CMSC 510 – Fall 2018

Homework Assignment 3

Announced: 11/12

Due: Monday, 12/3, noon

Nonlinear classification

- Use tensorflow library
- Two-class problem:
  - Take MNIST dataset
  - Partition digits into two classes
    - E.g. 0,1,2,3,4 vs 5,6,7,8,9
    - Or prime: 1,2,3,5,7 vs 0,4,6,8,9
- Implement and test:
  - Kernel Logistic regression
    - Use class variable for all training samples
  - Semi-supervised Kernel logistic regression
    - Assume only e.g. 10% of training samples have known class (you *forgot* the class for the others)

Nonlinear classification

- Implement and test:
  - Kernel Logistic regression

$$\arg\min_{\{\alpha_j\},b} \sum_{i=1}^{m} \ell(y_i(\sum_{j=1}^{m} \alpha_j y_j K[j,i] + b)) + \sum_{j=1}^{m}\sum_{k=1}^{m} \alpha_j y_j \alpha_k y_k K[j,k].$$

    - *loss*(yh(x))=ln(1+exp(-yh(x)))

  - Semi-supervised Kernel logistic regression

$$\min_{c\in\mathbb{R}^m,b} C\sum_{i=1}^{m} \ell(x_i, y_i, \sum_{j=1}^{m} c_j K(x_j, x_i), b) + \frac{1}{2}c^T K c + c^T K L_G K c$$

    - *loss*(x,y,h,b)=ln(1+exp(-y (h(x)+b)))

The same

Kernel and Graph

- Use RBF Kernel (Gaussian kernel):
  - K is a [samples x samples] matrix
  - K[i,j]=K(x_i, x_j)
  - K(x,y)=exp(-(||x-y||^2))

- Laplacian L is defined as a [samples x samples matrix]:

$$G_{j,k} = \mathrm{e}^{-\frac{||x_j - x_k||^2}{t}}$$

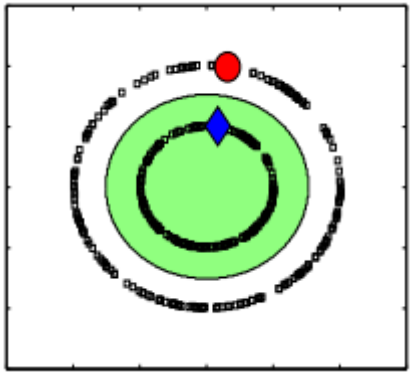
$$L_G = D_G - A_G$$
$$A_G(j,k) = G_{jk}$$
$$D_G(j,j) = D_j = \sum_{k=1}^{F} G_{jk}$$
$$D_G(j,k \neq j) = 0$$

$$\min_{c \in \mathbb{R}^m, b} C \sum_{i=1}^{m} \ell(x_i, y_i, \sum_{j=1}^{m} c_j K(x_j, x_i), b) + \frac{1}{2} c^T K c + c^T K L_G K c$$

Experiments

- Compare:
  - Linear LR (from HW2) on full dataset
  - Kernel LR on full dataset
  - Semi-supervised Kernel LR (only X% labels known, but all samples used for kernel/laplacian)
  - Kernel LR on reduced dataset (only X%, same as above)

  - Try different values of X (e.g. 10%, 5%, …)



Returning the Assignment

- Solution code should be written by you and you only (no web/book/friend/etc. code)
  - You can freely use the code provided on BB as your starting point

- Upload through Blackboard
  - A report in PDF
    - Results of your experiments (include you V#, and what are your two digits defining the two-class problem).
  - Code in python:
    - FamilyNameFirstName-KLR.py (kernel logistic regression)
    - FamilyNameFirstName-SKLR.py (semi-supervised KLR)
      - The files should have your name in a comment at the top