Homework 4 Questions

Instructions

- 4 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: Your answer here.

We usually transform dataset to higher-dimensional sapce when given data set are not linearly separable. When we use svm method, there are too many dot product between dataset. If we computation dot product between dataset after transforming dataset to higher-dimensional space, the computation cost is too much. However we can perform transforming and dot product at the same time, using kernel trick method that is cost cheaper and faster method.

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: In machine learning, expected test mse are composed of squared Bias and Variance of model and variance of error term. Here, variance of error term cannot be modified. Variance refers to the amount by which estimate result (ex. f hat) would change if we estimated it using a different training data set. Bias refers to difference between expected estimate result of model and correct value.

In general when p (number of parameter) less than n (number of data set), or less flexible, complex model, have high bias and low variance. But when p (number of parameter) more than n (number of data set), or more flexible, complex model, have low bias and high variance. Therfore, when model is simple, Test MSE is high, this is called underfitting. And model becomes more complex, Test MSE is lower, but model becomes too much complex, Test MSE becomes higher again this is called overfitting.

Q3: Suppose we are creating 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. Why might this affect classification accuracy?

Given the situation, describe *two* methods to improve classification accuracy, and explain why they would help.

A3: Because of hard assignment when distance between discripter and two or more cluster centroids are similar, little bit more closer cluster centroid are choosed to representation of the feature. We could use soft assignment methd.(every point are assigned to cluster number with different weight or probabilities about each cluster)

Example of clustering using soft assignment

Weighted K-means : For each point x, there are weight value w(k) about cluster k. This method allow each data not hardly belongs to single cluster.

GMM - EM : When do clustering, use Gaussian mixture model and EM algoritm. EM is composed of E-step and M-step using hidden variable. There is assumption about gaussian probability for each cluster and hidden variable that is more compotable. Using hidden variable and E-step, M-step than observation set's MLE about hidden variable and cluster number is maximize. Therefore, there is low probablity about equidistant situation because of using Gaussian model.

Q4: 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?

A4: Your answer here.

There are advantage about less effect by position and orientation of object in image, so could have good quality of classifier result. But there is disadvantage about pooring at localizing objects within image.

For better model we can use more complex deep-learning model ex) CNN, Auto-encoder etc.