Clothing sales time series forecast

Mariana Khachatryan

Outline

- ❑ Motivation
- ❑ Modeling framework
- ❑ Results
- ❑ Conclusions

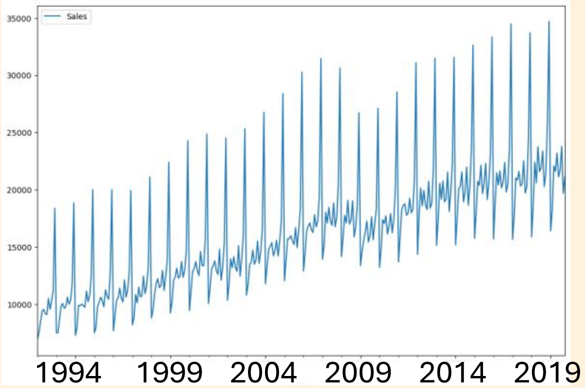
Motivation

Business Stakeholders

- ❑ Executive Leadership
  - Strategic planning, revenue goals.
- ❑ Merchandisers
  - Buying inventory ahead of demand.
- ❑ Marketing Team
  - Aligning campaigns with expected demand.

Modeling framework

Use clothing retail sale data from FRED (Federal Reserve Economic Database, https://fred.stlouisfed.org/series/RSCCASN)



Do time series forecasting for sales in the units of millions of dollars

Final data set:
334 data points
from 1992-2019

Split 70:30 and scale to [0,1]

Test set

Training set

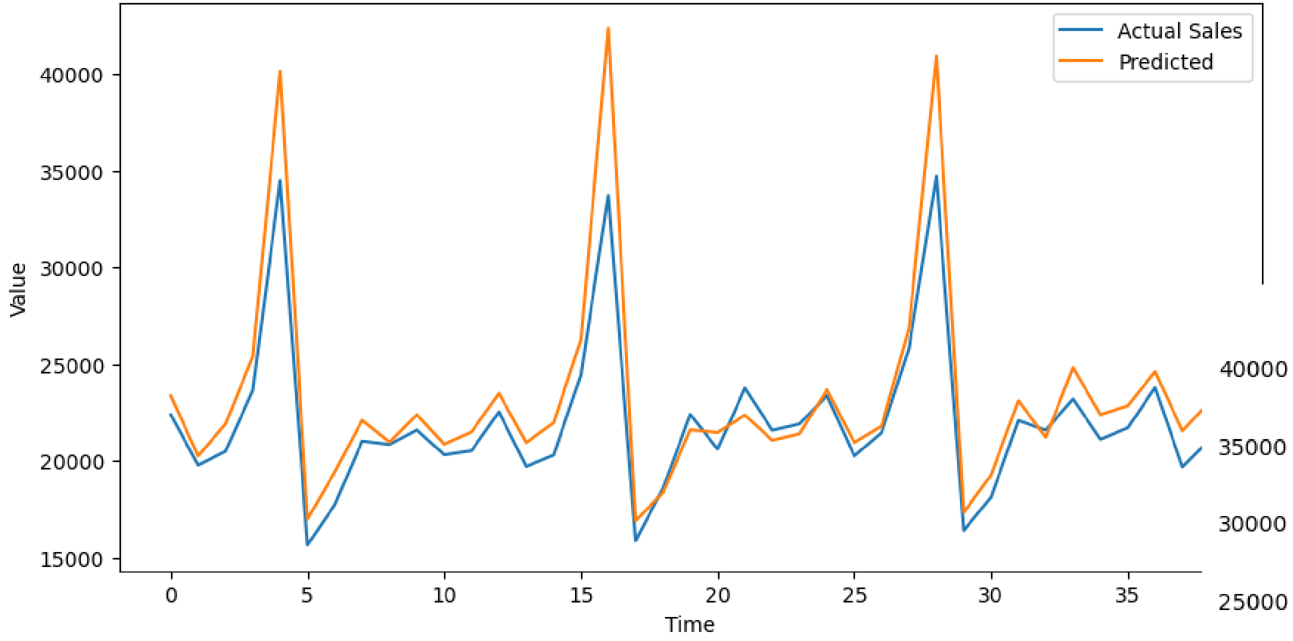
Use two different Machine Learning models

- LSTM
- Linear Regression (best performance, requires feature engineering)

