LSTM Time Series Prediction on the Avocado Dataset

Jonathan Carter
University of Oxford
jcarter@robots.ox.ac.uk

November 9, 2020

ABSTRACT

A neural network and optimisation library was implemented[1] using PyTorch tensors, including the Adam optimiser and its hyper-gradient variant. An LSTM [1] was then implemented and used on the Kaggle Avocado dataset[2] to predict the avocado price time series over various locations in the US. This is shown to give good predictive accuracy, including uncertainty estimates when trained to parametrise a Gaussian distribution over outputs.

1 Introduction

The Avocado Dataset gives the weekly average price of avocados and the volume sold for over 50 unique regions in the US. In this report the time series of sales volumes is used to predict the series of average prices. Simple economic theory states that these variables should be inversely proportional. Figure 3 of the appendix shows there is correlation between the variables but that this is imperfect, indicating that learning a simple model mapping from volume to price will give poor accuracy. The data displays trends including seasonality, indicating that a sequence-sequence time series model may give improved predictive accuracy. This motivates the use of an LSTM, a recurrent neural network architecture suited to time series data, particularly for tasks such as sequence-sequence regression. More details on the LSTM are given in the appendix.

2 Network architecture

The LSTM used a hidden state h_t of dimension 100. The regression network from h_t to the output y_t consisted of a fully connected network with 1 hidden layer of 100 neurons which used the swish [2] activation function. The hypergradient variant [3] of the Adam optimiser was used, with the standard initial Adam parameters[3] and a hypergradient learning rate $\beta = 5 * 10^{-9}$. This is compared with regular Adam in Section 4.1.

The network is trained to output a Gaussian distribution over the price y_t using the following loss function:

$$L(\theta) = \sum_i \sum_t -\log p(y_t^{(i)} | h_t^{(i)}, \theta) \tag{1}$$

where $p(y_t^{(i)} | h_t^{(i)}, \theta)$ is a Gaussian distribution whose mean and variance are conditioned on the hidden state:

$$p(y_t^{(i)} | h_t^{(i)}, \theta) \sim \mathcal{N}(\mu_\theta(h_t), \sigma_\theta(h_t)) \tag{2}$$

The model thus learns a distribution over the price at each time t for each input sequence of weekly sales volumes $x_{0:T}^{(i)}$ from each respective region denoted i. Each hidden state, h_t, is computed by forward propagation through the LSTM from the start of the sequence i.e. $h_t = f(x_t, h_{t-1}, \theta)$ from the initial hidden state $h_0 = 0$.

[1] https://github.com/joncarter1/AvoLSTM
[2] https://www.kaggle.com/neuromusic/avocado-prices
[3] $\alpha = 10^{-3}, \beta_1 = 0.9, \beta_2 = 0.99$

3 Results

3.1 Adam Optimisation

In figure 1 training curves using both the regular and hypergradient version (AdamHD) are shown.

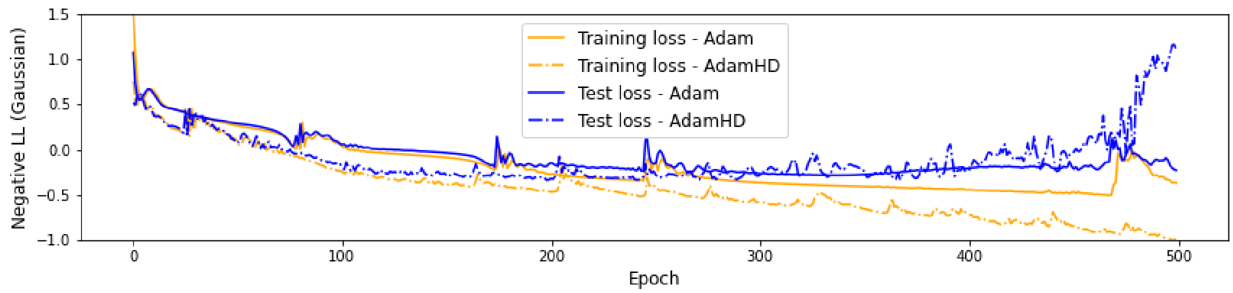

Figure 1: LSTM network convergence using the Adam optimiser inc. with hypergradient descent

3.2 Model predictions

In figure 2 model predictions are shown for the state of South Carolina, additional plots are given in the appendix.

