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: As for data which aren't linearly separable, linear classifiers return negative results. As using the fact that linear regression could be made more powerful using a basis function, or feature representation, we can apply it equally. A function mapping from input vectors to feature vectors could easily solve the problem as well. But the only problem is that where can we calculate the features from, which the kernel trick could help in such cases. Not calculating the lifting transformation, but calculating the kernel function, which is the multiple of two features. If the kernel function satisfies Mercer's condition, we can use this 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: Bias is the difference between average prediction and the correct answer which is intended. Variance is the scale what the estimate is varied. When the model is too simple which it couldn't represent all of the class characteristics, we call it as underfitting. It is majorly found at high bias, low variance situations. It contains both high training errors and test errors. On the other hand, if the model is too complicated, which irrelevant noises are found easily, we call it overfitting. It is usually found at low bias and high variance situations. It contains low training errors and high test errors.

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: BOW (Bag of Words) model discards large-scale spatial locations and its relative locations. As an advantage, strong sparsity of term vectors allows for efficient indexing schemes and other performance improvements. The example for such advantage could be various applications in improve of performance, including which used to overcome upcoming disadvantage as well. As for its disadvantage, by ignoring spatial relationships inside the image itself, correlation between features couldn't be described without any extensions. For improvements, new models and layers are introduced to capture spatial features within the image, or use spatial pyramid matches to enhance spatial matching features.

To evaluate whether BOW model is a good model or not, we can focus on its high potential on expansion and simplicity on the basic concept. It would address that the basic model wouldn't be good enough, but is good enough when sufficient extensions are applied.