"Machine Learning and Computational Statistics"

8^{th} Homework

Exercise 1 (Python code + text):

(a) Suppose you are given a data set $Y=\{(y_i,\mathbf{x}_i), i=1,\ldots,N\}$ where $y_i \in\{0,1\}$ is the class label for vector $\mathbf{x}_i\in R^l$. **Extract** the gradient descent logistic regression classifier for the two-class case (write in detail the algebraic manipulations using the hints in the relevant slides of its presentation).

(b) Consider a two-class, two-dimensional classification problem for which you can find attached two sets: one for training and one for testing (file *HW8.mat*). Each of these sets consists of pairs of the form $(y_i,\mathbf{x}_i)$, where y_i is the class label for vector $\mathbf{x}_i$. Let N_{train} and N_{test} denote the number of training and test sets, respectively. The data are given via the following arrays/matrices:

- ***train_x*** (a N_{train} x2 **matrix** that contains in its rows the training vectors $\mathbf{x}_i$)
- ***train_y*** (a N_{train}–dim. column **vector** containing the **class labels** (0 or 1) of the corresponding training vectors $\mathbf{x}_i$ included in ***train_x***).
- ***test_x*** (a N_{test}x2 **matrix** that contains in its rows the test vectors $\mathbf{x}_i$)
- ***test_y*** (a N_{test}–dim. column **vector** containing the **class labels** (0 or 1) of the corresponding test vectors $\mathbf{x}_i$ included in ***test_x***).

Train the Bayesian classifier using the training set given above and **measure** its performance using the test set (use the relevant code from Homework 6).

(c) **Train** the logistic regression classifier using the training set given above and **measure** its performance using the test set.

(d) Compare the results of the two classifiers and comment on them.

Exercise 2:

Suppose you are given a data set $Y=\{(y_i,\mathbf{x}_i'), i=1,\ldots,N\}$ where $y_i \in\{0,1\}$ is the class label for vector $\mathbf{x}_i'\in R^l$. Assume that y and $\mathbf{x}'$ are related via the following model: $y=f(\boldsymbol{\theta}^T\mathbf{x}'+\theta_0)$, where $\boldsymbol{\theta}$ and θ_0 are the model parameters and $f(z)=1/(1+\exp(-az))$.

(a) **Plot** the function $f(z)$ for various values of the parameter a.

(b) Propose a gradient descent scheme to **train** this model (that is, to estimate the values of the involved parameters), based on the **minimization** of the sum of error squares criterion, using Y.

(c) Can the model ever respond with a **"clear"** 1 or a **"clear"** 0, for a given $\boldsymbol{x}$?

(d) How can we interpret the response of the model for a given $\boldsymbol{x}$?

(e) Propose a way for leading the model responses very close to 1 (for class 1 vectors) or 0 (for class 0 vectors).

Hints:

(a) Use a more compact notation by setting $\boldsymbol{x}_i=[1\ \boldsymbol{x}_i']^{\mathrm{T}}$, i=1,…N, and $\boldsymbol{\theta}=[\theta_0\ \boldsymbol{\theta}]^{\mathrm{T}}$. The model then becomes $y=f(\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{x})$.

(b) The sum of error squares criterion in this case is $J(\boldsymbol{\theta}) = \sum_{n=1}^{N}\left(y_n - f(\boldsymbol{\theta}^T\boldsymbol{x}_n)\right)^2$.

(c) It is $f'(z) = \frac{df(z)}{dz} = af(z)\left(1-f(z)\right)$.