

Artificial Intelligence

(Introduction to ML)

Answer the following questions and write your answers in the predefined frames.

Student's Name:

Question 1:

(a) Which phases one has to go through to perform a machine learning task?

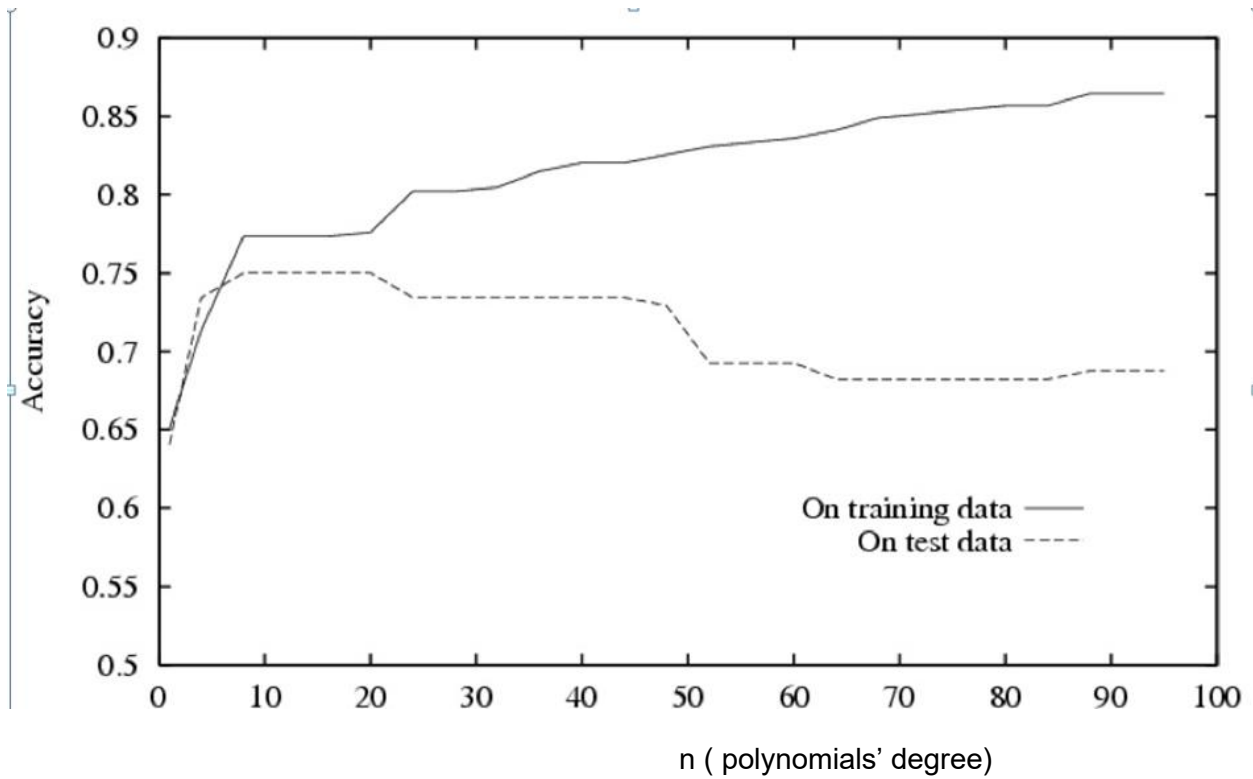
(b) Why we have separate test and validation sets?

(c) Give an example of the percentage splitting of the training set, with and without validation phase.

Question 2:

Consider the following graph, that depicts the training and test set accuracy during the linear regression learning, as a function of the polynomials' degree. This polynomial is used to learn a function $f: X \rightarrow Y$, where training and test set examples are drawn independently at random from

and underlying distribution $P(X)$, after which the trainer provides a noise free label Y . Note that $\text{error} = 1 - \text{accuracy}$.



1. Given the above plot of training and test accuracy, which polynomial's degree would you chose to classify future examples? Give a short justification.

2. What is the amount of overfitting in the polynomial you selected?

Question 3: Consider that you have to estimate the distribution of the grades of the students that attend this course and is given a limited list of students' grades. Let suppose that you have decided to model the distribution as a normal one. From your experience, you know that in most of the courses the mean grade is between 5.5 and 6.0 and the deviation is about 2. Which estimation method you will use and why?

Question 4:

Let suppose that for a classification problem you know that the data is not linearly separable. If the use of a non-linear classifier derives worse results than a linear estimator, what is the reason for this? What do you propose in order to alleviate this and get better results?

Question 5:

The following training set (examples) is given. Based on Information Gain, making calculations by hand, select the root of the corresponding decision tree as well as the feature of the next level left-hand node.

Example	<i>Comedy</i>	<i>Doctors</i>	<i>Lawyers</i>	<i>Guns</i>	<i>Likes</i>
e_1	false	true	false	false	false
e_2	true	false	true	false	true
e_3	false	false	true	true	true
e_4	false	false	true	false	false
e_5	false	false	false	true	false
e_6	true	false	false	true	false
e_7	true	false	false	false	true
e_8	false	true	true	true	true
e_9	false	true	true	false	false
e_{10}	true	true	true	false	true
e_{11}	true	true	false	true	false
e_{12}	false	false	false	false	false