Problem 1, 2, and 3 are fundamental questions. Please do on paper and submit in pdf format.

Problem 4 and 5 are implementing questions. Please work on notebook and submit in notebook file.

**Problem 1 – Gaussian Distribution** (15 pts)

Consider partitioning of the components ***x*** into three groups with a corresponding partitioning of the mean vector and of the covariance in the form of

Find an expression for the conditional distribution in which has been marginalized out.

**Problem 2 – Gaussian Process** (15 pts)

Consider a parametric model governed by the parameter vector ***w*** together with a data set of input values and a nonlinear feature mapping Suppose that the dependence of the error function on ***w*** takes the form

where is a monotonically increasing function. By writing ***w*** in the form

where for all *n*, show that the value of w that minimizes *J(****w****)* takes the form of a linear combination of the basis functions for

**Problem 3. Neural Network** (20 pts)

Consider classifying a single hidden layer neural network for the following data:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data |  |  |  |  |  |  |  |  |  |  |
|  | 4 | 4 | 1 | 2.5 | 4.9 | 1.9 | 3.5 | 0.5 | 2 | 4.5 |
|  | 2.9 | 4 | 2.5 | 1 | 4.5 | 1.9 | 4 | 1.5 | 2.1 | 2.5 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Hidden Layer | | | | | |
|  | 1 | 0 | 1 | 0 | 0 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Output Layer | | | | | | | | | | |
|  | 1 | 1 | -1 | -1 | 1 | -1 | 1 | -1 | -1 | 1 |

1. (14 pts) Implement a neural network using NumPy. Report the accuracy of the model.
2. (3 pts) Using Scikit-learn, implement a neural network classifier. Report the accuracy.
3. (3 pts) Classify the point using two models from above.

**Problem 4. KNN** (20 pts)

Download the yeast dataset from the link: <https://archive.ics.uci.edu/ml/datasets/Yeast>.

1. Implement a basic KNN model on the yeast dataset. The task is to predict the compartment in a cell that a yeast protein will localize to base on the properties of its sequence. Do not use Scikit-learn.
2. To optimize the results, test with Manhattan and Euclidean distance metrics.
3. Report the model's accuracy for both distance metrics with k values from 5 to 20.

**Problem 5. Gaussian Process** (20 pts)

Download the breast cancer data set from the UCI Machin Learning repository.

1. Construct Scikit-learn Gaussian Process models using basic, RBF, and Matern kernels.
2. Implement a 5-fold cross-validation algorithm to report the model and accuracy of each fold.
3. Report the result.