Project 3 Questions

Instructions

- 6 questions.
- Write code where appropriate; feel free to include images or equations.
- Please make this document anonymous.
- This assignment is **fixed length**, and the pages have been assigned for you in Gradescope. As a result, **please do NOT add any new pages**. We will provide ample room for you to answer the questions. If you *really* wish for more space, please add a page *at the end of the document*.
- **We do NOT expect you to fill up each page with your answer.** Some answers will only be a few sentences long, and that is okay.

Questions

Q1:

(a) Define these common terms in machine learning:

(i) Bias

(ii) Variance

(b) Define these terms in the context of evaluating a classifier:

(i) Overfitting

(ii) Underfitting

(c) How do overfitting and underfitting relate to bias and variance?

Please answer overleaf.

A1: Your answer here.

(a) Assume there is a relationship between Y and X: $Y = f(X) + constant$ and $\hat{f}(X)$ is the model for Y.

(i) Bias is the difference between the mean of predictions of the model and the correct mean of the targets. i.e. $bias = E[\hat{f}(X)] - f(X)$

(ii) Variance represents the variability of model predictions for a given data point. It is used to estimate of the target function will change if different training data was used. i.e. $variance = E[(\hat{f}(X) - E[\hat{f}(X)])^2]$

(b) (i) Overfitting is the case where the model performs well on training dataset but poor on test dataset. It means that the generalization of the model is unreliable.

(ii) Underfitting is the case where the model performs poor both on training dataset and test dataset, meaning that the model has "not learned enough" from the training data, resulting in low generalization and unrealiable predictions.

(c) Overfitting has low bias and high variance. The low bias is because the model can predict close value to the training data. But add new data points can change the model much, which results in high variance. For overfitting, it has high bias but low variance because it predicts poor compared to training data but won't change too much when add new points.

Q2: Suppose we create a visual word dictionary using SIFT and k-means clustering for a scene recognition algorithm. Examining the SIFT features generated from our training database, we see that many are almost equidistant from two or more visual words.

(a) Why might this affect classification accuracy?

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

A2: Your answer here.

(a) If the SIFT features are equidistant from two or more words, it would be difficult to categorize words based upon Euclidean distance. In this case, small difference in distance will result in different catorgary and thus, classification accuracy will be decreased.

(b)
- One solution is changing the k value. This will increase the number of words and hence might break the features that are quidistant to previous words.
- Another solution for improved clustering would amplify distances corresponding to regions of input space near cluster boudaries and attenuate distances corresponding to regions of input space further away from cluster boundaries[1]. The effect of such a range transformation would make clusters more compact and well separated thus resulting in better clustering performance.

[1] Sharma, Piyush Kumar, and Gary Holness. "L^2 L2-norm transformation for improving k-means clustering." International Journal of Data Science and Analytics 3, no. 4 (2017): 247-266.

Q3: The way that the bag of words representation handles the spatial layout of visual information can be both an advantage and a disadvantage.

(a) Describe an example scenario for each of these cases.

(b) Describe a modification or additional algorithm which might overcome the disadvantage.

(c) How might we determine whether bag of words is a good model?

A3: Your answer here.

(a) Adavantage: when the spatial information is not needed in the word clustering, i.e. the image might not include spatial information.
Disadvantage: when the images need have the same color histogram, e.g. the examples shown in lecture, the spatial information is needed for better construction of bag of words.

(b) One possible modification could be using the spatial pyramid representation. The image will be converted to several levels of resolution to construct the locally orderless representations. In this way, the spatial layout information of the image will be included in the bag of words representation.

(c) To determine whether the model is good or not, we could use precision and recall to evaluate its quality. For example, the bag of words representations are constructed using target image. And we will calculate the precision and recall values for the similar images that the model returns. The model with high precision and low recall is a good model.

Q4: Data bias affects machine learning, and recent national news has highlighted data bias in object detection. In one case, researchers discovered that current pedestrian detection models identified darker-skinned pedestrians with 5% less accuracy than lighter-skinned pedestrians. The researchers investigate multiple reasons for this inaccuracy, but one reason could be that the training dataset had 29% darker-skinned and 71% lighter-skinned labeled pedestrians. While measuring this difference in the data might be simple for pedestrians, other data biases can be harder to describe.

In Project 3, we will train a scene recognition model using data from Lazebnik et al. 2006. Please review its data to check for biases: observe image samples in the data/train and data/test directories and consider their class labels.

(a) Does this dataset contain potentially harmful biases? If so, describe them and why they might be harmful. (3–4 sentences.)

Releasing a dataset publicly can caused it to be used in unforeseen applications or by unforeseen actors.

(b) Describe a use of this scene dataset that may have been unanticipated. How might this reveal other overlooked biases? (3–4 sentences.)

Please read the following two short articles: Article 1 and Article 2.

