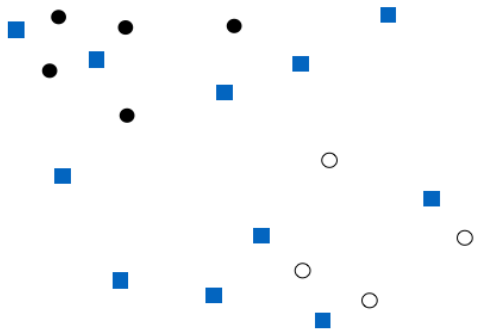
Semi-supervised Learning Project

Week 5 - Session 2

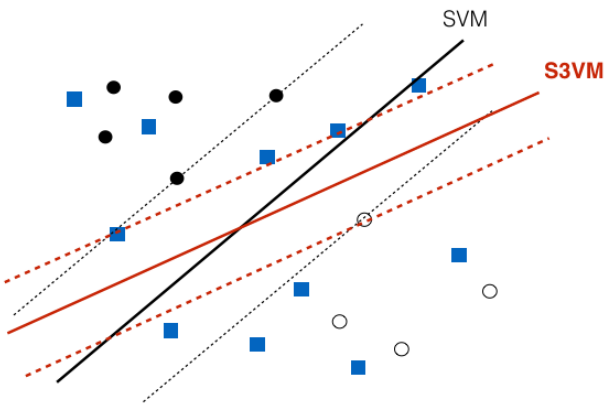
Semi-supervised Learning Exercise

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.

Solution: For supervised SVM, we only consider the labeled data and try to maximize the margins. The separating hyper-plane would split the margin into half (black line). For S^3VM, we start with SVM separator and greedily move it to low density area while maximizing margins for both labeled and unlabeled data points. The separating hyper-plane splits the width of the margin into half (red line).



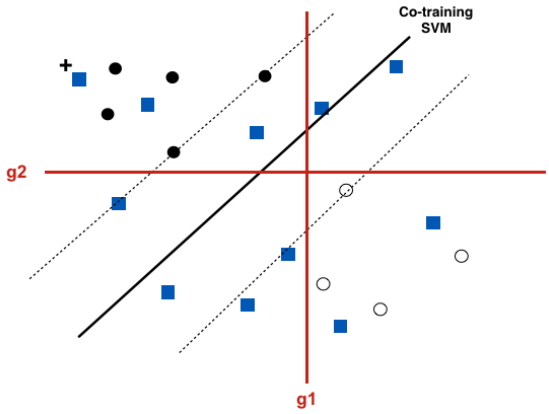
(b) In applying co-training, g_1 and g_2 are SVM classifiers, and p_i/n_i represents the number of positive or negative points to label at iteration i, respectively. In this example, assume that $p_1, n_1 = 2$ and $p_2, n_2 = 1$.

(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: Co-training partitions the feature space into two separate sets and its underlying assumption is for the two feature sets to be independent given the class. It also assumes that each set of feature are sufficient for classification. Here, the most natural way is to use one classifier (an SVM) for the x-axis and the second (another SVM) using the y-axis.

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

Solution: In literature, there are multiple versions of co-training algorithm for semi-supervised learning. Here we stick to the pseudo-code illustrated in lecture notes. In this algorithm, only the intersection of p and n points are added to the set of labeled points L. In this example only one sample is labeled after 2 iterations. (other versions of co-training are also acceptable solutions.)



(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: They perform differently because of multiple reasons: 1) in S^3VM all of the unlabeled data are considered for training while in co-traning algorithm, only a limited number of unlabeled data are contributing at a time. 2) the objective of S^3VM is to maximize the margin, while co-training in this example is based on the distance from separating hyper-plane. 3) The underlying assumption of independency of x and y given the class label might not hold in this example, thus resulting in less accurate classification.