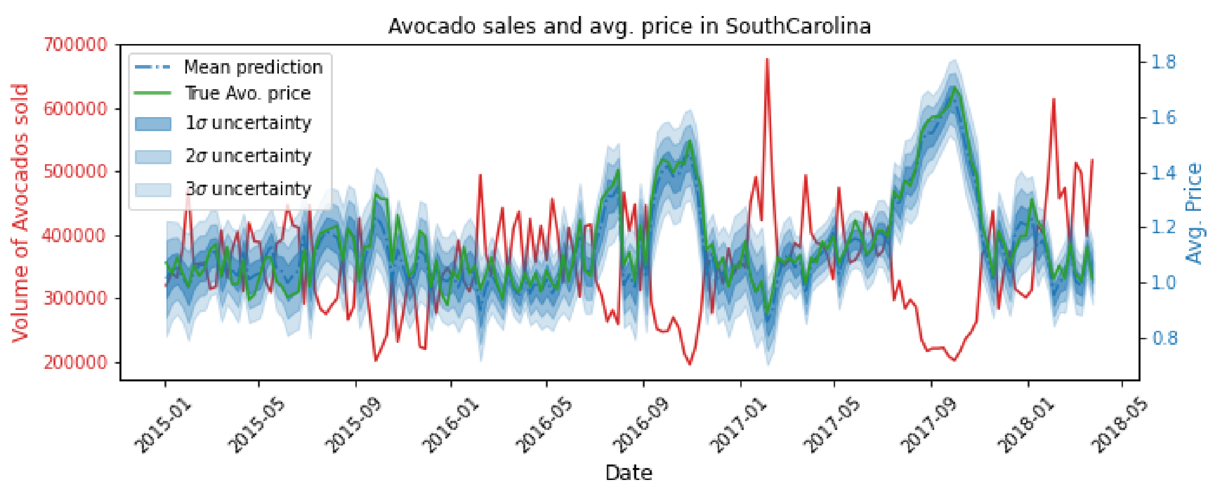

Figure 2: LSTM network predictive distribution over avocado prices.

4 Conclusions

A library for neural network training was written and used to implement an LSTM for probabilistic time series prediction. Regular SGD and Adam optimisers were implemented, including their hyper-gradient variants.

The hyper-gradient variant of Adam was found to give faster convergence as in Figure 1. These results were sensitive to the hyper learning rate: set too high it resulted in instability, however it was relatively simple to tune.

From Figure 2 the output distribution can be seen to give good predictive uncertainty estimates over the price, indicating the LSTM is well suited to this type of sequence-sequence regression task. This is as expected given their wide-spread use and success in tasks such as machine translation.

References

[1] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. Publisher: MIT Press.

[2] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions. *CoRR*, abs/1710.05941, 2017.

[3] Atilim Gunes Baydin, Robert Cornish, David Martinez Rubio, Mark Schmidt, and Frank Wood. Online learning rate adaptation with hypergradient descent, 2018.

A Correlation of price and volume

In the left hand plot all weekly price data is plotted against the weekly volume. We observe there is weak correlation between the variables but not a direct mapping. In the right hand plot we see that there are correlated trends in the time series, which the LSTM is able to learn.

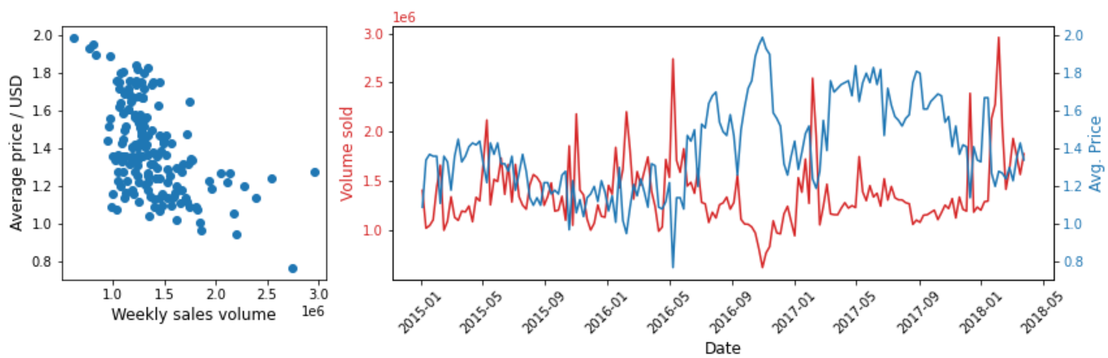

Figure 3: Weekly avg. price and volume of avocados sold in New York

B LSTM Architecture

LSTMs are a recurrent neural network architecture usually applied to problems with uniformly sampled time series data. At each time-step the input, x_t, results in a non-linear update of the internal hidden state, h_t. These hidden states can then be used for regression or classification in a downstream task. The internal gating of the LSTM gives the network memory and allow it to learn to weight new and old information for the task.

