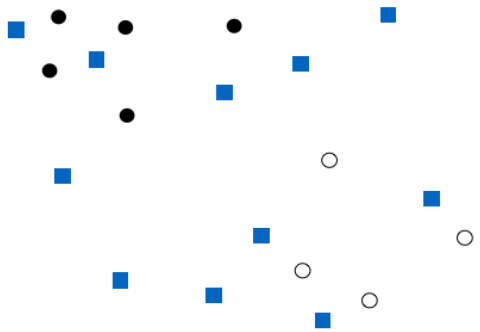
Semi-supervised Learning Project

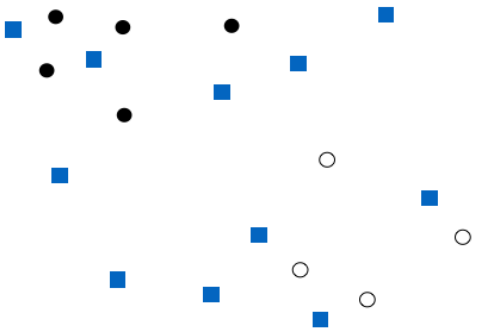
Week 5 - Session 1

Semi-supervised Learning Exercise

Consider the following figure which contains labeled (L) (class 1: black circles, class 2: hollow circles) and unlabeled (U) (blue squares) data. In this question, you will use two semi-supervised methods: S^3VM and co-training, to utilize the unlabeled data for further improvement of a SVM classifier.



(a) Explain how would the semi-supervised SVM (S^3VM) perform on this data as compared to the supervised SVM, by plotting the separating hyper-planes produced by both algorithms. For SVM, please draw marginal boundaries and separating hyper-plane by solid lines. Using a different color, draw the marginal boundaries and the separating hyper-plane of S^3VM using dash lines.



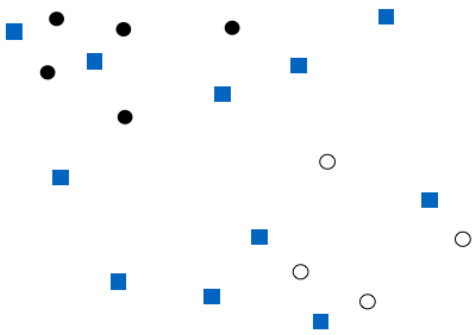
(b) In applying co-training, g_1 and g_2 are SVM classifiers, and p_i/n_i represents the number of positive or negative points to label at iteration i, respectively. In this example, assume that $p_1, n_1 = 2$.

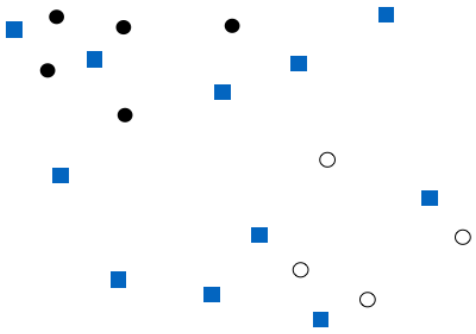
(b.1) Explain What is the main underlying assumption in applying co-training? How can we apply it to the above data (what are the two classifiers)?

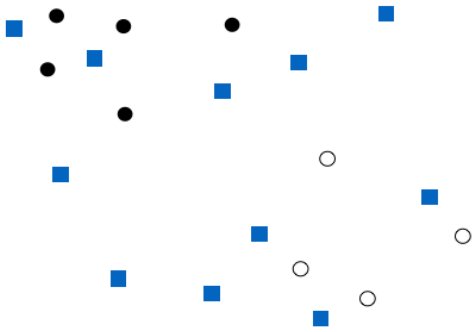
Solution:

(b.2) Identify the label of each point xi $\in$ U and the final SVM separating hyper-plane after applying 1 iteration of semi-supervised co-training. Assume that at 1st iteration , p1 and n1 points are labeled based on the farthest distance from the separation hyper-plane.

Solution:







(c) Compare the separating hyper-planes produced by S^3VM from part (a) and co-training from part (b.2). Explain why you think these two algorithms perform similarly/differently?

Solution:

Solution: