Homework 2

January 27, 2023

Description

We have been looking at the Linear Model Cannonical form

$$f_e(\boldsymbol{x}|w,b) = w^t\boldsymbol{x} + b$$

Although its power is somewhat limited, we can still have fun with it by going beyond the canonical solution. Take a look at the jupyter notebook provided for you. There you have:

1. Jax canonical linear model implementation.
2. A naive version of gradient descent that still requires some work for finding the learning rate.

Homework

1. You need to define one of the possible algorithms for learning rate to improve the base case at the jupyter notebook
   (a) Bisection
   (b) Golden Ratio
   (c) etc

   Test against the data produced at the Jupyter Notebook

2. You are going to upload a csv data into a database (It has been cleaned for you). The explanation of the dataset is at
   - https://www.openml.org/search?type=data&sort=nr_of_likes&status=any&id=1590 (Instructions to load it are being provided)

   Then, you are going to proceed to classify the data
   (a) Class 1 People making more than \$50,000
   (b) Class 2 People making less than \$50,000

   Hint: A lot of the features are categorical, it would be convenient to look at one shot representation.

   For this, you would use the new implemented Linear model to do the classification under Ridge Regularization (This is the serious part).

$$\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \left\{ \sum_{i=1}^{N} \left(y_i - w_0 - \sum_{j-1}^{d} x_{ij} w_j \right)^2 + \lambda \sum_{j=1}^{d} w_j^2 \right\}$$

   For testing you would use the following formulations:
   (a) Accuracy $= \frac{TP+TN}{TP+FP+FN+TN}$ where TP = True Positives are Class 1, TN = True Negatives are Class 2
   (b) Recall $= \frac{TP}{TP+FN}$ Here you can said how good is your algorithm

Thus, you will compare the non-regularized version to the regularized one.

3. Given the following definition,

   $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$ then

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \dots & \frac{\partial y_2}{\partial x_n} \\ \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \dots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

   (a) Prove that given $\boldsymbol{y} = A\boldsymbol{x}$, we have $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = A$

   (b) Given $\alpha = \boldsymbol{y}^T A \boldsymbol{x}$, prove $\frac{\partial \alpha}{\partial \boldsymbol{x}} = \boldsymbol{y}^T A$