Homework Assignment # 4

Due: Wednesday, December 9, 2016, 11:59 p.m.
Total marks: 100

Question 1. [30 MARKS]

In this question, you will implement a radial basis function network. The main difference is simply that the data is transformed using radial basis functions before being passed to the algorithms.

Implement a radial basis transformation. Use Gaussian kernels, and try three different bandwidths. Compare the performance on SUSY with the reduced feature set (8 features), comparing standard logistic regression and logistic regression with this radial basis transformation.

Question 2. [70 MARKS]

Implement any three learning methods (e.g., that you implemented from the last assignment) and run them on a dataset or problem of your choice. Use your knowledge about model comparison to formally conclude which of the two algorithms is better. This includes proper training-test splits, statistical significance tests, and proper meta-parameter selection techniques (e.g., cross-validation). You can use statistical significance tests built-in to python (or other languages). Provide a precise conclusion of your experiment. You can use any code and implementation you previously completed, but you are still prohibited from using packages that implement the learning algorithm directly. You can now use optimization software, such as lbfgs in scipy.

Homework policies:

Your assignment will be submitted as a single pdf document and a zip file with code, on canvas. The questions must be typed; for example, in Latex, Microsoft Word, Lyx, etc. or must be written legibly and scanned. Images may be scanned and inserted into the document if it is too complicated to draw them properly. All code (if applicable) should be turned in when you submit your assignment. Use Matlab, Python, R, Java or C.

Policy for late submission assignments: Unless there are legitimate circumstances, late assignments will be accepted up to 5 days after the due date and graded using the following rule:

on time: your score 1

1 day late: your score 0.9

2 days late: your score 0.7

3 days late: your score 0.5

4 days late: your score 0.3

5 days late: your score 0.1

For example, this means that if you submit 3 days late and get 80 points for your answers, your total number of points will be $80 \times 0.5 = 40$ points.

All assignments are individual, except when collaboration is explicitly allowed. All the sources used for problem solution must be acknowledged, e.g. web sites, books, research papers, personal communication with people, etc. Academic honesty is taken seriously; for detailed information see Indiana University Code of Student Rights, Responsibilities, and Conduct.

Good luck!