CS 461 – Fall 2021
Program 3 – Neural Networks / Machine Learning

For the final program of the semester, we're going to use a neural network to see the effects of a few different variables on exam scores.

You have a data file[1] of 1000 cases, listing for each student:

- Gender (female/male)
- Race/ethnicity (coded as Group A/B/C/D/E)
- Parental education (some high school, finished high school, some college, associate's degree, bachelor's degree)
- Whether on standard or free/reduced lunches, an indicator of family poverty
- Completion of a test preparation program (completed / not completed)

And 3 outcome variables, each an integer score from 0-100:

- Math score
- Reading score
- Writing score

Your task is to build a neural network to carry out regression, predicting each student's score on each of the 3 outcomes.

- Use one-hot coding for input variables.
- Your project must use TensorFlow or Keras. TF is installed on all Flarsheim labs, or is a free download if you want to install it yourself. (If your laptop uses the integrated graphics on the motherboard and doesn't have a separate GPU, you will need to compile the code from source, and will probably be disappointed with performance.)
- Start by using a single Gaussian to approximate each; then try 2 Gaussians, 3 Gaussians, etc., until performance drops; then 'drop back' one to the number of Gaussians that performed the best. Use that for testing your validation data.
- Use a 70/15/15 split for Training/Test/Validation data.
- Write a short report describing your project. Include:
  - How you developed the network
  - What the final network configuration is—how many Gaussians, how many layers, how many neurons/layer, etc.
  - What validation strategy did you use, and why?
  - What are your final results? How well did your network do?
  - Any ideas on how your network could be improved; ideas for further exploration; or thoughts about what possible follow-up studies might be done.
  - There's no required format for the report, and no references are required other than for any resources you copied code from in writing your program.
- Submit your report and your source code, either in a zipfile or via a GitHub link.

1 https://www.kaggle.com/spscientist/students-performance-in-exams