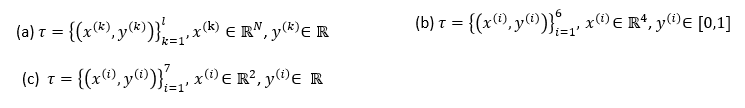
CSCI555:20 DMM

Homework-Graded

**Note:** Preferably, complete this as a team work with each team having at most three students. You can also complete the work individually.

**Problem 1.** Categorize the following training data to be either regression data or classification data. What would be the form of the regression models that would fit to the training data that are the regression data sets?



|  |  |  |  |
| --- | --- | --- | --- |
| (d) |  | (e) |  |

**Problem 2:** Refer to problem 1. Write down the datasets in cases (a), (b), and (c) in in tabular form (that is in the form used in (d) and (e)) whereas the datasets (d) and (c) in compact notation (that is in the form used in (a), (b), and (c)).

**Problem 3:** Assuming  to be the empirical loss, determine the gradient  of the linear regression models for the data that represent regression datasets in Problem 1. Remember so you have to determine the partial derivatives of the function with respect to the weights. Consult any calculus book or online resource to learn how to find partial derivatives of functions. There is also an example given in problem 3.

**Problem 3:** Find the gradient of the following functions at the specified points.

1. , at
2. , at
3. , at
4. , at

**Problem 4:** Refer to the dataset (e) in Problem 1. Assume that it is a regression dataset and you are fitting a linear regression model to the data. Define the empirical loss as . Find by hand and then write code in the form of a Python function that finds the value of when provided with any training example (also called data point or sample) from the dataset (e) in Problem 1. Assume .

**Problem 5:** Refer to the dataset in Problem 1(b). Assume the loss function to be as below.

Find by hand and then write code in the form of a Python function that finds the value of when provided with the following training. Assume .

**Problem 6:** Assume the gradient descent algorithm applied to the data . Assume, the algorithm starts with the weights , how it updates the weights for the following training examples assuming the first iteration of the training loop.

**Problem 7:** Go to the scikit learn datasets at

<https://scikit-learn.org/stable/api/sklearn.datasets.html>

to explore the pre-loaded datasets in scikit-learn. These datasets have loaders provided for them by the scikit-learn.org. You can recognize them by the fact that each of them has the prefix “load” in its name. Explore the various loaders and see what data sets do they load. Learn from the examples in the descriptions of the loaders as how to use them. Identify the individual datasets as regression data or classification data state what linear models (regression or classification) would fit to them.

**Problem 8:** Refer to the information in Problem 5, in the description provided in the link, learn the use of “**make\_regression**” and “**make\_classification**” to learn how to generate artificial regression and classification data.