

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: When we use linear classifier, we can see data that are not linearly separable. Then, we can separate them in a way that increases the dimension. A prime example is kernel trick.

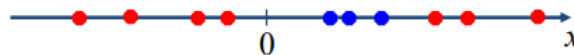


Figure 1: before kernel

In this case, we can separate them.

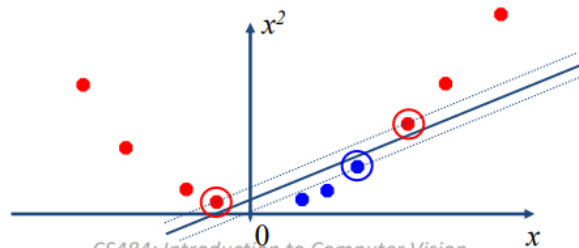


Figure 2: after kernel

After using kernel, we can separate them.

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 the expected prediction of our model and the correct value. And variance is the amount that the estimate of the target function will change if different training data was used. Underfitting is when model is too "simple" to represent all the relevant class characteristics. When we do underfitting, we get high bias and low variance like left figure. It also means high training error and high test error. Overfitting is when model is too "complex" and fits irrelevant characteristic (like noise) in the data. When we do overfitting, we get low bias but high variance like right figure. It also means low training error and high test error.

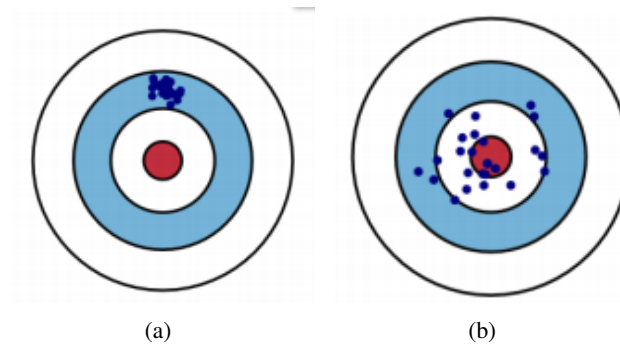


Figure 3: Underfitting and Overfitting

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: Your answer here.

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.