Bowen Weng
Department of Computer Science
Iowa State University

HW4 — Kernel Method

1 Clustering, SVM, & Kernel Method

1. 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).

   (b) What are your updated clusters after one step of iteration? Please explain your steps to derive your answer.

   (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.

   ii. In what situations would you prefer to use this cost instead of the standard K-means clustering?

   (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).

2. 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?

3. 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.

4. Consider the following data set with features $x_i \in \mathbb{R}^2, \forall i$ and binary class labels $y_i \in \{1,-1\}, \forall i$, $X = \{(0,2),(0.4,1),(0.6,1),(1,0)\}$, $y = \{-1,-1,1,1\}$:

   (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?

   (b) Compute the distance of the closest sample to the classifier boundary. Show me your equation and the sample(s) that are the closest.

   (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)?

5. Create a data set yourself with $X \subset \mathbb{R}^n$ 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.

(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).

(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.

Submitted by Bowen Weng on .