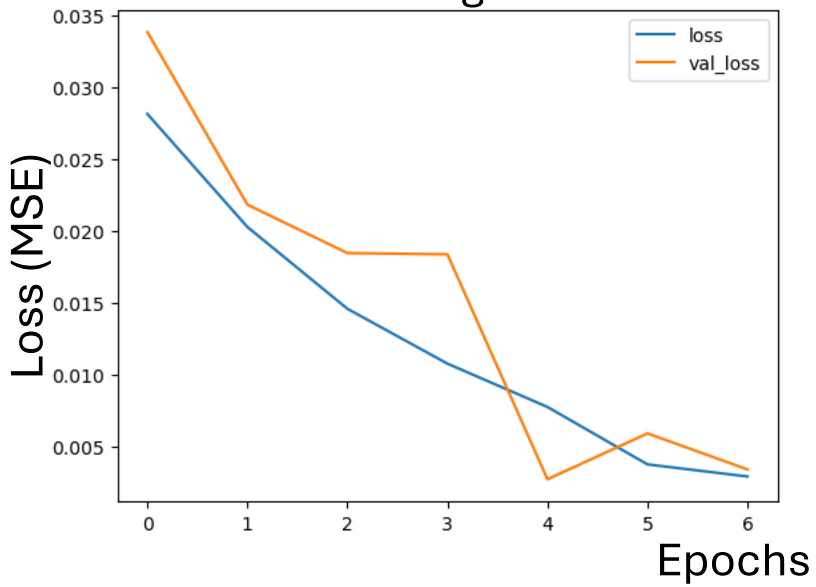
Results: LSTM

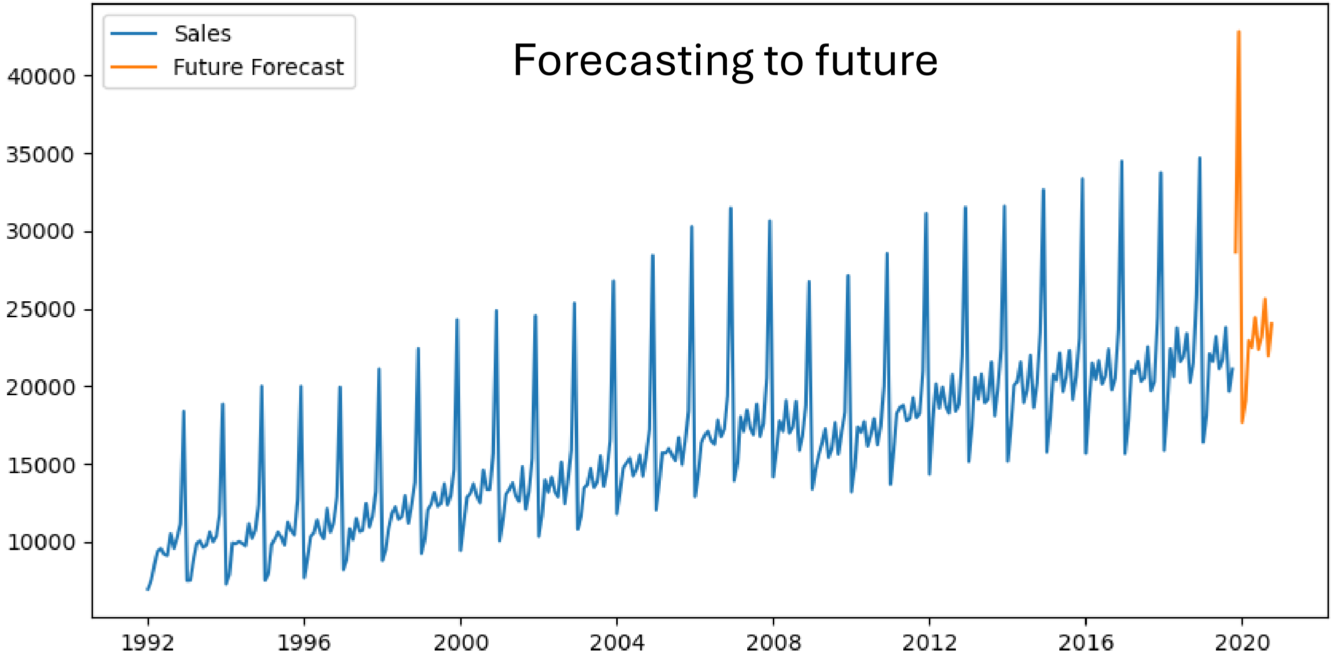
- LSTMs excel at capturing long-term dependencies in time series data
- Easy to implement
- Used early stopping based on validation loss