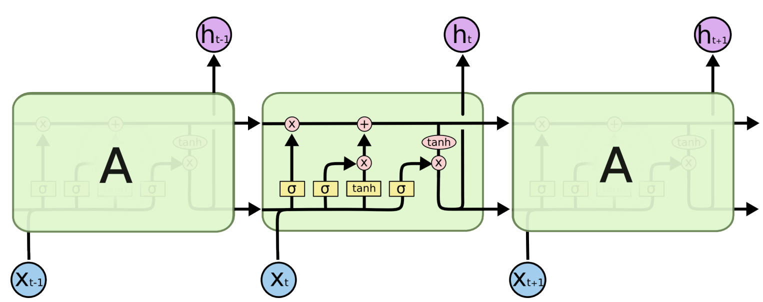

Figure 4: LSTM network architecture (https://colah.github.io/posts/2015-08-Understanding-LSTMs/)

C Additional predictive distributions

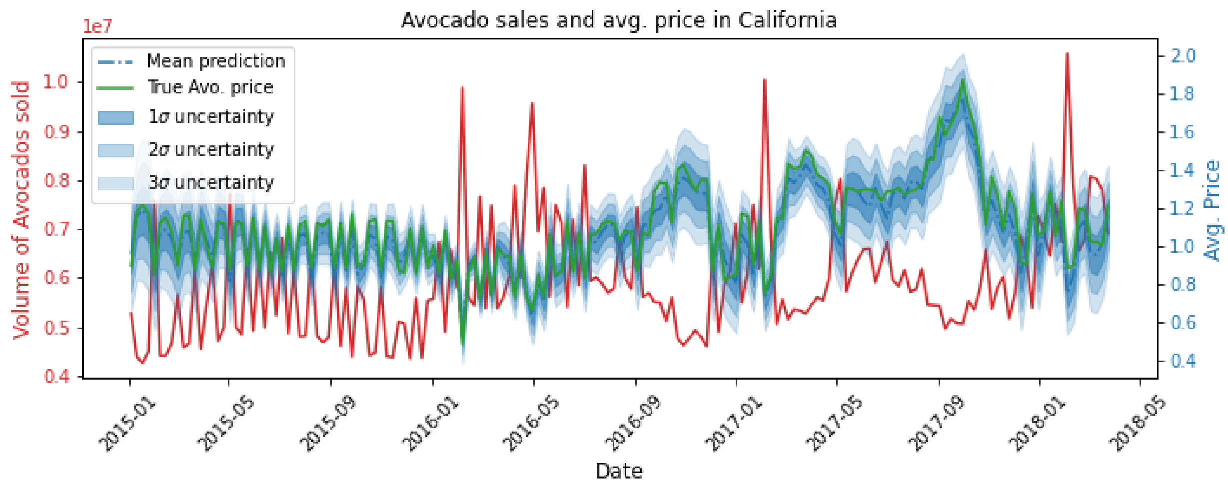

Figure 5: LSTM network predictive distribution over avocado prices in California.

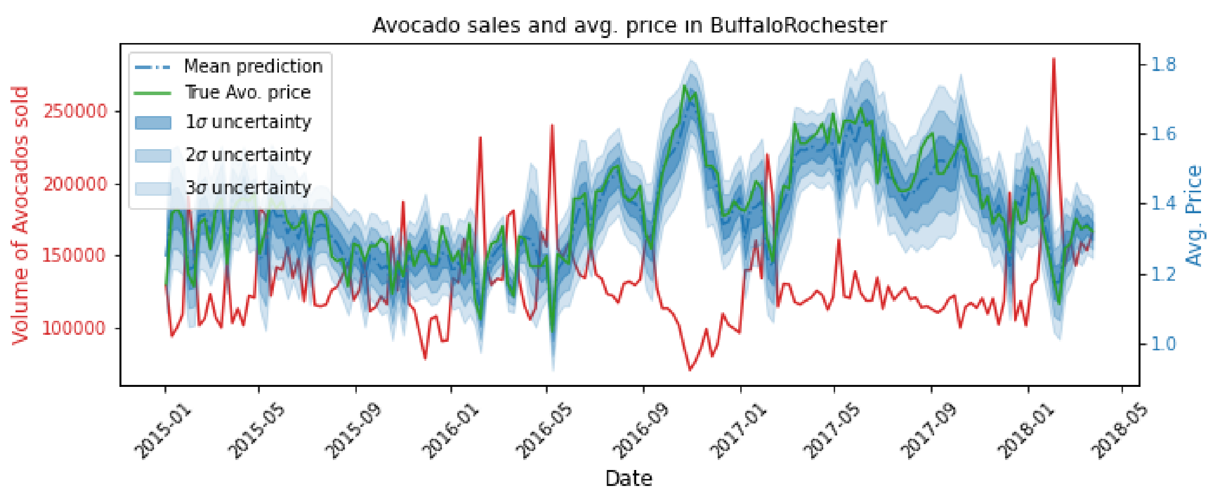

Figure 6: LSTM network predictive distribution over avocado prices in Buffalo Rochester.