CS181-S24

Assignment #4
Due: 11:59pm March 29, 2024

Collaborators:

Homework 4: SVM, Clustering, and In-Depth Ethics

Introduction

This homework assignment will have you work with SVMs, clustering, and engage with the ethics lecture.

Resources and Submission Instructions

We encourage you to read Chapter 5 in the textbook to learn more about SVMs, 6.2 to review k-means clustering, and 6.3 to review HAC.

Please submit the **writeup PDF to the Gradescope assignment 'HW4'**. Remember to assign pages for each question.

Please submit your **LaTeX file and code files to the Gradescope assignment 'HW4 - Supplemental'**.

You can use a **maximum of 2 late days** on this assignment. Late days will be counted based on the latest of your submissions.

Problem 1 (Fitting an SVM by hand, 10pts)

For this problem you will solve an SVM by hand, relying on principled rules and SVM properties. For making plots, however, you are allowed to use a computer or other graphical tools.

Consider a dataset with the following 7 data points each with $x \in \mathbb{R}$ and $y \in \{-1, +1\}$:

$$\{(x_i, y_i)\}_{i=1}^{7} = \{(-3, +1), (-2, +1), (-1, -1), (0, +1), (1, -1), (2, +1), (3, +1)\}$$

Consider mapping these points to 2 dimensions using the feature vector $\phi(x) = (x, -\frac{8}{3}x^2 + \frac{2}{3}x^4)$. The hard margin classifier training problem is:

$$\begin{aligned} &\min_{\mathbf{w}, w_0} \frac{1}{2}\|\mathbf{w}\|_2^2 \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \phi(x_i) + w_0) \geq 1, \ \forall i \in \{1, \ldots, n\} \end{aligned}$$

Make sure to follow the logical structure of the questions below when composing your answers, and to justify each step.

1. Plot the transformed training data in $\mathbb{R}^2$ and draw the optimal decision boundary of the max margin classifier. You can determine this by inspection (i.e. by hand, without actually doing any calculations).
2. What is the value of the margin achieved by the optimal decision boundary found in Part 1?
3. Identify a unit vector that is orthogonal to the decision boundary.
4. Considering the discriminant $h(\phi(x); \mathbf{w}, w_0) = \mathbf{w}^\top \phi(x) + w_0$, give an expression for *all possible* $(\mathbf{w}, w_0)$ that define the decision boundary. Justify your answer.
5. Consider now the training problem for this dataset. Using your answers so far, what particular solution to $\mathbf{w}$ will be optimal for the optimization problem?
6. What is the corresponding optimal value of w_0 for the $\mathbf{w}$ found in Part 5 (use your result from Part 4 as guidance)? Substitute in these optimal values and write out the discriminant function $h(\phi(x); \mathbf{w}, w_0)$ in terms of the variable x .
7. Which points could possibly be support vectors of the classifier? Confirm that your solution in Part 6 makes the constraints above tight—that is, met with equality—for these candidate points.
8. Suppose that we had decided to use a different feature mapping $\phi'(x) = (x, -\frac{31}{12}x^2 + \frac{7}{12}x^4)$. Does this feature mapping still admit a separable solution? How does its margin compare to the margin in the previous parts? Based on this, which set of features might you prefer and why?

Solution

Problem 2 (K-Means and HAC, 20pts)

For this problem you will implement K-Means and HAC from scratch to cluster image data. You may use `numpy` but no third-party ML implementations (eg. `scikit-learn`).

Your job is to implement K-means and HAC on MNIST, the collection of handwritten digits that you saw in HW3, and to test whether these relatively simple algorithms can cluster similar-looking images together.

The code in `homework4.ipynb` loads the images into your environment into two arrays – `large_dataset`, a 5000x784 array, will be used for K-means, while `small_dataset`, a 300x784 array, will be used for HAC. In your code, you should use the ℓ_2 norm (i.e. Euclidean distance) as your distance metric.

Important: Remember to include all of your plots in your PDF submission!

1. Starting at a random initialization and $K = 10$, plot the K-means objective function (the residual sum of squares) as a function of iterations and verify that it never increases.

2. For $K = 10$ and for 3 random restarts, print the mean image (aka the centroid) for each cluster. There should be 30 total images.

3. Repeat Part 2, but before running K-means, standardize or center the data such that each pixel has mean 0 and variance 1 (for any pixels with zero variance, simply divide by 1). For $K = 10$ and 3 random restarts, show the mean image (centroid) for each cluster. Again, present the 30 total images in a single plot. Compare to Part 2: How do the centroids visually differ? Why?

4. Implement HAC for min, max, and centroid-based linkages. Fit these models to the `small_dataset`. For each of these 3 linkage criteria, find the mean image for each cluster when using 10 clusters. Display these images (30 total) on a single plot.

   How do the "crispness" of the cluster means and the digits represented compare to mean images for k-means? Why do we only ask you to run HAC once?

   Important Note: For this part ONLY, you may use `scipy`'s `cdist` function to calculate Euclidean distances between every pair of points in two arrays.

5. For each of the HAC linkages, as well as one of the runs of your k-means, make a plot of "Number of images in cluster" (y-axis) v. "Cluster index" (x-axis) reflecting the assignments during the phase of the algorithm when there were $K = 10$ clusters.

   Intuitively, what do these plots tell you about the difference between the clusters produced by the max and min linkage criteria?

   Going back to the previous part: How does this help explain the crispness and blurriness of some of the clusters?

Problem 2 (cont.)

6. For your K-means with $K = 10$ model and HAC min/max/centroid models using 10 clusters on the `small_dataset` images, use the `seaborn` module's `heatmap` function to plot a confusion matrix between each pair of clustering methods. This will produce 6 matrices, one per pair of methods. The cell at the ith row, jth column of your confusion matrix is the number of times that an image with the cluster label j of one method has cluster i in the second method. Which HAC is closest to k-means? Why might that be?

7. Let's return to the postal service example from HW3. Do you think that clustering is a good way to identify digits, that is, first cluster the data, and then, for any new data point, classify it based on its cluster?

   (a) In particular, do you expect the clusters to correspond with the true labels? Is that a good way to evaluate clustering?

   (b) In the context of adversaries, how might clusterings be attacked? How is the process similar or different to the process for attacking classification models (like kNN and the generative classifiers that we saw in HW2)?

Solution

Problem 3 (Ethics Assignment, 5pts)

Consider the simulation we conducted in class. There you were part of a health-care administration company, Optimizing Health. Optimizing Health is a small startup and you were a developer reporting directly to the CTO. In particular, you were responsible for designing a system that predicts the healthcare needs of an individual. First, you needed to choose a way to predict healthcare needs. You chose between approximating need by:

- the predicted healthcare expenditure of that individual
- the predicted number of hours an individual is expected to visit the doctor's office.
- the number of diagnoses the patient has received in the past year.
- the average prognosis for individuals in that patient's age and racial groups.

As an organization, you settled on predicting healthcare need by predicted healthcare expenditure of that individual given their diagnosis.

Then, you used that predicted healthcare expenditure to assign a "Risk Score" to all patients, and, for patients above a certain cutoffs, you either automatically enrolled them in the a new program you created called the Medical Management Program (high risk) or referred them to their physician (moderate risk).

You discovered that, at a given Risk Score, Black patients are sicker than White patients. The algorithm was therefore less likely to recommend additional care to Black patients who needed it than White patients. 17.7% of Black patients were being recommended for additional care, while 46.5% of Black patients required additional care. After these recommendations were used, doctors provided disparate treatment, with Black patients not entering the Medical Management Program as frequently as they require.

All of these (and many unstated but salient) steps can be modeled using causal chains, backward looking responsibility, and forward looking responsibility (as we did in class).

After the terrible outcome, the CEO resigns and you are promoted to lead the organization! This means that any actions you take will be implemented and so you vow to do better.

The first thing you do is decide to rewrite the mission: "Optimizing Health's new mission is to provide care to individuals who need it but are not currently receiving it, without increasing healthcare disparities across groups."

With this new goal in mind, think through each important choice-point for success and outline (literally draw) a new causal chain (similar to what you did in Scenario Part 4 in class). This time, however, you can recommend choices other than those outlined in the in-class Scenario and above.

A sufficiently descriptive causal chain will include points that link the mission to, for instance, the choice to use ML, data collection, feature selection, ... , assignment of risk, what to do with risk scores, and so on. Remember to represent actual decision points as ovals and everything else (things that follow directly from decisions) as rectangles. Moreover, flag the decisions that carry moral or ethical responsibility to distinguish them from decisions that do not. At each point in the chain, if that point involves an ethically salient decision, provide your brief rationale for the decision, a brief description of the main argument against it, and your response to that objection. List the individuals who would be morally responsible at each contributing choice point. Finally, briefly describe how the outcome (the last node in your chain) meets the mission objective.

Solution

Name

Collaborators and Resources

Whom did you work with, and did you use any resources beyond cs181-textbook and your notes?