MLPR: Lab 1

Maarten Inja Patrick de Kok

September 11, 2012

Exercise 1

Our training and test data sets are generated through the Matlab code presented in Listing 1. A graphical representation is given in Figure 1.

Listing 1: The code which generates our data set

```
function [training, labels, test_A, test_B] = preprocess_data()
load("twoclass.mat");
training = [A(1:750,:); B(1:750,:)];
labels = [ones(750,1), zeros(750,1); zeros(750,1), ones(750,1)];
test_A = A(751:1000,:);
test_B = B(751:1000,:);

clf;
hold on;
plot(training(1:750,1), training(1:750,2), "@+1");
plot(training(751:1500,1), training(751:1500,2), "@x3");
plot(test_A(:,1), test_A(:,2), "@o4");
plot(test_B(:,1), test_B(:,2), "@*0");
hold off;
```

Exercise 2

Confusion matrices of the KNN with $1 \leq k \leq 20$:

Exercise 3

Table 1: TA = True A; TB = True B; CA = classified as A; CB = classified as B.

k		CA	CB
1	TA	199	51
	TB	34	216
2	TA	205	45
	TB	34	216
3	TA	212	38
	TB	33	217
4	TA	217	33
	TB	35	215
5	TA	214	36
	TB	30	220
6	TA	217	33
	TB	31	219
7	TA	216	34
	TB	31	219
8	TA	221	29
	TB	32	218
9	TA	216	34
	TB	30	220
10	TA	217	33
	TB	32	218
11	TA	214	36
	TB	29	221
12	TA	215	35
	TB	28	222
13	TA	215	35
	TB	31	219
14	TA	216	34
	TB	31	219
15	TA	216	34
	TB	31	219
16	TA	213	37
	TB	31	219
17	TA	215	35
	TB	35	215
18	TA	215	35
	TB	34	216
19	TA	215	35
	TB	36	214
20	TA	216	34
	TB	33	217

Figure 1: A plot of the dataset. The red "+" and magenta "○" represent the training and test data points of class A. The blue "×" and black "∗" represent the training and test data points of class B.