Statistical inference and machine learning
Midterm project

- This project must be solved individually.
- Deadline: Monday 16 November 2020, 23:59 (No late submissions will be accepted)
- Upload a single zip file on Moodle containing your solution and your code implementations. You can use any programming language, but it is recommended to use MATLAB for the part 1 as you are given some functions that can help you.

1 Using SVM for Spam Classification

In the first half of this project, you will be using support vector machines (SVMs) with two examples. Experimenting with these datasets will help you gain an intuition of how SVMs work and how to use a Gaussian kernel with SVMs. In the next half, you will be using support vector machines to build a spam classifier.
For SVM hypothesis we use the following notation (linear kernel):

$$\min_{\theta} C \sum_{i=1}^{m} [y^{(i)} \text{cost}_1(\theta^T x^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T x^{(i)})] + \frac{1}{2} \sum_{j=1}^{n} \theta_j^2.$$

If we use non-linear kernels then we use this form:

$$\min_{\theta} C \sum_{i=1}^{m} [y^{(i)} \text{cost}_1(\theta^T f^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T f^{(i)})] + \frac{1}{2} \sum_{j=1}^{m} \theta_j^2$$

where $f^{(i)}$ is the feature vector, and for $1 \leq j \leq m$ we have

$$f_j^{(i)} = \text{similarity}(x^{(i)}, x^{(j)}).$$

The function "similarity" depends on the choice of kernel.

1.1 SVM with Linear kernels

We begin with a 2D example dataset (data1.mat) which can be separated by a linear boundary. (Figure 1)