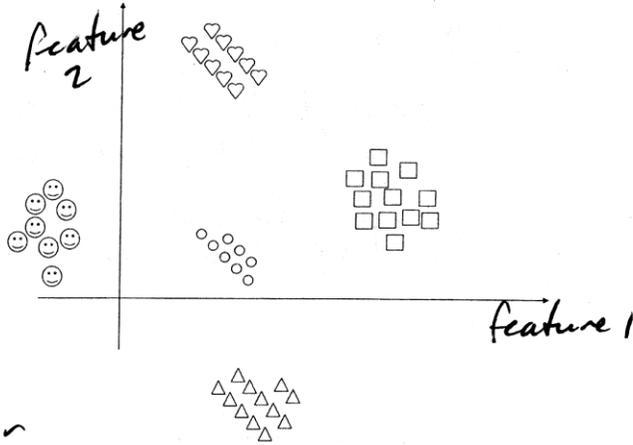


Name: Key

EE435 Spring 2012 Quiz 3

Given the scatterplot to the right, answer the questions that follow:



1. How many features are represented?

2

2. How many classes are represented?

5

3. What does each individual shape represent?

a sample from a particular class

4. Would you say that based on this scatterplot, the classification problem is probably easy or hard? Why?

Looks relatively easy - there is space between each class, so that a boundary could be drawn and no errors made (at least with this data).

5. Assume that the axes cross at the origin (0,0). Name the easiest way to differentiate between smiley faces and the other shapes. Name the easiest way to differentiate between the triangles and the other shapes.

- check the feature 1 value - if < 0 , must be smiley face.
- check the feature 2 value - if < 0 , must be triangle

6. In terms of pattern recognition, what does it mean if in the scatter plot above that there were some smiley faces among the triangles, and some hearts among the squares? That is, how would it affect the design of a pattern recognition system?

- in this case, expect that any decision boundary will cause recognition errors.
- It will be more challenging to determine the best decision boundaries.

The questions on the reverse side do not relate to the scatterplot above.

7. Name 2 measurable features that might be useful in differentiating Dachshunds from Great Danes.

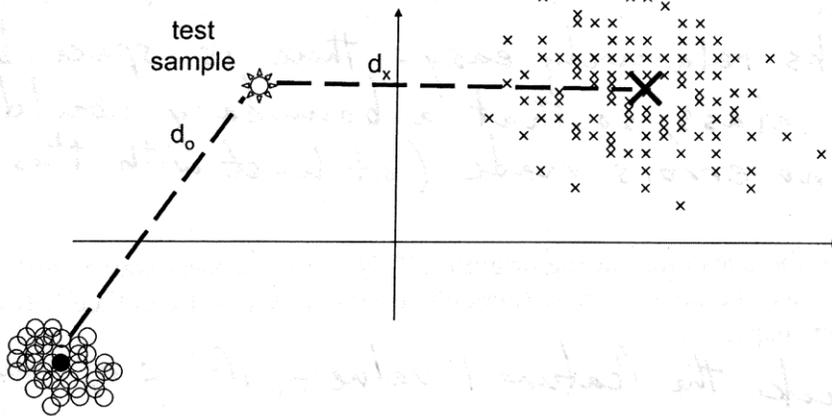
e.g. length, width, weight

8. Give a good reason why using the above two features, that there still may be errors in differentiating Dachshunds from Great Danes.

- Expect Dachshunds to be narrow width, but longer in length, for adult dogs. If all ages are considered, you may have Great Danes that look like dachshunds, and vice versa.

9. A pattern recognition system must decide if a sample is an "x" or an "o". In the scatterplot below, d_o and d_x are Euclidean distances from a test sample, and are equal ($d_o = d_x$). The large X and the filled in circle are the models for each. Should the pattern recognition system decide that the test sample vector is an "x" or an "o"? Why?

if $d_x = d_o$, it doesn't matter what class the sample is assigned to, just be consistent.



10. In the system of problem 9, if a weighted Euclidean distance is used instead of a Euclidean distance, should the pattern recognition system decide that the test vector is an "x" or an "o"? Why?

Why?

- Now the spread of the classes must be considered.
- The sample should be classified as x, because the o's are tightly packed and it is less likely to be an o

Bonus: What is the capital of Cambodia?

Phnom Penh