Q-learning and K-means

Course: Machine Learning
Instructor: Dr. Mirela Popa

Student name: Parand Mohri Student ID: i6234307

Handing in
Upload a single report in form of a PDF. E.g. make a scan. Hand in code in form of a single zip file. Submissions by email or other types of archives are not accepted. Thank you for your understanding.

For the first part (a) include in the report a short description of your result, the best policy and your interpretation of the role of the two parameters alpha and gamma. For the second part (b) include the required explanations.

Filling in
You can use this Word file to answer your questions in a digital form. Alternatively, you can print the document, fill it in, and upload a scan. Make sure that we can read your hand-writing.

Graded: Code and Paper assignment: Q-learning

Your task is to implement the SARSA algorithm for a simple single player game, in which an agent explores the environment, collects rewards and eventually arrives in the destination state, finishing the game (e.g. snake game, PacMan). Your goal is to maximize the final score (which is obtained by arriving in the shortest time to the destination state), while also exploring the environment. The grid is 4x4 and the set of valid actions are move up, down, right, left, except for the boundary walls, where only specific actions are possible. All the other values are currently initialized, but you can adjust them as you consider. A part of the code is provided for you in Canvas (tutorial6.ipynb); your task is to complete the missing steps, including the update of the value function.

The algorithm is the following:

For each s, a initialize the state Q(s,a) to zero

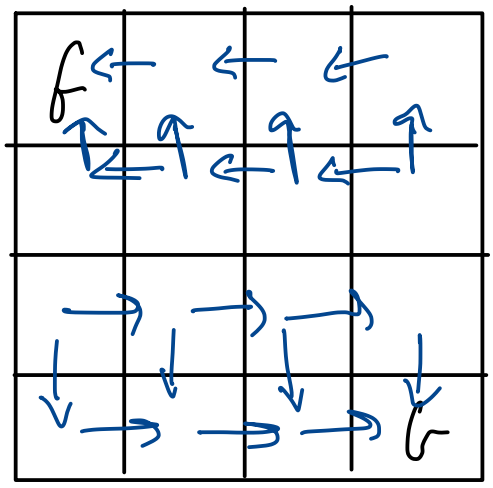
Start from a random state s

Do forever:

- Select an action a randomly and execute it
- Receive immediate reward r
- Observe the new state s'
- Update the table entry for Q(s,a) as follows
- $$Q(s,a) = (1-\alpha) \cdot Q(s,a) + \alpha \cdot (r + \gamma Q(s',a'))$$
- Make the transition $s \leftarrow s'$
- If s' is the destination state then stop

Include in this report your observations about the process, the obtained Q matrix and your interpretation about the role of the two parameters alpha and gamma and how do they affect the final policy.

Best policy for maximizing the score (include it as a matrix/drawing)



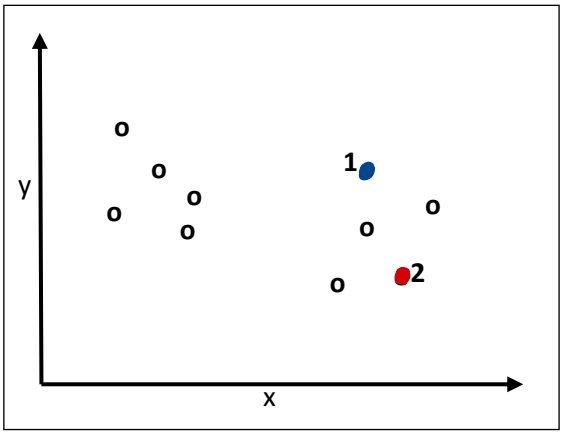
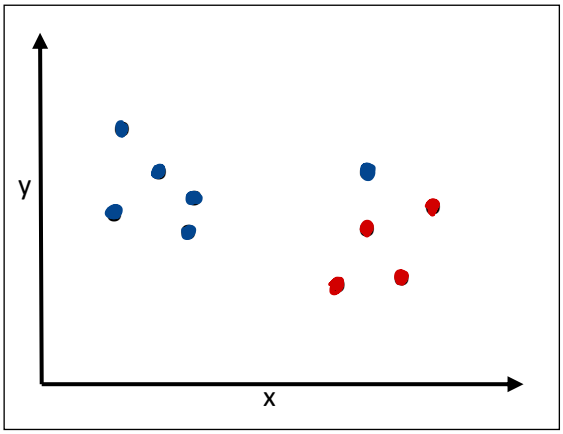
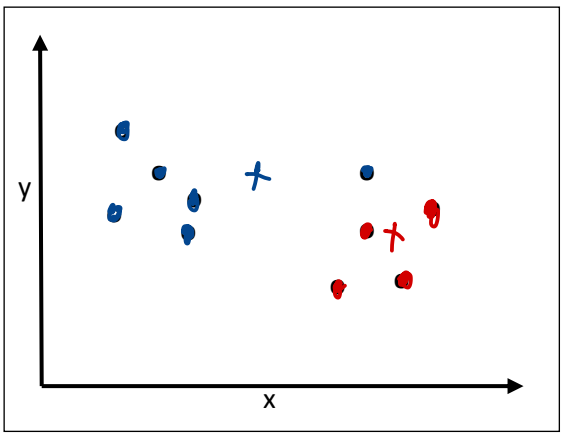
Explanation of the role of the parameters:

bigger gamma we get bigger value and that is correspond to reward and with each squar closer to find goal we get really large difference → Bacuse we are delying the reward

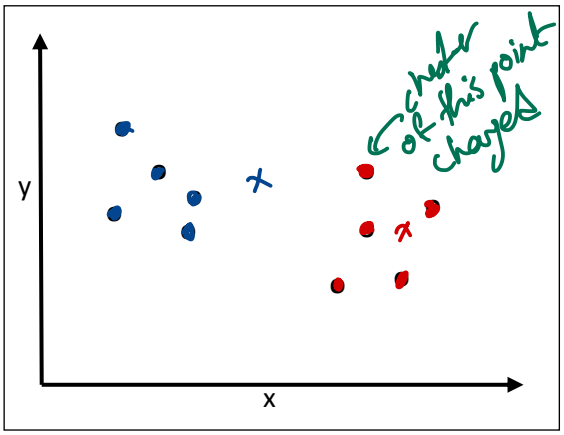
with big alpha agent is making more diviation from its privius experience rather than the right answer so if its really big we get wrong values sometimes.

Graded: Paper assignment: K-Means

Given the following data set, show (with drawings) and explain (with your own words) the different steps of a k-means algorithm when k=2. Show and explain individual steps of the algorithm – not just full iterations.
(Explanation of symbols: o = data points; 1 = marker for first centroid, 2 = marker second centroid)

Step 1 (not iteration!): 	Explanation: *Initialization: Centroids get assigned to random locations. Here two random points from the data set are picked as initial seeds.*
Step 2 (not iteration!): 	Explanation: then we check the distance of all data points with these two centroid and assign a cluster to them depending on which of these two point they are closer too
Step 3 (not iteration!): 	Explanation: then for each cluster we find the average of that cluster by summing the value of data points and devide them by how many are they then we have two new points that we put as cetroid points

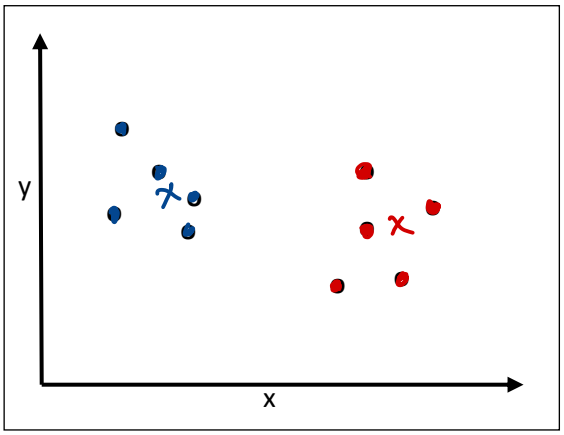
Step 4 (not iteration!):



Explanation:

we again calculate the distance of each point with these new centroids and re cluster the points

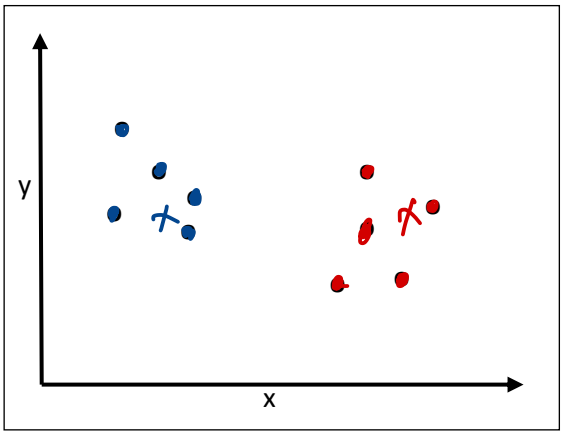
Step 5 (not iteration!):



Explanation:

again we find the new centroid by finding the mean.

Step 6 (not iteration!):



Explanation:

continue till it converge