(c) Do these articles change your answers? Why or why not? (2–3 sentences).

Please answer overleaf.

A4: Your answer here.

(a) Yes, this dataset contains potentially harmful biases. For example, in highway catorgary, if in the training dataset, the number of images without cars are more than images with cars, it would be easier for the classifier to detect the images without cars is in highway catorgary than the images with cars.

(b) As the dataset includes scences such as bedroom, kitchen and living rooms, where might include some privacy information for the people living there, it could be used to analyze the living behavior of people. This can also cause bia as the images might come from certain countries and with such dataset, the model would perform poor in images taken in other countries or other cultures.

(c) No, these articles prove my answers. Analyzing the standard open source data sets such as ImageNet shows the poor distribution in geo-diversity. The geo-diversity, the income-level for different households, differences among different culture and countries will all contribute to the bias in the training dataset, which at last will make the model biased. The training dataset often reflects the life and backgound of the engineers responsible.

Q5: 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.

Please answer overleaf.

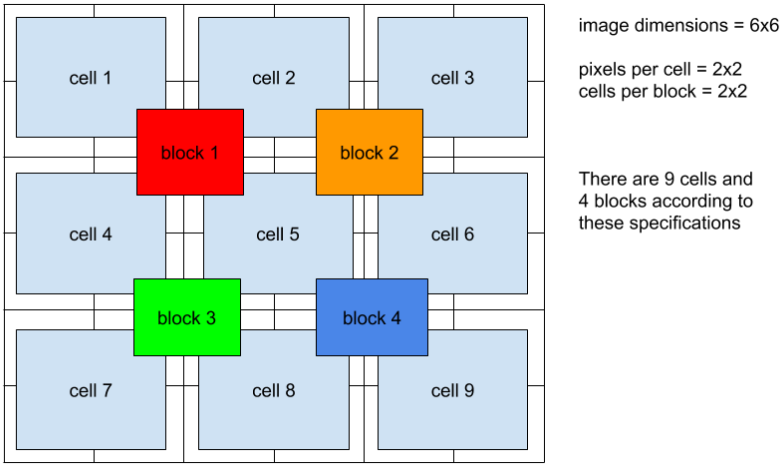
A5: Your answer here.
Here I use one v.s. others approach for multi-class classifications.

```
# 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)
#
# Please turn this into a multi-class classifier for k classes.
# 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
#

import numpy as np
import copy

def classify(train_feats, train_labels, test_feat):
# Train classification hyperplane
weights, bias = np.zeros(d, k), np.zeros(k)
for i in range(k):
    single_train_label = copy.deepcopy(train_label)
    single_train_label[train_label != i] = 0
    single_train_label[train_label == i] = 1
    weights[:, i], bias[i] = train_linear_classifier(train_feats,
                                        single_train_label)
# Compute distance from hyperplane
test_score = test_feats * weights + bias

return np.argmax(test_score)
```

Q6: In Project 3, 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 corners) to create visual words.

HOG creates a 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 $z \times z \times 9$.

Blocks can overlap as displayed in the diagram below.



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.

(a) Given a 72×72 image, calculate the number of cells, blocks, and feature vector size that will occur when we extract HOG features with the following parameters.

Scenario 1: Pixels per cell = 4×4, cells per block = 4×4
Calculate:
Number of cells:
Number of blocks:
Dimensions of resulting feature descriptor:

Scenario 2: Pixels per cell = 8×8, cells per block = 2×2.
Calculate:
Number of cells:
Number of blocks:
Dimensions of resulting feature descriptor:

(b) What are the pros and cons of the two parameter combinations? Which might you expect to have better performance?

Note: You may find it useful to look at the thesis of Navneet Dalal (co-inventor of HOG) for more on this topic. [Link to thesis] (pages 39, 41 in Section 4.3).

Please answer overleaf.

A6: Your answer here.

(a) Scenario 1:
Number of cells: 324
Number of blocks: 225
Dimensions of resulting feature descriptor: 32400

Scenario 2:
Number of cells: 81
Number of blocks: 64
Dimensions of resulting feature descriptor: 2304

(b) Scenario 1:
Pros: contains more information from the image.
Cons:

- too many dimensions for the feature, which can be slow for following operations such as matching, clustering, etc.
- adaptivity to local imaging conditions is weakened as the block is bigger.

Scenario 2:
Pros: proper number of dimensions of the featrue.
Cons: can miss some important local information in the image.

Comparing these two combinations, the number of cells for Scenario 1 is more than the number of cells in Scenario 2 while the numbers of blocks are the same for both Scenarios. This results in feature in Scenario 1 has more dimensions than the feature in Scenario 2. According to the reference, Navneet Dalal's thesis, the second one might has better performance. They found that 6-8 pixel wide cells do best irrespective of the block size and 2×2 and 3×3 cell blocks work best.

Feedback? (Optional)

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