Assignment 5 Supervised Learning

Due Date: 1400/Bahman/05

Foundations of Data Science

Supervisor Dr. SaeedReza Kheradpisheh

Teaching Assistants Hesam Damghanian, Ali Rahimi

Shahid **B**eheshti **U**niversity

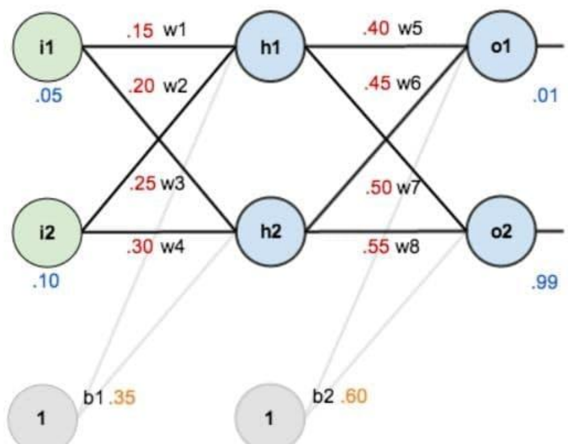
Optimization (3/20)

Theory

1. Discuss these questions in-length:
    a. Why do we usually consider cost function as a negative of log-likelihood?
    b. Explain L1 and L2 regularization; compare them to each other.
    c. What is the effect of the momentum in a learning method?
2. Learning in a Neural network is done by updating the weights. “Backward propagation of errors” known as backpropagation is the common method to do the updating efficiently. Now consider the following neural network with Sigmoid as neurons’ activation function and “Mean Squared Error” as the cost function.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2$$

    a. Calculate the network error after one step of feed-forward.
    b. Calculate one step of backpropagation with a learning rate equal to 0.3.



3. Suppose batch gradient descent in a deep network is taking excessively long to find a value of the parameters that achieves a small value for the cost function J(W[1],b[1],...,W[L],b[L]). Which of the following techniques could help find parameter values that attain a small value forJ? (Check all that apply)
    - ☐ Try better random initialization for the weights
    - ☐ Try mini-batch gradient descent
    - ☐ Try using Adam
    - ☐ Try initializing all the weights to zero
    - ☐ Try tuning the learning rate α

Implementation

In the last assignment, you implemented the gradient descent algorithm for optimizing linear regression, this time you are asked to implement the following optimization algorithms in python for the same model as previous.

1. SGD
2. SGD with momentum
3. AdaGrad
4. RMSProp
5. Adam

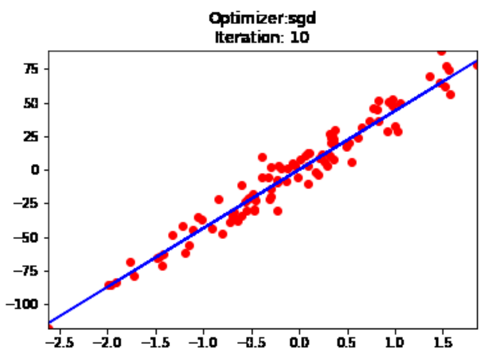
Tasks:

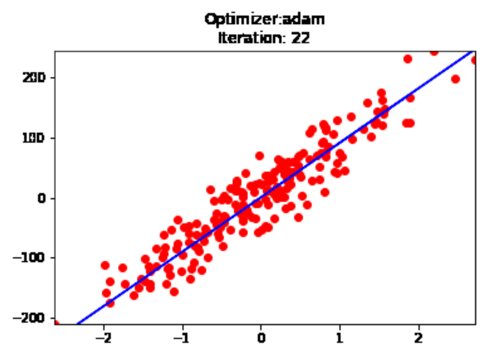
1. Use the template code provided, only modify the TODO parts and for reproducibility reasons, do not change the random seed.

2. You should find and **report** the right hyperparameters for these algorithms to converge, i.e., learning rate and **explain** each hyperparameter for each algorithm (1-2 lines).

3. Discuss and compare each algorithm's pros and cons (2-3 lines each).

You should expect outputs like these:





Bonus:

1. Visualize the gradient descent process (Check out the template code for guides).
2. Implement new optimization methods to solve the toy regression problem, from recent papers, and test your model with 2d inputs; in addition, visualize the regression process on 3d plots.

Note:

- The template code is just to accelerate your work, you may change it or implement yours.
- See [here] for more information about optimization algorithms.

Time-Series (1.5/20) +Bonus

Implementation

In this assignment you are requested to forecast environmental data and discuss and compare performance of conventional time series algorithms in forecasting these data.

Steps:

1. Define seasonality, trends, and stationarity in time series. Report the ADF statistics[1] of your target data.
2. Acquire air quality data or historical weather data from up to 3 major cities, at least with 60 months of historical data, if available.
3. Preprocess and feature engineer data[2] and split the data into train and test sets, time-series data should not be shuffled so keep the most recent part of the data as a test set. Besides, you might as well check time series split[3] from the sklearn package.
4. Briefly explain and use the following algorithms to forecast a target, e.g., temperature, AirQualityIndex (AQI), and report results.
    a. Autoregressive[4] models variants, i.e., ARIMA.
    b. RNN variants
    c. CNN (+check out 1d convolutions)
    d. FBProphet[5]
    e. Others…

Dataset suggestions:

1. Air quality: Air Quality Historical Data Platform
2. US Weather: US Weather Events (2016 - 2020) | Kaggle
3. Historical Hourly Weather Data 2012-2017 | Kaggle
4. UCI Machine Learning Repository: Air Quality Data Set
5. More: Find Open Datasets and Machine Learning Projects | Kaggle

Useful links:

1. Lesson 1: Time Series Basics | STAT 510
2. Start Here with Machine Learning: Time Series
3. Time Series analysis tsa — statsmodels

Note: Don't limit yourself only to the aforementioned methods, based on the quality of your work, extra scores may be granted for observing and testing other TS forecasting algorithms. Once again, we emphasize the report; it should contain all your questions and your innovative findings. Use figures, pictures, and tables, and DO NOT PUT ANY CODE IN THE REPORT.

---

[1] Wikipedia: Augmented Dickey-Fuller test
[2] Time-related feature engineering — scikit-learn 1.0.2 documentation
[3] TimeSeriesSplit
[4] Classical models
[5] Prophet