**Q4 Investigating the Features**

1. In the paper by Sagaceta Mejia et al., the researchers constructed several technical analysis features that were calculated using the Python Pandas TA (Technical Analysis) library.
2. When deploying a neural network it is necessary to configure a layer of inputs (Hardesty 2017). These inputs form an array and, in the context of artificial neural networks are referred to as ‘features’. These features can be values of a sample of external data or the inputs of other neurons. The difference between features and methods are that features are ‘raw data’, by virtue of the fact that it forms the input layer in MLP (Multi-layer Perceptron).

A method in programming on the other hand is a function that transforms inputs. A method typically takes a number of arguments (inputs) and the method or function will produce a particular output through the transformation of the input (argument) (Telles 22).

Lastly, a model is a mathematical or programmatic representation of a real-world situation in which it supports decision making. (Winston 1, 5)

1. The features drawn from the Pandas-TA (Technical Analysis) library are divided into the following indicator categories (Johnson 2021):
2. Candles
3. Cycles
4. Momentum
5. Overlap
6. Performance
7. Statistics
8. Trend
9. Utility
10. Volatility
11. Volume
12. The researchers employed two techniques in the optimization of their MLP model, in line with their stated objective of formulating an emerging market ETF trend predictor:
13. Early stopping to prevent overfitting
14. Using a subset of 5 features to train the model. Deep learning neural networks are notoriously resource hungry and this step was introduced to ensure the efficient use of computer processing resources.

The researchers summarise the results of these two optimization techniques in Table 6 on page 9. From the table, it is clear that the training time is significantly reduced when compared to training the dataset on all the identified features and, it is demonstrated, that the predictive accuracy of the limited feature set is improved, most likely due to the increase in epochs (training iterations).

**Q5 Optimization**

1. Cross-validation is a method used in machine learning whereby a dataset is divided into a pre-specified number of blocks which are then used to train and test the outcome of several predictive algorithms. (StatQuest 2018)
2. The number of blocks used in cross-validation, is referred to as the number of folds. For instance, a common number of folds is 10 and this is what the researchers used as well. This means, the data was divided into 10 equally sized blocks and the number of k is 10, meaning 10-fold cross validation was used to test and train the model.
3. Distance measures are methods to quantify the distance between data points in one (or more dimensions). The Jaccard distance is calculated as the ratio of the intersection and union of a combination set. ()
4. The Jaccard distance is equal to 1 minus the Jaccard similarity (Rajaraman 94). As such, it is a measure of *dissimilarity*. Other measures encountered in lessons previously are the Euclidean and Manhattan distance measures. In Euclidean distance, an application of Pythagorean geometry is encountered whereby two points in the same space or dimension are set equal to the hypotenuse of a right-angled triangle. For the distance between two points this will be:

Closely related to this is the ‘Manhattan’ distance measure which relies on absolute values (and not squaring) to calculate distance. For comparison, this is how one would calculate the Manhattan distance between two points:

1. The researchers indicate that an optimal solution would be one where the MLP model assists investors in timing their entry and exit points into emerging market ETFs as well as serving as an indicator of risk to ensure that investors attain maximum profit for minimum loss.

Due to the limitations of the model they do warn however that investors would be best advised to also implement the model into a broader investment framework but provide no details on how such a framework should be constructed.

**References**

Hardesty, Larry. Explained: Neural Networks. 14 April 2017. <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>

Telles, Matt. *C# Black Book*. Coriolis, 2002.

Winston, Wayne. *Operations Research Applications and Algorithms Fourth Edition*. Brooks/Cole Cengage Learning. 2004.

Johnson, Kevin. *Pandas – Technical Analysis*, 2021. <https://github.com/twopirllc/pandas-ta>

StatQuest with Josh Starmer. “Machine Learning Fundamentals: Cross Validation” Youtube, 24 April 2018. <https://www.youtube.com/watch?v=fSytzGwwBVw>

Rajaraman and Ullman. *Mining of Massive Datasets*. Cambridge University Press, 2014. <http://infolab.stanford.edu/~ullman/mmds/book.pdf>