

**HW for 2019-06-12**

(due: 2019-06-24)

**1: Multi-class learning** In class, we discussed multi-class learning using the following method: let  $x^1, \dots, x^k$  be indicators for the training data assigned to classes 1 through  $k$ , and compute soft indicators  $y^1, \dots, y^k$  for the unlabeled data by solving

$$y^j = -L_{22}^{-1}L_{21}x^j.$$

Argue that if each training example is assigned to exactly one data item, then for each  $i$  we must have  $\sum_{j=1}^k y_i^j = 1$ .

*Hint:* Use that the vector of all ones is a null vector for  $L$ .

**2: Label by first encounter** Consider the following alternate method of semi-supervised learning: for any unlabeled node  $i$ , we start a random walker; we assign to node  $i$  the label of the first training example encountered. If our data is at the first group of nodes and our unlabeled examples are in the second group, we can express this in terms of the *absorbing Markov chain* with transition matrix

$$P = \begin{bmatrix} I & P_{12} \\ 0 & P_{22} \end{bmatrix} = \begin{bmatrix} I & A_{12}D_{22}^{-1} \\ 0 & A_{22}D_{22}^{-1} \end{bmatrix}.$$

This Markov chain has many stationary states. Which one do we reach starting from node  $j$  of the unlabeled set? That is,

$$P^k \begin{bmatrix} 0 \\ e_j \end{bmatrix} \rightarrow \begin{bmatrix} \pi^j \\ 0 \end{bmatrix}.$$

What is  $\pi^j$ ? Show that if the training examples are labeled according to a zero-one vector  $x$ , then  $y_j = (\pi^j) \cdot x$  (the expected value of the first label encountered) is exactly the value of the soft label discussed in class. That is, this is another explanation of the same Laplacian-based soft-labeling scheme we saw before.

Note: You should assume that all nodes are connected, so that the largest eigenvalue of  $P_{22}$  is less than one.