Sales Prediction with Time Series Modeling

Gautam Shine, Sanjib Basak

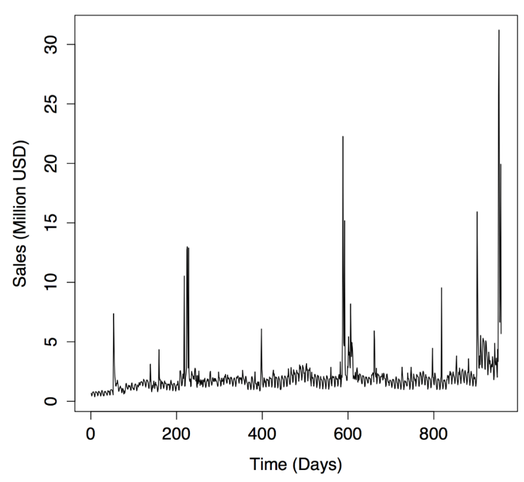
Predicting Sales

Sales forecasting is critical for retailers

- optimal stocking of products
- website stability under peak traffic
- planning, customer support, and marketing

Anomalies like Black Friday are especially difficult to capture with models

Example data from online sales of tech products:



Conventional Time Series Models

Autoregression (AR)

$$r_t = c + \epsilon_t + \sum_{i=1}^{p} \phi_{t-i}\, r_{t-i}$$

Moving Average (MA)

$$r_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_{t-i}\, \epsilon_{t-i}$$

Autoregressive Integrated Moving Average (ARIMA)

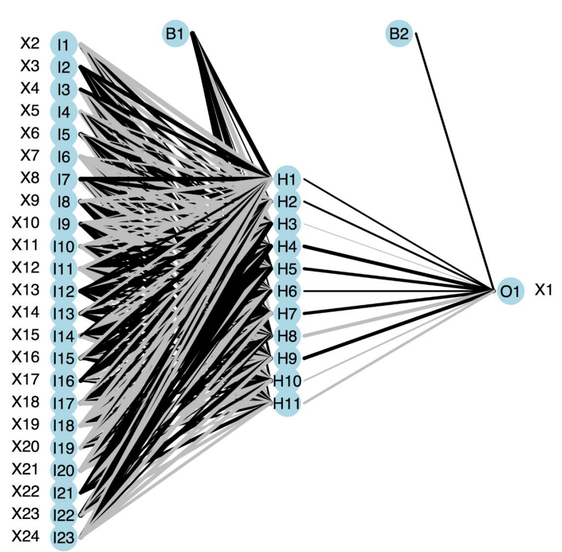
$$d_t = c + \epsilon_t + \sum_{i=1}^{p} \phi_{t-i}\, d_{t-i} + \sum_{i=1}^{q} \theta_{t-i}\, \epsilon_{t-i}$$

Feed-Forward Neural Networks

Nonlinearity induced by hidden layer

$$y_j = b_j + \sum_{i=1}^{n} w_{ij}\, x_i$$

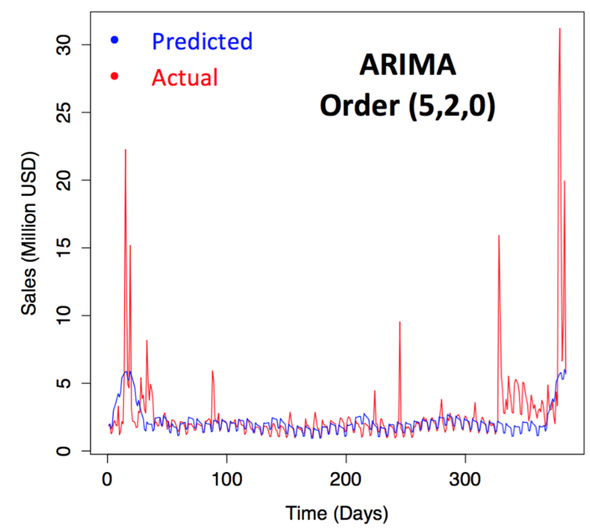
Parameters b_j and w_{ij} learned from data



Autoregression can be included through lagged inputs

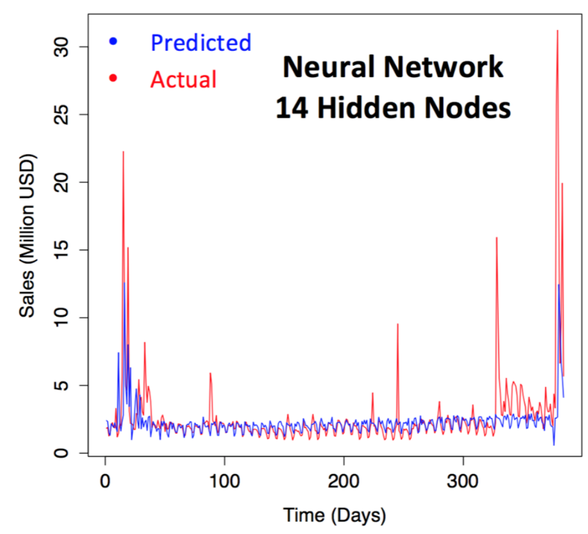
Optimization is non-convex, averaging needed

Forecast Results



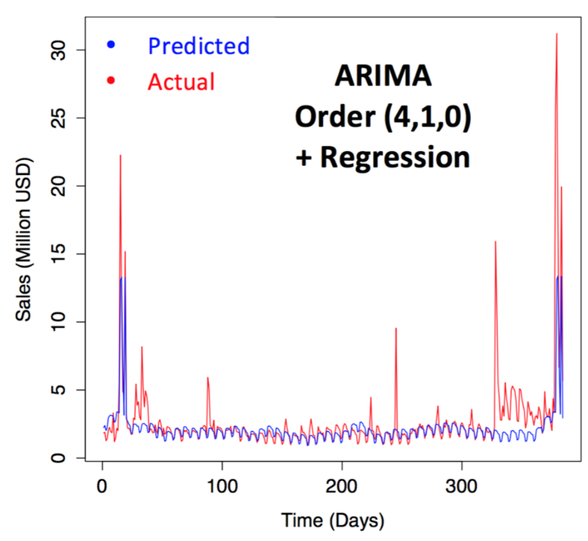
ARIMA

- Order (5,2,0) chosen by minimizing Akaike information, which is a regularized maximum likelihood estimate
- Fourier terms used to introduce multiple-seasonality
- Predictions are too smooth to capture sales spikes



Neural Net

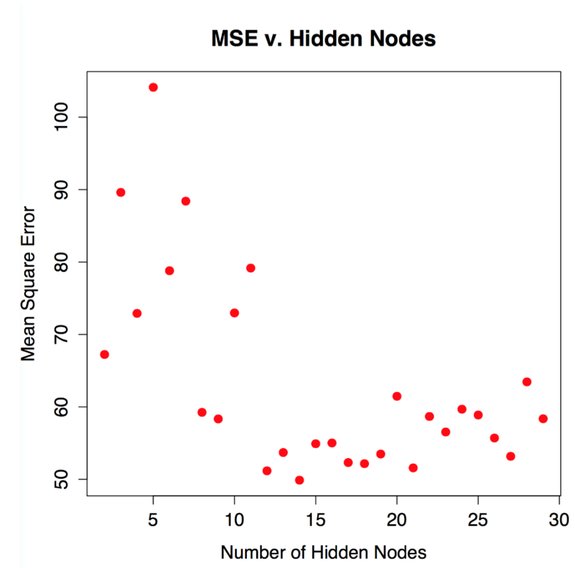
- 10 autoregression lags and 14 hidden layers, chosen by minimizing generalization error
- Averaged prediction of 100 nets initialized with random seeds
- Qualitatively captures sales spikes with lower MSE than ARIMA, but low daily accuracy



ARIMA + Regression

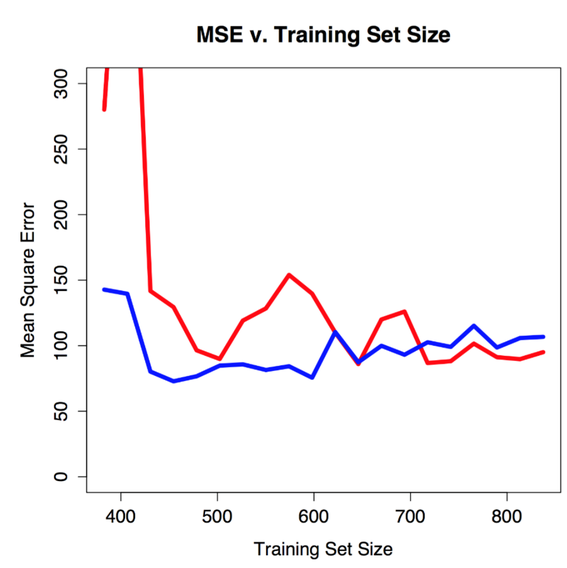
- Regression with indicator vectors for special days like Black Friday and Christmas greatly improves ARIMA
- Captures sales spikes while retaining daily accuracy
- Lower MSE than the neural net

Model Fitting



Hidden Node Count and Autoregression Order

- Erratic behavior can be observed when the number of hidden nodes is low
- Autoregression order made little difference beyond 5 or so
- Both chosen to minimize test MSE, but nearby values could also have been reasonable



Learning Curve

- Curve is non-monotonic since data is not i.i.d. and set size affects prediction interval
- Neural nets require more data to achieve the same forecast accuracy, but can exceed ARIMA with large data sets