University of Tehran

School of Electrical and Computer Engineering

Pattern Recognition

Assignment 5

Due Date: 7th of Khordad

Corresponding TAs:

Hamidreza Aftabi (Q1, Q2, Q5) – hamid.aftabi@gmail.com

Salar Nouri (Q3, Q4, Q5) – salar.nouri74@gmail.com

Spring 98

PROBLEM 1

L2 Norm SVMs[۱]

In this assignment another procedure for solving SVMs will be discussed. This approach, known as the l_2 norm SVM, is given by the following minimization problem (be informed that the slack variables are squared in this assignment compared with the l_1 norm SVMs which was discussed in the class):

$$\min_{\omega,b,\xi} \frac{1}{2}\|\omega\|^2 + \frac{C}{2}\sum_{i=1}^{m}\xi_i^2$$

$$s.t.\ y^{(i)}(\omega^T x^{(i)} + b) \ge 1 - \xi_i,\ \ i = 1,\ldots,m$$

(a) As you can see, we have dropped the $\xi \ge 0$ constraint in the l_2 problem. Show that the optimal value of the objective will be the same whether or not this constraint is present.

(b) Write the *Lagrangian* of the l_2 soft margin SVM.

(c) Minimize the *Lagrangian* with respect to ω, b and ξ by the taking the following gradients: $\nabla_\omega L, \frac{\partial L}{\partial b}, \nabla_\xi L$ and then setting them equal to 0. Here, $\xi = [\xi_1, \xi_2, \ldots, \xi_m]^T$.

(d) Write the dual of the l_2 norm SVM optimization problem.

PROBLEM 2

Hinge Loss[۲]

Another approach for solving SVMs is minimizing hinge loss. This is a loss function used for training classifiers. The hinge loss is used for maximum-margin classification, most notably for SVMs. The hinge loss is defined as:

$$C(x, y, f(x)) = (1 - y * f(x))_+$$

C is the loss function, x is the sample, y is the true label and f(x) is the predicted label. We can rewrite C as:

---

[۱] https://scholarscompass.vcu.edu/cgi/viewcontent.cgi?article=5248&context=etd

[1] http://www2.kobe-u.ac.jp/~abe/pdf/SVMsforPC.pdf

[۲] http://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf

$$C(x, y, f(x)) = \begin{cases} 0 & y * f(x) \geq 1 \\ 1 - y * f(x) & else \end{cases}$$

Now we can write objective of SVM which consists of two terms:

$$\min_{\omega} \lambda \|\omega\|^2 + \sum_{i=1}^{n} (1 - y_i < x_i, \omega >)_+$$

The first term is a regularizer and the second one is the hinge loss. The regularizer balances between margin maximization and loss.

The dataset for this assignment is **Haberman's Survival Dataset** which contains 306 cases from a study that was conducted at the University of Chicago's Billings Hospital on the survival of patients who had undergone surgery for breast cancer. For more information visit https://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival.

(a) Use Scikit-Learn' PCA to reduce the dimension of data to 2-D.

(b) Plot the samples of each class. Are these 2 classes linearly separable? If not use a nonlinear function that maps 2-D feature space into 3-D feature space to make two linearly separable classes.

$$\begin{pmatrix} x_1 \\ x_2 \\ \varphi(x_1, x_2) \end{pmatrix} \rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

(c) Find the weight of the decision boundary by minimizing the SVM objective function using gradient descent algorithm and plot the results (change α and λ in order to get the best results).

$$\omega^+ = \omega^- - \alpha \frac{\partial J}{\partial \omega}$$

(d) Now, repeat part (c) by using Scikit-Leran's SVM and compare the results.

PROBLEM 3

Support Vector Regression and Logistic Regression[3]

1. In this problem, at first we study Support Vector Regression (SVR) which works on Classification. We have training data $(x_i, y_i)_{i=1}^n$ and wish to find f , such that $f(x) \approx y$.

---

[2] https://pdfs.semanticscholar.org/e923/9469aba4bccf3e36d1c27894721e8dbefc44.pdf Chapters 4.2, 4.3 and 10

[3] https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12.pdf

For **linear SVR**, we take $f(x)=w^T x$. The loss function in SVR is the deadzone linear penalty $l_\varepsilon = \max(0,|u|-\varepsilon)$. Accordingly, the SVR cost function is,

$$J(w)=\frac{1}{2}\|w\|^2 + c\sum l_\varepsilon(w^T x_i - y_i)$$

Here the $\frac{1}{2}\|w\|^2$ terms acts as a regularizer. We will drop an intercept term (i.e. $f(x)=w^T x+b$) here for simplicity.

