CSCI 6968/4968: Homework 1

Assigned Friday January 27 2023. Due by 11:59pm Friday February 10 2023.

Create a Jupyter notebook for this assignment, and use Python 3. Write documented, readable and clear code (e.g. use reasonable variable names). Submit this notebook interspersing any textual answers in markdown cells (using LaTeX), clearly labeled, along with your code.

1. [30 pts] Compute the gradient and Hessians of the loss function for ℓ_2-regularized logistic regression, given as

$$f(\boldsymbol{\omega}') = \frac{1}{n}\sum_{i=1}^{n} \log\left(1 + \exp(-y_i(\omega_0 + \boldsymbol{\omega}^T \mathbf{x}_i))\right) + \lambda \|\boldsymbol{\omega}\|_2^2.$$

   Note that the bias term is not regularized. Here, $\boldsymbol{\omega}' = [\omega_0; \boldsymbol{\omega}]$ is the vertical concatenation of the bias and the feature coefficients, for convenience. Write $\nabla_{\boldsymbol{\omega}'} f$ and $\nabla^2_{\boldsymbol{\omega}'} f$ as concisely as you can.

   Argue that logistic regression is a convex optimization problem.

2. [70 pts] Try multinomial logistic regression on the Fashion-MNIST data set.

   - Download the Fashion-MNIST data set at `https://github.com/zalandoresearch/fashion-mnist` and create your Jupyter notebook in the base directory of this repo so that the relative paths will be consistent for the TA.
   - Load the training and testing splits of the data set according to the instructions for loading the Python given in the README; use the same variable names.
   - Preprocess the training data so that all 784 features (pixel values) look essentially like standard Gaussians (this helps with accuracy and convergence). Do this by fitting an sklearn `StandardScaler` on the training data and applying it to the training and test data sets; see `https://scikit-learn.org/stable/modules/preprocessing.html`. Overwrite the original training and test data sets with these processed data sets.
   - Use Matplotlib and the `helper.get_sprite_image` function from the Fashion-MNIST repo to display one of each of the ten image classes in the training data set. Show them in a 2×5 grid.
   - Use sklearn to fit a multiclass logistic regression model to predict the image labels, using the SAGA solver (you may need to increase the number of maximum iterations and/or decrease the convergence tolerance — be reasonable). Report the top-1 and top-5 classification accuracies and confusion matrices on the test and train data sets: see `https://scikit-learn.org/stable/modules/generated/sklearn.metrics.top_k_accuracy_score.html` and `https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html`.
   - What conclusions do you draw about the performance of the model on the various classes, given the confusion matrix on the test set?
   - Select and display one of the misclassified images in the training data set: what class should it have been classified as, and what class was it misclassified as?

3. (CSCI 6961 students) [50 pts] Compute the gradients $\nabla_{\mathbf{W}} f$ and $\nabla_{\mathbf{b}} f$ of the loss function for ℓ_2-regularized multinomial logistic regression,

$$f(\mathbf{W}, \mathbf{b}) = -\frac{1}{n}\sum_{i=1}^{n} \left[\mathbf{y}_i^T(\mathbf{W}\mathbf{x}_i + \mathbf{b}) - \log\left(\mathbf{1}^T \exp(\mathbf{W}\mathbf{x}_i + \mathbf{b})\right)\right] + \lambda \|\mathbf{W}\|_F^2.$$

   Recall that $\|\mathbf{W}\|_F^2$, the squared Frobenius norm of $\mathbf{W}$, is the sum of its squared entries: $\|\mathbf{W}\|_F^2 = \sum_{i,j=1}^{k,d} W_{i,j}^2$. Give as concise expressions for these gradients as you can. It will help to use the chain rule.