CSE 575: Statistical Machine Learning: Mid-Term 2

Instructor: Prof. Hanghang Tong

March 28th, 2017

First Name:			
Last Name:			
Email:			
ASU ID:			
Q	Topic	Max Score	Score
1	SVM	30	
2	Kmeans	30	
3	GMM and EM	20	
4	Mincut	20	
Total:		100	

- This exam book has **10** pages, including this cover page and 1 blank page at the end.
- You have 75 minutes in total.
- Good luck!

1 Support Vector Machines [30 points]

Given the following dataset in 1-d space, which consists of 2 positive data points at the following coordinates $\{-1, -3\}$ and 3 negative data points at the following coordinates $\{1, 3, 4\}$. Suppose we use a soft-margin SVM **without** kernel, i.e., $f(x) = w \cdot x + b$. Recall that in the soft-margin SVM classifier, we aim to min $\frac{1}{2} w \cdot w' + C \sum_{i=1}^{n} \epsilon_i$ with some additional constraints, where ϵ_i $(i = 1, ..., n)$ is the slack variable for the i^{th} training example, n is the number of training examples, and C is regularization parameter.

1. [4 points] In this SVM classifier, how many independent parameters are there (2 points)? What are they (2 points)?

   Sol: 2 (2 pts). w and b (1 pt each).

2. [4 points] Draw the dataset in 1-d space.

   Sol:

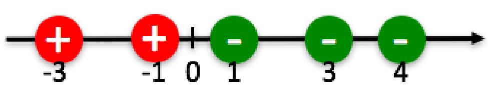

Figure 1

3. [4 points] **[True or False]**. With a larger C, the resulting classifier favors more on the size of the classier margin, compared with the training error. This is because the slack variable ϵ_i approximates the training error.

   Sol: F

4. [6 points] Suppose $C = \infty$. Draw the decision boundary of your SVM classifier (2 points). How many support vectors are there (2 points)? What are they (2 points)?

   Sol: at the origin. 2 support vectors. -1 and 1.

5. [4 points] Suppose $C = \infty$. What is your SVM classifier (2 points)? (i.e., what are w and b, respectively?) What is the size of the margin of your classifier (2 points)?

   Sol: Margin is 2 (based on geometry). decision boundary: $x = 0$. Since the margin is $2/\|w\|$, we have w=1. $b = 0$ (based on geometry).

6. [4 points] Suppose $C = 0$. What is w in your SVM classifier (2 points)? What is the size of the margin of your classifier (2 points)?

   Sol: $w = 0$, and margin is infinity.

7. [4 points] **[True or False]**. Now add an additional positive example at $\{2\}$. If we use the same soft-margin SVM formulation described as above, we might end up with a **non-linear** classifier, depending on the specific value of C.

   Sol: F

2 Kmeans [30 points]

Given N data points x_i $(i = 1, ..., N)$, Kmeans will group them into K clusters by minimizing the distortion function $J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{n,k} \|x_n - \mu_k\|^2$, where μ_k is the center of the k^{th} cluster; and $r_{n,k} = 1$ if x_n belongs to the k^{th} cluster and $r_{n,k} = 0$ otherwise. In this question, we will use the following iterative procedure.

- Initialize the cluster center μ_k $(k = 1, ..., K)$;
- Iterate until convergence
  - Step 1: Update the cluster assignments $r_{n,k}$ for each data point x_n.
  - Step 2: Update the center μ_k for each cluster k.

1 [4 points]. Given 6 data points in 1-d space: $x_1 = -3$, $x_2 = -1$, $x_3 = 0.5$, $x_4 = 2$, $x_5 = 3$, $x_6 = 4$. Plot these six data points in 1-d space.

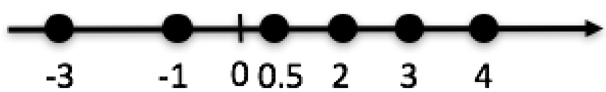

Figure 2: input data points

2 [8 points] Suppose the initial cluster centers are $\mu_1 = -1$ and $\mu_2 = 3$ (i.e., $K = 2$). If we run Kmeans on the above dataset only for one iteration, what is the cluster assignment for each data point after Step 1 (4 points)? What are the updated cluster centers after Step 2 (4 points)?

sol: the left 3 belong to the first cluster; and the right three to the second cluster. $\mu_1 = -7/6$ and $\mu_2 = 3$.

