

Abstract

With ever growing amounts of data collected by companies in an attempt to gain a competitive edge by leveraging data analytics, a challenge arises to extract useful signal correlated with values of interest from data wide and big. Such challenge is also pertinent to tasks of automated financial forecasting, where historical data is limited in size, and abundance of data can be used as input to forecasting model, hence making input arbitrary "wide". Based on a study of financial forecasting for 10 different commodity futures, we test different approaches for joint selection of relevant historical market data and predictive model parameters. We discuss an algorithm that can efficiently perform such selection. We evaluate such algorithm empirically and show that analysis of inputs selected by the algorithm can provide deeper insights into the relationship between target value and inputs. Our results also indicate that there is no single "best" model for financial forecasting, albeit we find that recurrent neural networks perform best for forecasting of most commodities. Outlined approach and best practices can be directly transferred to other corporate forecasting tasks.