Model prediction on test sample



Learning Curve

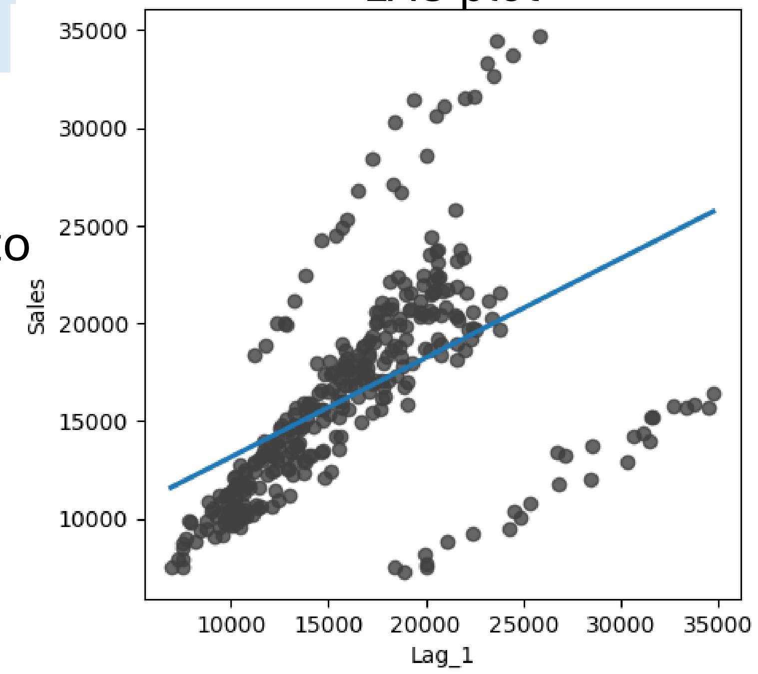




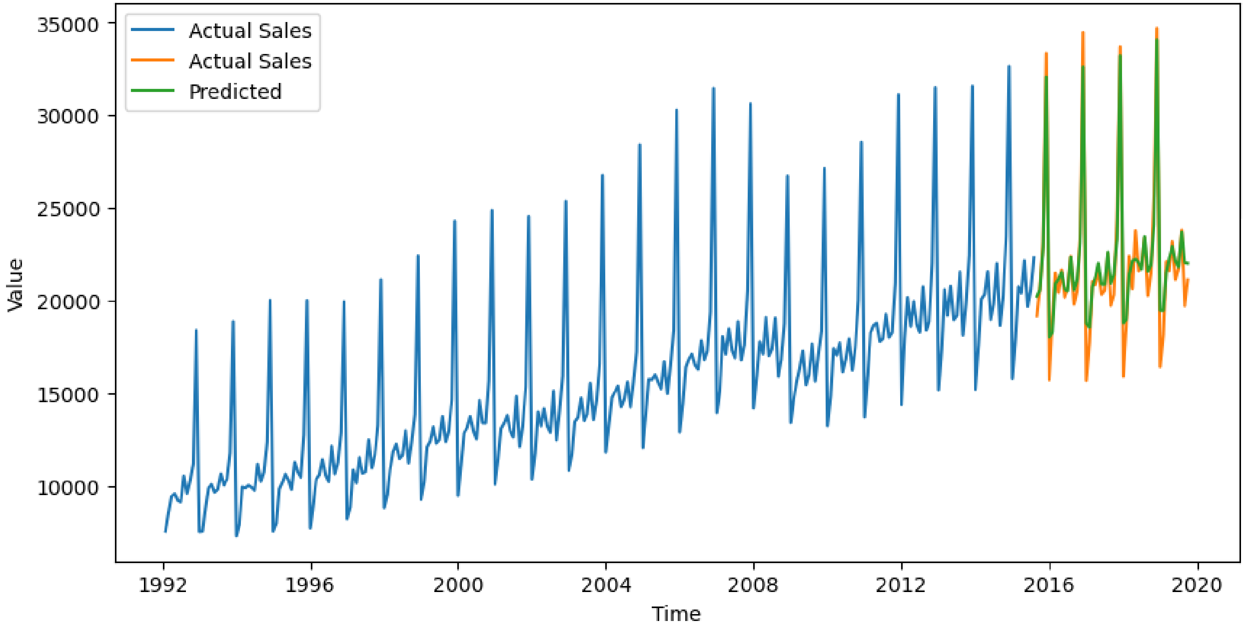
Results: Linear Regression

- Need to do some feature engineering
  - Create a time-step feature to describe trend
  - Add lag feature (value of the time series at the previous time step) to describe autocorrelation
  - Add month categorical variable to describe seasonality

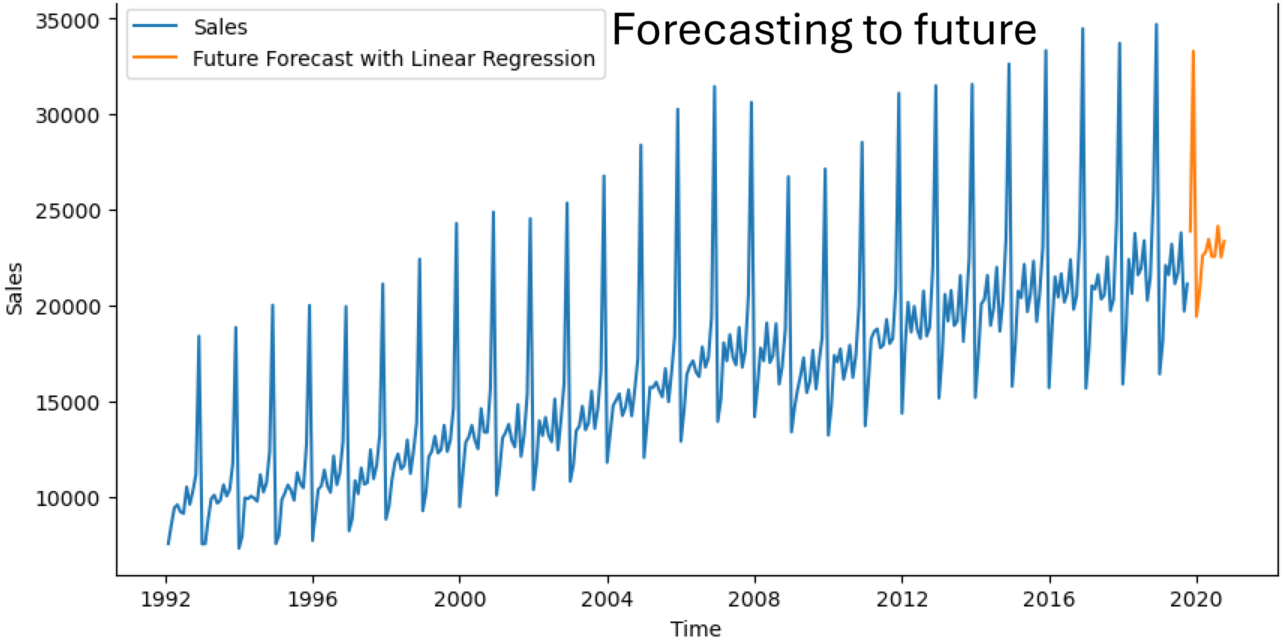
LAG plot



Model prediction on test sample



Forecasting to future



Results: Summary

Used early stopping based on validation loss for LSTM based Deep Learning model training.
Used feature engineering for Linear Regression based model.
Overall best model performance was obtained with Linear Regression. It is also a simple model and easy to interpret.

Model	**Root Mean Squared Error (RMSE) in** the units of **millions of dollars**	R^2
Linear Regression	1172 (5.4% of mean value)	0.92
LSTM	2212 (9.5% of mean value)	0.72

Conclusions

- ❑ Used LSTM and Linear Regression based models for time series forecasting.
- ❑ Linear Regression gives better performance, though both perform well.
- ❑ When dataset is small Linear Regression is better choice.

- ❑ Other models one can try

  - Try adding more lag features (look at partial autocorrelation)
  - Fit with hybrid model consisting of Linear Regression + XGBoost models.
  - Use statistical model called SARIMA (Seasonal Autoregressive Integrated Moving Average.)