Homework 5 Questions

Instructions

- 3 questions.
- Write code where appropriate.
- Feel free to include images or equations.
- Please make this document anonymous.
- **Please use only the space provided and keep the page breaks.** Please do not make new pages, nor remove pages. The document is a template to help grading. If you need extra space, please use and refer to new pages at the end of the document.

Questions

Q1: Given a linear classifier, how might we handle data that are not linearly separable? How does the *kernel trick* help in these cases? (See course slides in supervised learning, plus your own research.)

A1: We can handle data that are not linearly separable by using kernel tricks. Kernel trick is an inner product of two linear transformations that map original input space into some higher-dimensional feature space, where the training set is separable. Kernel function K is given as:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \tag{1}$$

This kernel K allows us to use it instead of explicitly computing the transformation $\phi(x)$. When testing a new point, we can use the following equation:

$$\sum_i \alpha_i y_i \phi(x_i) \cdot \phi(x) = \sum_i \alpha_i y_i K(x_i, x) + b \tag{2}$$

Instead of computing individual ϕ linear transformation values, we can compute the kernel K directly to separate data in a higher dimension.

Q2: In machine learning, what are bias and variance? When we evaluate a classifier, what are overfitting and underfitting, and how do these relate to bias and variance?

A2: Bias indicates how accurate the data are, or the difference between the predicted value and the correct value. Variance indiciates how precise the data are, or how much the target function will change if different training data were used.

Overfitting occurs when the model is too complex and is too fit to the training data. It contains irrelevant characteristics in the data. Since the model fits well with training data, it has **low bias**, but it is sensitive of training data, meaning that it has **high variance**.

Underfitting occurs when the model is too simple to represent enough characteristics in the training data. Since the model takes too many assumptions, it has **high bias**, but because it is not flexible, it has **low variance**.

Q3: The way that the bag of words representation handles the spatial layout of visual information can be both an advantage and a disadvantage. Describe an example scenario for each of these cases, plus describe a modification or additional algorithm which can overcome the disadvantage.

How might we evaluate whether bag of words is a good model?

A3: An advantage of bag of words is that the visual information is invariant to scale and orientation, since it uses feature extraction methods such as SIFT or Harris corner detection in order to extract feature descriptors. An example scenario would be comparing two images of different rotations. We if extract feature points using SIFT, for example, the features being compared will be invariant to rotation, and thus we will be able to compare the two images using the detected feature points.

A disadvantage of bag of words is that it ignores spatial relationship between patches, or the actual words. For example, consider a scenario of comparing two images where one image is showing cars travelling in the background, whereas the other image is showcasing a specific model of car. Since bag of words methods will put all the cars in the same bag without discerning between foreground and background, the two images may be declared to be similar, even though they may contain very diffferent contents. A possible modification to be problem is also adding the spatial relationship of the feature into the bag. By storing spatial relationship and feature along with the feature descriptors themselves into the bag, we will be able to discern spatial difference of similar features across the images.

We might evaluate whether bag of words is a good model by testing the advantage and disadvantage shown above. We may test its performance across images with different geometric and/or phometric transformations and see whether the results produce correct classification. Also, to test its flaws, we can test different images of similar features but with different spatial relationships. If the method classifies two different images as similar, then we can say that bag of words is not a good model for that scenario.