Please provide complete and well-written solutions to the following exercises.

Due November 14, 4PM PST, to be uploaded as a single PDF or Jupyter Notebook document to brightspace.

Homework 7

Exercise 1. The electricity prices we analyzed in class behaved a bit differently over the last 7 years. Perform a linear regression just on data from 2017 to the present. What slope do you get?

Our electricity prices did not reflect inflation. Perform the same linear regression analysis (for the full dataset, and for the past 7 years) by adjusting for inflation. (This part of the exercise is intentionally open-ended. To get some usable inflation numbers, see e.g. https://fred.stlouisfed.org/series/FPCPITOTLZGUSA . For example, if electricity rates increased by three percent in one year when inflation is three percent for that whole year, then the inflation adjusted electricity rate would be constant over that year.)

Exercise 2. Let

$$h(x) := \frac{1}{1+e^{-x}}, \qquad \forall\, x \in \mathbf{R}.$$

Fix $x \in \mathbf{R}$ and $y \in [0,1]$. Define $t\colon \mathbf{R}^2 \to \mathbf{R}$ by

$$t(a,b) := \log\Big([h(ax+b)]^y[1-h(ax+b)]^{1-y}\Big), \qquad \forall\, a,b \in \mathbf{R}.$$

Show that t is concave. Conclude that t has at most one global maximum.

Exercise 3. Use logistic regression, a two layer and a three layer neural network to classify an IMDB dataset found here:

https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews

(As we did in class, use the TfidfVectorizer to vectorize the text documents.)

Compare the performance and speed of each of these three methods. (Optional: use another classification method that you think will do better.)

Use a training set size of 45000 and a test set size of 5000.

Exercise 4. Adjust the parameters such as batch size for the deep learning approaches to classifying digits in the MNIST dataset. Are you able to significantly improve the percentage of correct classifications above 99.3% ?