Exercise 1  Regularized logistic regression

We consider the following training data:

<table>
<thead>
<tr>
<th>Index</th>
<th>( \mathbf{x}_i )</th>
<th>( y_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>([1, 1]^\top)</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>([2, 1]^\top)</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>([4, 3]^\top)</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>([4, 4]^\top)</td>
<td>1</td>
</tr>
</tbody>
</table>

1. Plot the training data in 2D, with red dots for class 1 and blue crosses for class \(-1\).
2. Our goal is to minimize the following cost function:

\[
J(\alpha) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \mathbf{x}_i^\top \alpha)) + \frac{\lambda}{2} \| \mathbf{w} \|^2 .
\]

Compute the gradient of this function with respect to \( \alpha \). You can split the computation in two parts: with respect to \( b \) and with respect to \( \mathbf{w} \).
3. Recall the gradient descent algorithm. Write the algorithm for the minimization of \( J(\alpha) \).
4. Compute the first 2 iterations of this algorithm for \( \alpha = [-0.5, 0, 0]^\top \), \( \mu = 0.1 \) and \( \lambda = 1 \). Compute for each iteration the cost \( J(\alpha) \) ad the corresponding \( \alpha \).
5. Compute the gradient norm after the second iteration.
6. After 500 iterations, we obtain a vector \( \alpha = [-3.55, 0.71, 0.71]^\top \). Compute the gradient \( \nabla_{\alpha} J(\alpha) \) at this point and its norm. Can this vector be considered as a solution of our problem?
7. Add to the plot of Question 1 the decision boundary for the solution obtained after 2 iterations, and then for the \( \alpha \) obtained after 500 iterations.

Exercise 2  Perceptron

In this exercise, we will work on the same data as in Exercise 1.
1. Plot the training data.
2. Recall the perceptron algorithm.
3. Compute the first 4 steps of this algorithm with \( \alpha = [-3, 0, 0]^\top \), \( \mu = 0.5 \), and parsing the examples in sequential order \((1 \to 2 \to 3 \to 4)\).
4. Compute the first 4 iterations of this algorithm with the same parameters, but this time parsing the examples in the order \( 3 \to 4 \to 2 \to 1 \).
5. Add to the plot of Question 1 the decision boundary obtained for both solutions.