(a) Write down the dual problem as a Quadratic Programming, by following the same procedure as SVM.
(b) Develop a kernelized version of the algorithm, using dual problem. Specify how you may obtain the prediction for a new point x.

2. Second, **Logistic Regression** is a method of fitting a probabilistic classifier that gives soft linear thresholds. It is common to use logistic regression with an objective function consisting of the negative log probability of the data plus a L_2 regularizer:

$$L(W)=-\sum_{i=1}^{N}\log\left(\frac{1}{1+e^{y_i(W^T x_i+b)}}\right)+\lambda\|w\|_2^2$$

(Here **w** does not include the "extra" weight w_0.)

(a) Find the partial derivatives $\frac{\partial L}{\partial w_j}$ and partial second derivatives $\frac{\partial^2 L}{\partial w_j \partial w_k}$.
(b) Show that L (**w**) is a convex function, with using partial derivatives results.

Hint: A function L is convex if it is **Hessian**. A matrix **H** is PSD if and only if
$a^T Ha \equiv \sum_{j,k} a_j a_k H_{j,k} \geq 0$ **;** for all real vectors **a**.

PROBLEM 4

Positive Definiteness of Gaussian Kernel and SVM[۴]

(a) In this part, you will show that the Gaussian kernel $k: R^d \times R^d \to R$ defined by

$k(x,x')=\exp(-\delta\|x-x'\|^2), \delta>0$, is a valid kernel.

[۴] https://pdfs.semanticscholar.org/e923/9469aba4bccf3e36d1c27894721e8dbefc44.pdf Chapters 4.2, 4.3 and 5.

If k_1, k_2 are positive definite kernels, show that $\alpha k_1 + \beta k_2$ is a positive definite kernel for $\alpha, \beta \geq 0$ and $k_1 k_2$ is a positive definite kernel too.

(b) Now, as we know decision function learned by SVM can be written as $f(x) = \sum_{i=1}^{m} \alpha_i y^{(i)} K(x^{(i)}, x) + b$. For the training data $\{(x^{(i)}, y^{(i)})\}$ which are separated by at least a distance of ε :($\left\| x^{(j)} - x^{(i)} \right\| > \varepsilon$). Find values for the weighted parameters $\{\alpha_i, b\}$ and Gaussian kernel width, such that for all $i = 1, \ldots, m$, $x^{(i)}$ is correctly classified.

PROBLEM 5

A) Poly and Linear Kernels[δ]

Use polynomial kernel with appropriate order to implement a soft margin SVM classifier for the given dataset, concerning two different approaches (one class against one class, and one class against all classes). Order of your polynomial is important since it might affect the generalizability of your model. Generally, for choosing the optimum parameter of your classifiers in any case, you can use GridSearchCV (please peruse the Scikit-learn instruction and find out about its attributes and parameters).

1- Compare the results of the two approaches with each other concerning the CCR (Correct Classifier Rate) on the test data, and the time required for training and testing the dataset.

2- Explain completely advantages and disadvantages of each approach.

3- Use linear SVM with appropriate value of parameter "C". Compare the performance of the linear SVM with your results in section (1) using the polynomial kernel.

B) RBF kernel

Test the effect of parameter tuning for RBF-kernel SVM. You need to tune the values of the parameters C and γ using GridSearchCV. What is the best accuracy you get?

Note: Use **Reduced Fashion-MNIST Dataset** for this assignment which is available at CECM.

[δ] https://scikit-learn.org/stable/modules/svm.html

NOTES

1. Please make sure you reach the deadline because there would be no extra time available.
2. Late policy would be as bellow:
   - Every student has a budget for late submission during the semester. This budget is two weeks for all the assignments.
   - Late submission more than two weeks may cause lost in your scores.
3. Analytical problems can be solved on papers and there is no need to type the answers. The only thing matters is quality of your pictures. Scanning your answer sheets is recommended. If you are using your smartphones you may use scanner apps such as CamScanner or google drive application.
4. Simulation problems need report as well as source codes and results. This report must be prepared as a standard scientific report.
5. You have to prepare your final report including the analytical problems answer sheets and your simulation report in a single pdf file.
6. Finalized report and your source codes must be uploaded to the course page as a ".zip" file (not ".rar") with the file name format as bellow:

   PR_Assignment #[Assignment Number]_Surname_Name_StudentID.zip

7. Plagiarisms would be strictly penalized.
8. You may ask your questions from corresponding TAs.