
Solutions

All problems are worth 2 points.

1. What is one real-world application that last week's guest lecturer's pattern recognition technology has enabled or can enable?

There was a wide range of viable responses here, including:

- (a) Improved handwriting recognition
 - (b) Computer-assisted handwriting instruction for children
 - (c) Signature recognition
 - (d) Detection of neurological disorders
2. Given the general linear discriminant function $g(x) = w^T \phi(\mathbf{x}) + w_0$, with weight vector w , bias term w_0 and input features \mathbf{x} , what is the purpose of the ϕ function?

The ϕ function is used (optionally) to transform the input feature space into some other space, where (hopefully) the classes will be linearly separable, or at least more separable than they were in the original space.

3. What is an ROC curve?

*An ROC (or Receiver Operating Characteristic) curve displays the performance of a binary classifier by plotting the true positive rate against the false positive rate **at a range of classification thresholds**.*

4. Why is it problematic for a classifier to have a high bias? (Note that I am not referring to bias *parameters* as in problem 1.)

A classifier with a high bias—that is, one with a decision boundary that is very strongly fitted to the precise boundaries observed in the training data—will tend to have poor generalization performance when applied to test data, because it will have overfit to idiosyncrasies and random perturbations in the training data that likely do not reflect the true class membership probabilities throughout the feature space.

5. How are individual trees in a random forest different from traditional decision trees?

*Each tree in a random forest is trained in a semi-random fashion. Rather than seeking a globally optimal tree, random forest trees sample **random subsets of possible splitting rules** (at each node), and pick the best rules from among those subsets. Additionally, each tree is trained on a **random resampling of the original data**, in which some points may appear multiple times while others do not appear at all.*