Simon Yoon

ECE469 Artificial Intelligence

Professor Sable

The project demonstrates the training and testing of a neural network using backpropagation and is based on the pseudo-code from Figure 18.24 in the 3rd edition of our textbook. The dataset is from the UCI ML repository for predicting whether income exceeds $50K/yr based on census data.[1] Some features with missing values or ones that were sparse such as capital-gain/loss were removed. The final list of attributes are age, working status*, census weight, education*, education number, marital status*, occupation*, relationship*, race*, sex*, hours worked per week, and native country*.[2] Most of these elements were mapped to values between 0 and 1 with some exceptions. Some were normalized such as age which was divided by 100, hours worked per week is divided by the total number of hours in a week (168) and final census weight (fnlwgt) is the number of people the census believes the entry represents which was divided by two million. The filenames go as follows: income.init (initial neural network), incomes.test (test set), incomes.train (training set), NNincomes.1.100.trained (trained file), NNincomes.1.100.results (results) as per the file-naming of your examples. The network was trained with a learning rate of 0.1 over 100 epochs/iterations and resulted in an accuracy 82.5%. Via Cygwin, built using make, run via ./nn.exe.

[1] UCI Machine Learning Repository: Adult Data Set

[2] * = categorical else continuous attribute