3 [10 points] Now, suppose the initial cluster centers are $\mu_1 = -3$, $\mu_2 = 0$ and $\mu_3 = 3$ (i.e., $K = 3$). If we run Kmeans on the above dataset only for one iteration, what is the cluster assignment for each data point after Step 1 (4 points)? What are the updated cluster centers after Step 2 (3 points)? What is the distortion function J after this iteration (3 points)?

sol: $\{-3\}$, $\{-1, 0.5\}$ and $\{2, 3, 4\}$ $\mu_1 = -3$, $\mu_2 = -0.25$, $\mu_3 = 3$. $J = 0 + (\frac{3}{4})^2 + (\frac{3}{4})^2 + 1 + 0 + 1 = 3.125$

4 [8 points] If we run Kmeans on the above dataset to find six clusters (i.e., $K = 6$). What is the **optimal** cluster assignment for each data point (4 points)? What is the corresponding distortion function J for the optimal clustering assignment (4 points)?

sol: each data point forms its own clusters. $J = 0$.

3 GMM and EM Algorithm [20 points]

Given four data points in 1-d space, i.e., $x_1 = 4, x_2 = 2, x_3 = 1, x_4 = 0$, we want to use GMM (Gaussian Mixture Models) and EM algorithm to find two clusters. Suppose the covariance of each Gaussian component is known and fixed. Recall that in M-step of the EM algorithm, we want to update the model parameters, including (a) $P(\mu = \mu_j)$ and (b) μ_j, where $\mu_j,\ (j = 1, 2)$ is the mean (i.e., the center) of each Gaussian component. Suppose in the previous step of EM algorithm, we have the following estimation for $E(z_{i,j}) = P(\mu_j|x_i)$ $(i = 1, 2, 3, 4;\ j = 1, 2)$: $E(z_{1,1}) = 1$, $E(z_{1,2}) = 0$, $E(z_{2,1}) = 1$, $E(z_{2,2}) = 0$, $E(z_{3,1}) = 0.5$, $E(z_{3,2}) = 0.5$, $E(z_{4,1}) = 0$, $E(z_{4,2}) = 1$.

1. [4 points]. Plot the data points in 1-d space.

   Solution: four points at 4, 2, 1, 0, respectively.

2. [8 points]. Put all the $E(z_{i,j})$ numbers into a 4×2 table, whose rows correspond to different data points (x_1, x_2, x_3, x_4) and columns correspond to two different clusters/Gaussian components (μ_1, μ_2).

   Solution:

1	0
1	0
0.5	0.5
0	1

3. [4 points]. After the M-step, what is the new estimation of $P(\mu = \mu_1)$ (2 points)? What is the new estimation of $P(\mu = \mu_2)$ (2 points)?

   Solutions: $2.5/4 = 5/8$ and $1.5/4 = 3/8$ (2 pts each).

4. [4 points]. After the M-step, what is the new estimation of μ_1 (2 points)? What is the new estimation of μ_2 (2 points)?

   Solutions: $\mu_1 = (4 + 2 + 1 \times 0.5)/(1 + 1 + 0.5) = 2.6$, $\mu_2 = (0.5 \times 1 + 0 \times 1)/1.5 = 1/3$ (2 pt each)

4 Mincut [20 points]

Given a graph with 6 nodes (i.e., data points) in Figure 3 where the weight for each edge/link is *1*. We want to run *MinCut* algorithm to find two clusters.

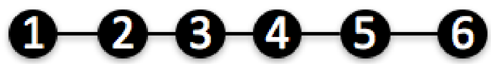

Figure 3: The Input Graph

1. [4 points] Write down the adjacency matrix W of this graph.

   Sol:

$$\mathbf{W} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

2. [4 points] Write down the graph Laplacian matrix L of the adjacency matrix of this graph.

   Sol:

$$\mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

3. [6 points] Suppose we have the following partion/clustering result: $\{1, 2, 3\}$ as partition-1 and $\{4, 5, 6\}$ as partition-2. What is the corresponding clustering membership vector q (3 points)? What is the *cutsize* for this partition result (3 points)?

   Sol: $q = [1,\ 1,\ 1,\ -1,\ -1,\ -1]$ (scaling the numbers and/or fliping the sign is fine). cutsize is 2 (the edge 3-4 and edge 4-3 are both counted. it is fine if the answer is 1).

4. [6 points] Now, suppose we have the following partion/clustering result: $\{1\}$ as partition-1 and $\{2, 3, 4, 5, 6\}$ as partition-2. What is the corresponding clustering membership vector q (3 points)? What is the cutsize for this partition result (3 points)?

   Sol: $q = [1,\ -1,\ -1,\ -1,\ -1,\ -1]$ (scaling the numbers and/or fliping the sign is fine). cutsize is 2 (the edge 1-2 and edge 2-1 are both counted. it is also fine if the answer is 1).

BLANK PAGE