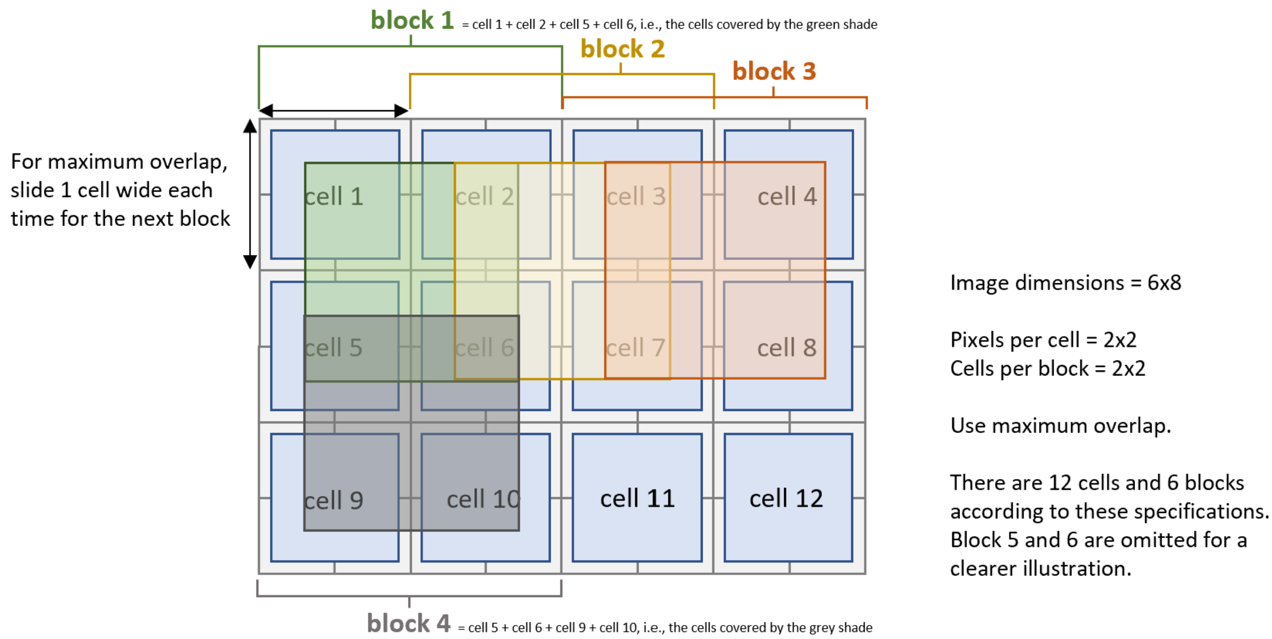
```
# Inputs
#   train_feats: N x d matrix of N features each d descriptor
                            long
#   train_labels: N x 1 array containing values of either -1
#               (class 0) or 1 (class 1)
#   test_feat: 1 x d image for which we wish to predict a label
#
#   Outputs: -1 (class 0) or 1 (class 1)
#
# Inputs:
#    As before, except
#    train_labels: N x 1 array of class label integers from 0 to
                             k-1
# Outputs:
#    A class label integer from 0 to k-1
#

# TODO: Turn this into a multi-class classifier for k classes.
def classify(train_feats, train_labels, test_feat)
    # Train classification hyperplane
    weights, bias = train_linear_classifier(train_feats,
                                    train_label)
```

```
    # Compute distance from hyperplane
    test_score = weights * test_feats + bias

    return 1 if test_score > 0 else -1

    ################################################
    # YOU MAY USE THIS ADDITIONAL PAGE

    # WARNING: IF YOU DON'T END UP USING THIS PAGE
    # KEEP THESE COMMENTS TO MAINTAIN PAGE ALIGNMENT
    ################################################
```

Q5: **[10 points]** In this homework, we will use a feature descriptor called HOG—'Histogram of Oriented Gradients'. As its name implies, it works similarly to SIFT. In classification, we might extract HOG features across the entire image (not just at interest points).

HOG creates one feature descriptor per image 'block'. Each block is split into 'cells' covering pixels. HOG outputs a 9-bin histogram of oriented gradients per cell. We append these together to obtain the feature descriptor for each block. As a result, if we have (z, z) cells per block, the feature descriptor for each block will be of size $z \times z \times 9$. In other words, computing HOG over the whole image produces a matrix where each row is a descriptor.

Blocks can overlap as displayed in the diagram below.



(a) **[6 points]**

Given a 72×72 image, calculate the number of cells, blocks, and final feature vector size that will occur when we extract HOG features over the whole image with the following parameters using maximum overlap between blocks.

(i) **[3 points]** Scenario 1: Pixels per cell = 4×4, cells per block = 4×4

1. **[1 points]** Number of cells:

TODO: Your answer for (a) (i) (1) here

2. **[1 points]** Number of blocks:

TODO: Your answer for (a) (i) (2) here

3. **[1 points]** Dimensions of feature vector for the whole image:

TODO: Your answer for (a) (i) (3) here

(ii) **[3 points]** Scenario 2: Pixels per cell = 8×8, cells per block = 2×2.

1. **[1 points]** Number of cells:

TODO: Your answer for (a) (ii) (1) here

2. **[1 points]** Number of blocks:

TODO: Your answer for (a) (ii) (2) here

3. **[1 points]** Dimensions of feature vector for the whole image:

TODO: Your answer for (a) (ii) (3) here

(b) **[4 points]** When using HOG, the parameters such as pixels per cell and cells per block impact the resulting feature descriptor and so our performance on a classification task.

What are the pros and cons of the two parameter combinations? Which might you expect to have better performance? **[3-6 sentences]**

TODO: Your answer for (b) here

Discussion Attendance:

Extra Credit: [2 points]

Please mark this box only if you've attended the discussion session in person.

☐ I attended the discussion session on DATE

Feedback? (Optional)

Please help us make the course better. If you have any feedback for this assignment, we'd love to hear it!