Udacity Machine Learning Nanodegree 2020

Capstone Proposal

Childhood Autistic Spectrum Disorder Screening using Machine Learning

Mukkul Kurwa

March 2020

The project's **domain background** — the field of research where the project is derived is Healthcare. The early diagnosis of neurodevelopment disorders can improve treatment and significantly decrease the associated healthcare costs. In this project, I will use supervised learning to diagnose Autistic Spectrum Disorder (ASD) based on behavioral features and individual characteristics. More specifically, I will build and deploy a neural network using the Keras API.

A problem statement — The task is to diagnose Autistic Spectrum Disorder (ASD) based on behavioral features and individual characteristics. This is a classification problem where the model takes behavioral features and individual characteristics and detects whether the patient has Autistic Spectrum Disorder (ASD) or not.

The **datasets and inputs** — This project will use a dataset provided by the UCI Machine Learning Repository that contains screening data for 292 patients. The dataset can be found at the following
URL: https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++

Abstract: Children screening data for autism suitable for classification and predictive tasks

Data Set Characteristics:	Multivariate	**Number of Instances:**	292	**Area:**	Life
Attribute Characteristics:	Integer	**Number of Attributes:**	21	**Date Donated**	2017-12-24
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	39319

A **solution statement** — I'm going to use Keras to build and train our network. This model will be relatively simple and will only use dense (also known as fully connected) layers. This is the most common neural network layer. The network will have one hidden layer, use an Adam optimizer, and a categorical cross entropy loss. I won't worry about optimizing parameters such as learning rate, number of neurons in each layer, or activation functions in this project.

A **benchmark model** — I'll use the default vanilla model as the benchmark. Hyper parameter tuning my final model will result in significant improvements over this benchmark.

A set of **evaluation metrics** — Once the model is trained, I need to test its performance on the testing dataset. The model has never seen this information before; as a result, the testing dataset allows me to determine whether or not the model will be able to generalize to information that wasn't used during its training phase. I will use some of the metrics provided by Scikit-learn for this purpose such as classification reports and accuracy score.

An outline of the project design — The project design includes the following phases:

- Data Preprocessing: This dataset is going to require multiple preprocessing steps. First, I have columns in our DataFrame (attributes) that I don't want to use when training our neural network. I will drop these columns first. Secondly, much of the data is reported using strings; as a result, I will convert the data to categorical labels. During the preprocessing, I will also split the dataset into X and Y datasets, where X has all of the attributes I want to use for prediction and Y has the class labels.

- Data Splitting: Split the data into a training set and validation set with an 80-20 split.

- Model Training and evaluation: I will start with the simple model architecture first before training and evaluation. Then iterate this process trying different architectures and hyper-parameters to reach an accuracy I'm happy with.

Source:

Fadi Fayez Thabtah
Department of Digital Technology
Manukau Institute of Technology,
Auckland, New Zealand
fadi.fayez '@' manukau.ac.nz

Relevant Papers:

1) Tabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. Proceedings of the 1st International Conference on Medical and Health Informatics 2017, pp.1-6. Taichung City, Taiwan, ACM.
2) Thabtah, F. (2017). ASDTests. A mobile app for ASD screening. www.asdtests.com [accessed December 20th, 2017].
3) Thabtah, F. (2017). Machine Learning in Autistic Spectrum Disorder Behavioural Research: A Review. To Appear in Informatics for Health and Social Care Journal. December, 2017 (in press)