Machine Learning Assignment 2 – Alexandros Sivris (03627456)

Exercise 1

The Learned GMM parameters for the 4 clusters are:

	1	2	3	4
π	0.201826972008895	0.261616498294994	0.240132859507759	0.296423670188352
μ	[-0.0146 -0.0796]	[0.0262 0.0617]	[-0.0432 0.0446]	[-0.0432 0.0446]
Σ	1.0e-03 * 0.3961 0.2181 0.2181 0.1287	0.0011 -0.0004 -0.0004 0.0002	1.0e-03 * 0.1749 0.2617 0.2617 0.3978	1.0e-03 * 0.7426 -0.5897 -0.5897 0.6073

Exercise 2

Of the test set **55** sequences are classified as train sequences and **5** as test sequences.

Exercise 3

(a) ***WalkPolicyLearning***

1.) The reward matrix is the following:

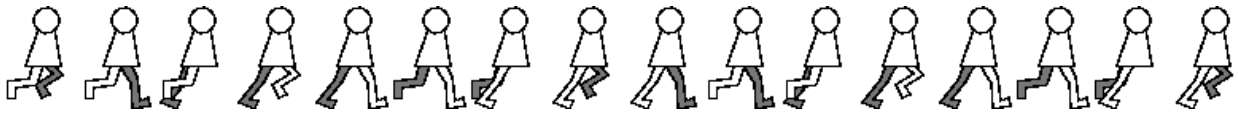
0	-1	0	-1
0	0	-1	-1
0	0	-1	-1
-1	-1	0	-1
-1	-1	-1	0
0	0	0	0
0	0	0	0
0	1	0	0
-1	-1	0	-1
0	0	0	0
0	0	0	0
0	1	0	-1
0	-1	0	-1
-1	0	0	1
-1	-1	0	1
0	-1	0	-1

2.) The value of γ is **0.75**. As γ approaches 0 only immediate rewards are considered. As it grows towards 1 future rewards will be considered with higher weight, willing to delay the reward (Source: http://mnemstudio.org/path-finding-q-learning-tutorial.htm)

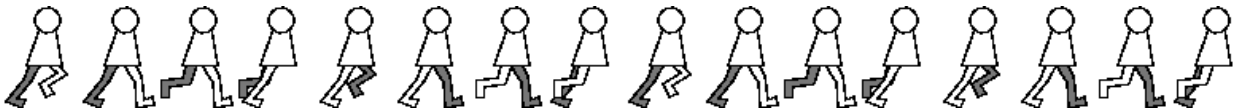
3.) The program needs about 5 to 6 iterations (varies with initial state).

4.)

Starting state 10:



Starting state 3:



(b) ***WalkQLearning***

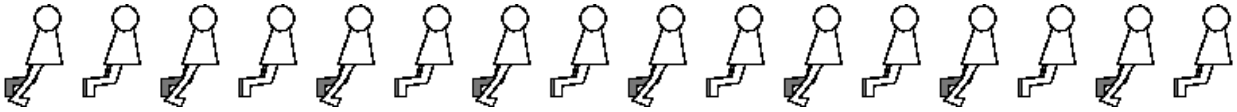
1.) ε=0.01, α=0.1

2.) In a greedy-only approach the convergence happens pretty slow.

3.) The Q matrix changes very little after about 30.000 steps

4.)

Starting State 5:



Starting State 12: