EXAM TEST 2 2021/22

Question 1

1). We have an MDP = $\langle X, A, \delta, r \rangle$ with:

$X = \{Side_1, Side_2\}$ Set of states that the agent can be in;

$A = \{pass\_bridge_1, pass\_bridge_2, pass\_bridge_3\}$ Set of actions.

$\delta: X \times A \longrightarrow X$ transition function

$r: X \times A \longrightarrow \mathbb{R}$ reward function

2). We want to find an optimal policy function, a policy function is optimal when maximizes the expected cumulative reward. We can use Q-learning:

foreach x, initialize $\hat{Q}(x,a) \leftarrow 0$
we observe the current state
foreach time $t = 1 \dots T$ do:
choose an action a
execute a
observe the new state x'
collect the reward $\bar{r}$
$\hat{Q}_{(t)}(x,a) \leftarrow \bar{r} + \gamma \max_{a' \in A} \hat{Q}_{(t-1)}(x',a')$
$x \leftarrow x'$

$\hat{\pi}^*(x) = \underset{a \in A}{\operatorname{argmax}}\, Q(x,a)$

3). We can choose the ε-greedy policy. We select exploration with probability ε and exploitation with probability $1-\varepsilon$.

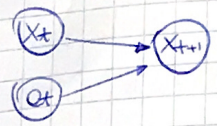
Question 2

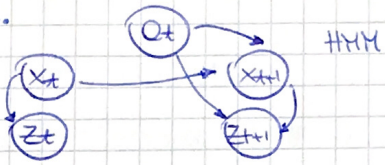
The Markov property says that:

- Once the current state is known the evolution of our dynamic system does not depend on previous states, actions, observations.
- The current state contains all the informations needed to predict the future.
- Future states are conditionally independent from past states and observations given the current state.
- Given the current state past, present and future observations are statistically independent.

The difference between MDPs and HMMs is the property of full observability. In MDPs states are fully observable, also if we have non-deterministic actions after the execution of the action we can see the resulting state.

In HMMs states are not observable.





Is possible to see how in the model of the MDP the new state depends only on the previous state and action while in HMM the ~~new~~ action influences also the future observation.

Question 3

1). $\dim(W_1) = 50 \times 100 = 5000$ $\quad \dim(W_2) = 100 \times 10 = 1000$

2). $\#\text{params} = 5000 + 1000 = 6000$

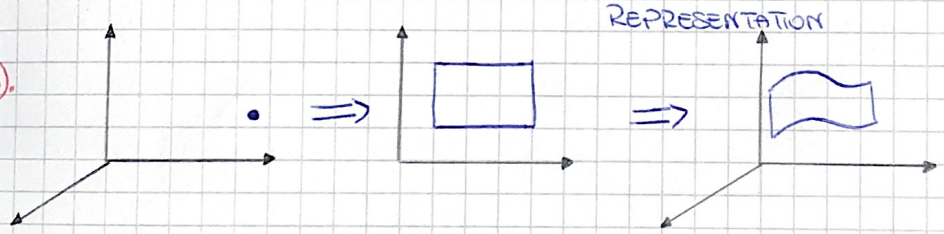
3). $h = g(W^T x + c)$ with $g(a) = \max(0, a)$ ReLU function

$y(x, w) = w^T \max(0, W^T x + c) + b$

Question 4

1). An autoencoder is a combination of 2 neural networks: an encoder and a decoder. The training is based on reconstruction loss and we can see a low dimensional representation. In an autoencoder we have hidden layers with reduced size that are called bottlenecks. Autoencoders use the same sample from the dataset $\{x_n\}$ as input and output.

3).

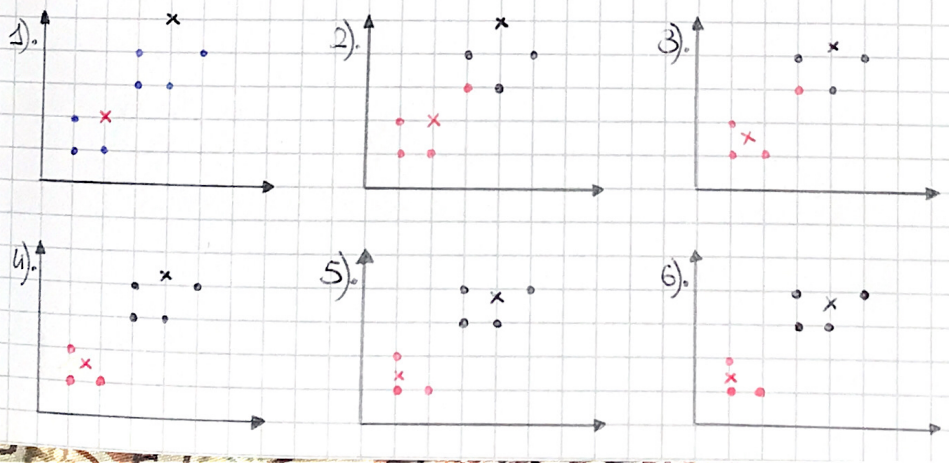


Input: 3D

Hidden Layer: 2D

Output: 3D

Question 5



In the second plot i assign each cluster to the nearest centroid. In the third i recompute the centroids of the clusters. In the plot 4 i recompute the assignmen and is possible to see that one sample is switched. In the plot 5 i recompute the centroids of the clusters and in plot 6 i have the final solution.

Question 6

1). Bagging is an ensemble method that uses different learners in sequence.

In bagging we have different steps:

- We split our dataset D into M bootstraped datasets $D_1 \dots D_M$.
- I train each model $y_i(x)$ using the bootstraped dataset D_i $\forall i = 1 \dots M$
- $y_{BAGGING}(x) = \frac{1}{M} \sum_{n=1}^{M} y_k(x)$

2). I have a dataset D and i split this dataset into 4 bootstraped datasets D_1, D_2, D_3, D_4.

I train each classifier $y_i(x)$ using $D_i(x)$ $\forall i = 1 \dots 4$

$y_{BAGGING}(x) = \frac{1}{4} \sum_{k=1}^{4} y_k(x)$