Image Based Recognition and Classification - DD2427

Exercise 5

Jim Holmström - 890503-7571

April 10, 2012

Exercise 1. Show that the decision boundary induced by $1NN$ for the two feature vectors $\bar{x}_1, \bar{x}_2$ is a straight line.

Solution. Decision boundary is defined as: $|\bar{x}_1 - \bar{x}| = |\bar{x}_2 - \bar{x}|$ since neither side can be negative we can simple square both side and expand the L_2-norm which gives us:

$$\sum (x_{1,i} - x_i)^2 = \sum (x_{2,i} - x_i)^2$$
$$\sum x_{1,i}^2 - 2x_{1,i}x_i + x_i^2 = \sum x_{2,i}^2 - 2x_{2,i}x_i + x_i^2$$
$$\sum x_{1,i}^2 - x_{2,i}^2 + 2(x_{2,i} - x_{1,i})x_i + 0 = \sum 0$$
$$2\sum (x_{2,i} - x_{1,i})x_i + \sum x_{1,i}^2 - x_{2,i}^2 = 0$$
$$\sum (x_{2,i} - x_{1,i})x_i + \frac{\sum x_{1,i}^2 - x_{2,i}^2}{2} = 0$$
$$\sum c_i x_i + d = 0$$

With c_i, d is all constants given $\bar{x}_1, \bar{x}_2$, this is the form for multidimensional linear equation with an $(n-1)$-hyperplane as solution. $\square$

Exercise 2. Draw decision boundaries for $1NN$-classifier with the training sets:
Class1: $\{(7,11),(15,9),(15,7),(13,5),(14,4),(9,3),(11,3)\}$
Class2: $\{(11,11),(13,11),(8,10),(9,9),(7,7),(7,5),(15,3)\}$

Solution. See (1) at page 3

Exercise 3. Why might using a too large value of k be bad for this data set? Why might using a too small value be bad for this data set?

Solution. Generally a large k will reduce noise in the classification in our case $5NN$ would result in an linear classifier between the 2 camps and in reality the two oddballs at each camp wont be taken in to consideration and this might be a good or bad thing depending on if the oddballs are noise or relevant data. In the same way but inverted holds for too small values of k say for example $2NN$ we would have an island behind the camp which might be relevant or just missclassification due to the noise (the oddballs).
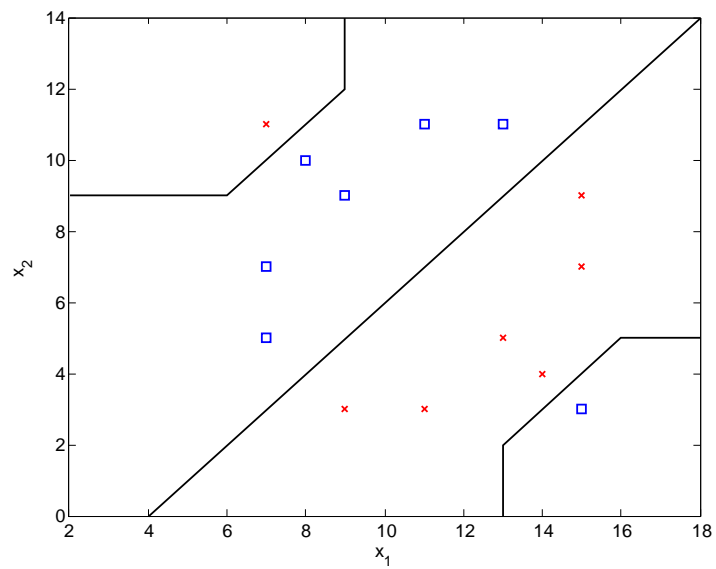
Exercise 4.

Solution.

Figure 1: The decision boundary for exercise2