Homework 3, due January 31st, 11:59pm

January 24, 2024

1. We will consider a slightly modified MAP learning for Logistic Regression (slide 9 of the Logistic Regression slides) that minimizes the following loss function:

$$C(\mathbf{w}) = -\frac{1}{N}L(\mathbf{w}) + \lambda \mathbf{w}^T\mathbf{w}$$

where $L(\mathbf{w}^{(t)})$ is the log-likelihood from page 7 of the Logistic Regression slides. Minimizing this loss by gradient descent has the following update equation:

$$\mathbf{w}^{(t+1)} \leftarrow \mathbf{w}^{(t)} - \eta\lambda\mathbf{w}^{(t)} + \frac{\eta}{N}\frac{\partial L(\mathbf{w}^{(t)})}{\partial \mathbf{w}}$$

Observe that there is an extra factor of $1/N$ in the loss term compared to the class notes.

Before using logistic regression, be sure to normalize the variables of the training set to have zero mean and standard deviation 1, and to use the exact same transformation on the test set, using the mean and standard deviation of the training set.

a) Using the `Gisette` data, train a logistic regressor on the training set, starting with $\mathbf{w}^{(0)} = 0$, with 300 gradient descent iterations and shrinkage $\lambda = 0.0001$

   Find a good learning rate η such that the loss converges in at most 300 iterations and is monotonically decreasing. Plot the training loss vs iteration number. Report in a table the misclassification error on the training and test set. On the same graph, plot the Receiver Operating Characteristic (ROC) curve of the obtained model on the training and test set. (2 points)

b) Repeat point a) on the `madelon` dataset. (2 points)

c) Repeat point a) on the `dexter` dataset. (2 points)

2. For the `Gisette` data, minimize analytically the following loss function:

$$C(\mathbf{w}) = \frac{1}{N}\sum_{i=1}^{n}(y_i - \mathbf{w}^T\mathbf{x}_i - w_0)^2 + \lambda(\mathbf{w}^T\mathbf{w} + w_0^2)$$

where $\lambda = 0.0001$ and $(\mathbf{x}_i, y_i)$ are the training examples. (Hint: this is exactly the linear regression loss, used for classification.)

We will predict the classification label as $\hat{y} = \text{sign}(\mathbf{w}^T\mathbf{x} + w_0)$. (see next page)

a) Report in a table the misclassification error on the training and test set. (2 point)

b) On the same graph, plot the Receiver Operating Characteristic (ROC) curve of the obtained model on the training and test set. (1 point)