Homework 3

Description

We have been looking at the Logistic Model form

$$L\left(\boldsymbol{x}_1,...,\boldsymbol{x}_N|\theta\right) = \log \prod_{i=1}^{N}\prod_{l=1}^{K} p\left(\boldsymbol{x}_i|\boldsymbol{w}_l\right)^{I\{\boldsymbol{x}_i\in\omega_l\}}$$

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 at

- https://github.com/kajuna0amendez/Class_MachineLearning_Jax/tree/main/notebook/Class_05

There you have a Jax canonical logistic implementation.

Homework

1. You need to implement the regularization version of the Logisitc Regression

$$L\left(\boldsymbol{x}_1,...,\boldsymbol{x}_N|\theta\right) = \log \prod_{i=1}^{N}\prod_{l=1}^{K} p\left(\boldsymbol{x}_i|\boldsymbol{w}_l\right)^{I\{\boldsymbol{x}_i\in\omega_l\}} + \lambda trace\left(W^TW\right)$$

$$\text{s.t. } W = \begin{pmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_k \end{pmatrix}$$

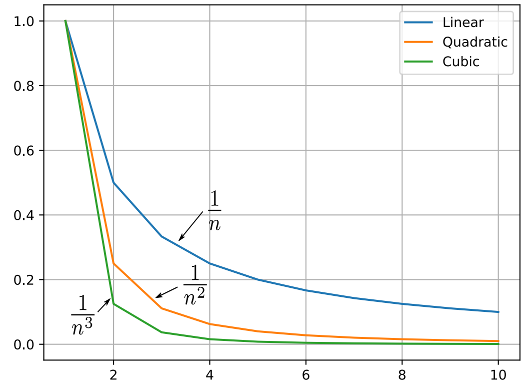
   You will implement:

   (a) The search for the λ hyper-parameter.
   (b) The Quasi-Newton Method described at the notes to accelerate the algorithm.
   (c) The linear search for the learning rate parameter.

2. Remember the data set

   - https://www.openml.org/search?type=data&sort=nr_of_likes&status=any&id=1590 (Instructions to load it are being provided)

   Using accuracy and recall you will compare the classic and Newton algorithms convergency rates as the one show in the figure:



3. In the Robbins-Monro (Stochastic Gradient Descent) proof, we have some stuff that was left to you to do:

   (a) Prove that if $Pr\left[y_n \leq y|\boldsymbol{w}_n\right] = H\left(y|\boldsymbol{w}_n\right)$ a distribution then if

   $$e_n = E\left[\int_{-\infty}^{\infty} \left(y-\alpha\right)^2 dH\left(y|\boldsymbol{w}_n\right)\right]$$

      i. Now, assuming that exist C such that (i.e. zero out of the range $[-C, C]$)

      $$Pr\left[\left|Y\left(\boldsymbol{w}\right)\right| \leq C\right] = \int_{-C}^{C} dH\left(y|\boldsymbol{w}\right) = 1 \ \forall x$$

   Then, we have that $0 \leq e_n \leq \left[C + |\alpha|^2\right] < \infty$

10 Extra Points:

Improve Performance of the algorithm by using

1. A Sparse representation of Y_{hot}
2. Doing the direct product at the Correct Place
3. Imagine a metric to show that