EXERCISE 1

1. Provide a formal definition of the Reinforcement Learning (RL) problem. Describe formally what are the inputs and the outputs of a RL algorithm.

2. Describe the main steps of a RL algorithm. Provide an abstract pseudo-code of a generic algorithm for RL (e.g., Q-learning).

In reinforcement learning, we have to learn the policy for a MDP, the input is in the form $\{(x_i, a_i, r_i)\}_{i=1}^{N}$ where x_i is a state, a_i the action that led to x_i, and r_i the given reward. The output should be a policy function $\pi : X \to A$ that maximize the cumulative reward. The main steps are:

```
For t=1...T {
    choose a∈A      ← ε greedy → with prob ε choose a random action.
                                  else choose the best action
    compute r
    update a data structure Θ      ε decreases over time
}
choose π by reasoning about Θ
```

EXERCISE 2

Describe two different methods to overcome overfitting in Convolutional Neural Networks (CNN).

One method is *dropout*: with some probability we ignore portions of the network. Another metodh is to add a reg. term in the loss function like: $\lambda \sum_{i=1}^{l} |W^{(i)}|$ with $\lambda \in \mathbb{R}^+$.

EXERCISE 3

1. Describe the principle of maximal margin used by SVM classifiers. Illustrate the concept with a geometric example.

2. Draw a linearly separable dataset for binary classification of 2D samples. Draw two solutions (i.e., two separation lines): one corresponding to the maximum margin, the other one can be any other solution.

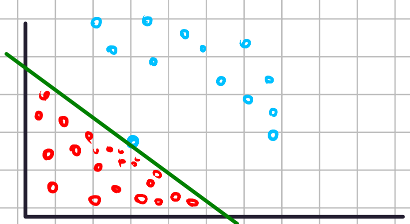
3. Discuss why the maximum margin solution is preferred for the classification problem.

We consider as margin, the distance between the linear discriminant and the closest sample to it: $\min_{x_i \in D} \frac{|y(x_i)|}{\|W\|}$. The SVM aims to find the w that maximize such margin, while classificating correctly all the samples:
$D = \{(x_i, t_i)\}_{i=1}^{N}$, $t_i \in \{+1, -1\} \Rightarrow y(x_i)t_i \geq 0 \ \forall i$. We can add a relaxation in the optimization problem:

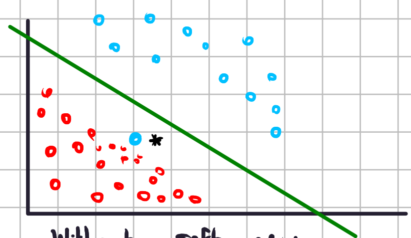
$$\begin{cases} w^* = \arg\max_{w} \left(\min_{x_i \in D} \frac{|y(x_i)|}{\|W\|} \right) + C \sum_{i=1}^{N} \xi_i^+ \\ \text{s.t.} \quad t_i\, y(x_i) \geq 0 \quad \forall i \\ \quad\quad \xi_i \geq 0 \quad \forall i \end{cases}$$

$\Rightarrow$ we allow some samples to overcome the margin

$$\begin{cases} \xi_i = 0 & \text{in his side} \\ \xi_i \in [0,1] & \text{violates margin} \\ \xi_i > 1 & \text{missclassified} \end{cases}$$



With soft margin



Without soft margin

* may be a very noisy sample.

EXERCISE 4

1. Provide the definition of *Confusion matrix* for a multi-class classification problem.
2. Provide a numerical example of a confusion matrix for a 3-classes classification problem with a balanced data set including 100 samples for each class. Show the confusion matrix in two formats: with absolute values and with the corresponding percentage values.
3. Compute the accuracy of the classifier for the numerical example provided above.

Hint: use simple numerical values, so that you do not need to make complex calculations.

The conf. matrix is the matrix C where C_{ij} is the number of samples of class i classified as j, the accuracy is $\mathrm{trace}(C) \cdot \left(\sum_{i=1}^{K}\sum_{j=1}^{K} C_{ij}\right)^{-1}$.

Example

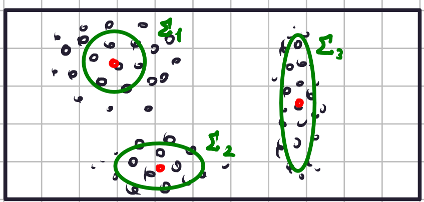
$$C = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 20 & 3 & 2 \\ 2 & 8 & 21 & 2 \\ 3 & 5 & 9 & 30 \end{array} \Rightarrow \text{accuracy} = \frac{71}{100} = 71\ \%$$

EXERCISE 5

Given an unsupervised dataset $D = \{\mathbf{x}_n\}$

1. Define the Gassian Mixture Model (GMM) and describe the parameters of the model.
2. Draw an example of a 2D data set (i.e., $D \subset \Re^2$) generated by a GMM with $K = 3$, qualitatively showing in the picture also the parameters of the model.
3. Determine the size of the model (i.e., number of independent parameters) for the dataset illustrated above.

In this model we assume that we have K classes and the dataset is generated by K normal distr. such that: $\mathbb{P}(x \mid C_i) = \mathcal{N}(x; \mu_i, \Sigma_i)$ and $\mathbb{P}(C_i) = p_i$. The parameters are: μ_i, Σ_i, p_i $1 \le i \le K$. Notice how only $K-1$ priori probabilities p_i are free since $\sum_{i=1}^{K} p_i = 1$. An example for $K=3$ is:



the FREE parameters are:

$p_1, p_2 \in \mathbb{R}$

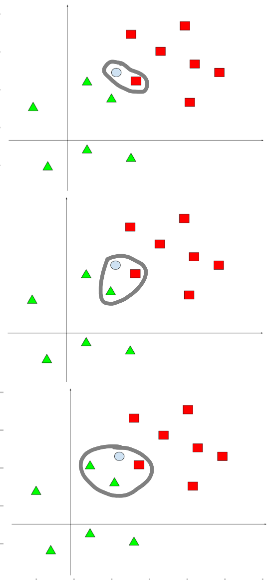
$\mu_1, \mu_2, \mu_3 \in \mathbb{R}^2$ $\Rightarrow$ total size: 20

$\Sigma_1, \Sigma_2, \Sigma_3 \in \mathrm{Mat}(2\times 2)$

EXERCISE 6

1. Describe the K-nearest neighbors (K-NN) algorithm for classification.
2. Given the dataset below for the two classes $\{square, triangle\}$, determine the answers of K-NN for the query point indicated with symbol o for K=1, K=3, and K=5. Motivate your answer, showing (with a graphical drawing) which instances contribute to the solution.

```
KNN (K, D, x) {
    D_K = { }
    For x_i ∈ D {
        if |D| < K {
            D.append(x_i)
        }
        else {
            if ∃ x_j ∈ D_K : ||x - x_j|| > ||x - x_i|| {
                substitute x_i with x_j in D_K
            }
        }
    }
    return  c* = argmax_c  Σ_{x_i ∈ D_K} δ(x_i is in class c)
}
```



K=1
output: □

K=2
output: Δ

K=3
output: Δ