

## Data Mining Quiz February 18, 2016

1. What is the difference between *supervised* and *unsupervised* learning? Name and describe one example of each type of learning.
2. Suppose we want to perform hierarchical clustering on the dataset  $B$  below. In this case we have 5 instances and one feature. Using the Euclidean distances find the  $5 \times 5$  distance matrix  $d(x_i, x_j)$ . Use the distance matrix to perform hierarchical clustering for both *single* and *complete* linkages. How is the result different between these two cases?

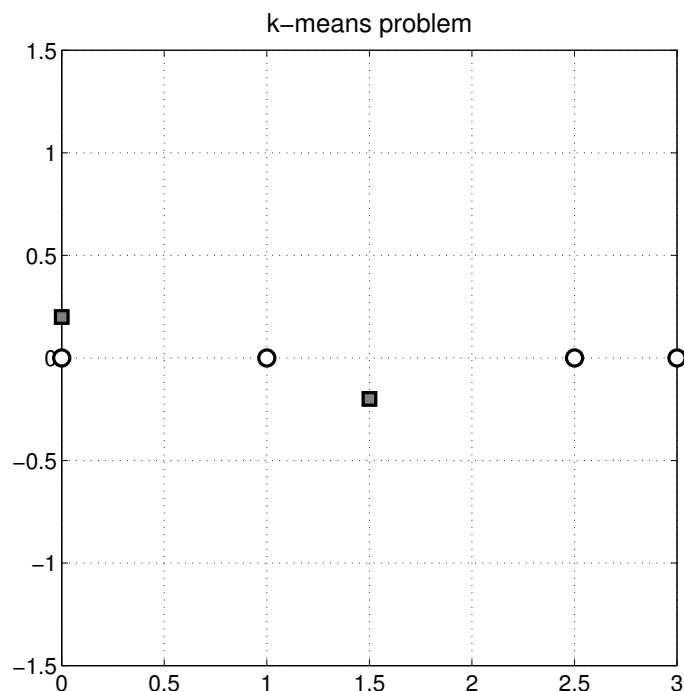
$$\mathbf{B} = \begin{bmatrix} -1 \\ 2 \\ 6 \\ 11 \\ 12 \end{bmatrix} \tag{1}$$

3. Specify the *regularization* terms that appear in both LASSO and Ridge Regression.

4. Run the  $k$ -means algorithm on the dataset in the figure below until convergence. The figure is titled “ $k$ -means problem”. Draw the cluster centers at convergence on the plot. Assume that the circles represent the data instances and the squares represent the initial cluster centers (in this example  $k = 2$ ). Be sure to answer how many iterations it takes for the algorithm to converge.

For convenience the data is also given in matrix form.  $X$  are the data instances. The initial cluster centers are  $[0, 0.2]$  and  $[1.5, -0.2]$

$$X = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 2 & 0 \\ 2.5 & 0 \\ 3 & 0 \end{bmatrix} \quad (2)$$



5. A binary classification problem is shown in the figure titled “SVM Question”. Suppose you are given the following possible solution to the linear SVM problem for this data:  $\alpha = [0, 1, 3/2, 0]$  and a bias term of 1. Find the classification function for this possible solution and use it to classify the origin (i.e. you must first determine the coefficients  $w$ ). Answer the following, which of the data instances are support vectors and do you think this represents an optimal solution, why or why not.

$$\mathbf{X} = \begin{bmatrix} -2 & 3 \\ -1 & 2 \\ 1 & 0 \\ 2 & 0 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \quad (3)$$

