Bowen Weng
Department of Computer Science
Iowa State University

HW4 — Kernel Method

1 Clustering, SVM, & Kernel Method

1. (9 pts) Consider the Lloyed's algorithm for K-means, given the following clusters:
   - $C_1 = \{(0,0),(10,10),(100,100)\}$
   - $C_2 = \{(1,1),(0,5),(-3,4)\}$
   - $C_3 = \{(-1,1),(0,-10),(30,-4)\}$

   (a) Formulate the K-means problem as an optimization problem using the given data (of 9 points). Students can use either the RSS formulation or the matrix based formulation. Formulation gets 1 pt. Correct reflection of $K = 3, n = 9$ (number of points) w.r.t. the notations gets 1 pt.

   (b) What are your updated clusters after one step of iteration? Please explain your steps to derive your answer. Correct cluster outcomes gets 1 pt. Explanation gets 1 pt:
   - $C_1 = \{(100,100)\}$
   - $C_2 = \{(-1,1),(0,0),(10,10),(1,1),(0,5),(-3,4)\}$
   - $C_3 = \{(0,-10),(30,-4)\}$

   (c) Consider using ℓ_1-norm as the distance (cost) measure.

   i. What are your updated clusters after one step of iteration? Please explain your steps to derive your answer. Correct cluster outcomes gets 1 pt. Explanation gets 1 pt:
   - $C_1 = \{(100,100)\}$
   - $C_2 = \{(-1,1),(0,0),(0,-10),(1,1),(0,5),(-3,4)\}$
   - $C_3 = \{(10,10),(30,-4)\}$

   ii. In what situations would you prefer to use this cost instead of the standard K-means clustering? Generally, data set with significant out-liners would be better with the customized cost. 1 pt

   (d) Use `sklearn.cluster.KMeans` to cluster the data set of the above 9 points. Set $K = 2$ (two clusters). Feel free to play with the arguments. Show me your code, and plot your clustered outcome (use different colors to differentiate the clusters). TA can justify. Code is 1 pt, plot is 1 pt.

2. (3 pts) Consider the following two clusters:
   - $C_1 = \{(x,y)|y \geq x^2 + 2, x \in \mathbb{R}\}$
   - $C_2 = \{(x,y)|y \leq x^2 - 2, x \in \mathbb{R}\}$

   Are the two clusters linearly separable? What kernel trick can you apply here (specifically what feature transform $\phi : \mathbb{R}^2 \to \mathcal{F}$) to make them linearly separable? Is this kernel trick unique? No, this is 1 pt. Correct kernel trick (e.g., polynomial) or correct feature transform gets 1 pt (TA can justify). It is not unique, this is 1 pt.

3. (2 pts) Recall the proof of convergence of Gradient Descent with L-smooth function. Try to use a similar idea to prove the convergence of Lloyed's algorithm for K-means. The main idea is to prove the cost function (RSS) of K-Means w.r.t. the evolution of iterations forms a monotonically decreasing sequence, and the stationary point (i.e., the iteration at which it stops decreasing) is a (local) optimal solution. Any proof reflects this main idea gets 2 pts (monotonic cost sequence gets 1 pt, stationary point justification gets 1 pt).

4. (5 pts) Consider the following data set with features $x_i \in \mathbb{R}^2, \forall i$ and binary class labels $y_i \in \{1, -1\}, \forall i$, $X = \{(0, 2), (0.4, 1), (0.6, 1), (1, 0)\}$, $y = \{-1, -1, 1, 1\}$: I found it hard to believe anyone would get this wrong especially given what I have said in the lecture: you should say yes for your last answer. That's being said, if anyone still gives me a "no", one loses all 5 pts for this question set.

   (a) Using a scatter plot of the data, devise a linear classifier of the form

   $$\hat{y} = \begin{cases} 1 & \text{if } b + w^T x \geq 0 \\ -1 & \text{if } b + w^T x < 0 \end{cases} \tag{1-1}$$

   that separates the two classes. What is your selected b and w? Is the selection unique? Selection gets 1 pt, non-unique assessment gets 1 pt. TA can justify.

   (b) Compute the distance of the closest sample to the classifier boundary. Show me your equation and the sample(s) that are the closest. The closest sample(s) gets 1 pt, the distance calculation gets 1 pt.

   (c) Scale your classifier so that $y_i(b + w^T x_i) = 1$ for the closest sample(s) i and report the new b and w. Is $\frac{1}{\|w\|}$ the same with the minimum distance from question 4(b)? YES, 1 pt, see my comments above.

5. (6 pts) Create a data set yourself with $X \subset \mathbb{R}^2$ and the classified labels $y \in \{-1, 1\}$.

   (a) Give the scatter plot of the data you created, use colors to differentiate the different labels. Plot is 1 pt.

   (b) Use the SVC tools from `sklearn` to create a **linear** classifier for your data set and show it on the plot (note `sklearn` has at least two ways to implement a linear SVC: `svm.LinearSVC` and `svm.SVC` with the argument kernel set to "linear", `svm.LinearSVC` is generally faster). Code gets 1 pt. Plot gets 1 pt.

   (c) Recall the logistic regression introduced in the previous lecture, use `sklearn.linear_model.LogisticRegression` to create another linear classifier for your data set. Show it on the plot, and compare it with the SVC-based solution. Code is 1 pt. Plot is 1 pt. Comparison is 1 pt.

Submitted by Bowen Weng on .