Supervised Learning Lab

Aim
The aim of this lab is to learn the concepts and the techniques for supervised learning in particular Linear Regression, K-Nearest Neigbour (KNN), and Support Vector Machines (SVM). This lab is divided into three parts.

To pass the lab, write down your answers to the questions in Part 1, 2 and 3. During the lab demonstration, explain your answers to the lab assistant. The lab assistant will ask you some general question about the parts of the lab and how you solved them.

Part 1
Read the tutorial Linear Regression available on the course website. Execute yourself the Python scripts of the document and modify the parameters of the algorithm to understand better what happens. To pass this lab you need to answer the questions below. Write down your answer and discuss them with the lab assistant.

1. When can you use linear regression?
2. How can you generalize linear regression models to account for more complex relationships among the data?
3. What are the basis functions?
4. How many basis functions can you use in the same regression model?
5. Can overfitting be a problem? And if so, what can you do about it?

Part 2
Read the tutorial K-Nearest Neighbor (KNN) available on the course website. Execute yourself the Python scripts of the document and modify the parameters of the algorithm to understand better what happens. To pass this lab you need to answer the questions below. Write down your answer and discuss them with the lab assistant.

1. Why choosing a good value for k is important in KNN?
2. How can you decide a good value for k?
3. Can you use KNN to classify non-linearly separable data?
4. Is KNN sensible to the number of features in the dataset?
5. Can you use KNN for a regression problem?
6. What are the Pros and Cons of KNN?

Part 3

To complete this part read the tutorial Understanding Support Vector Machines available on the course website, and test the Python code of the examples. To pass this lab you need to answer the questions below. Write down your answer and discuss them with the lab assistant.

1. What is the basic idea/intuition of SVM?
2. What can you do if the dataset is not lenearly separable?
3. Explain the concept of Soften Margins
4. What are the pros and cons of SVM?