CS 4780/5780 Homework 5

Due: Thursday 10/25/18 11:55pm on Gradescope

Problem 1: Derivation for Hard-margin Linear SVMs

a) Assume your data is linearly separable. What is the strength for the hard-margin linear SVM solution over the solution found by the Perceptron?

b) In class we mentioned there are two equivalent formulation of linear SVM, which is shown below.
Formulation A:

$$\min_{\mathbf{w},b} \mathbf{w}^\top \mathbf{w}$$
$$\text{s.t.} \begin{array}{ll} \forall_i \; y_i(\mathbf{w}^T\mathbf{x}_i + b) & \geqslant 0 \\ \min_i \left|\mathbf{w}^T\mathbf{x}_i + b\right| & = 1 \end{array}$$

Formulation B:

$$\min_{\mathbf{w},b} \; \mathbf{w}^\top \mathbf{w}$$
$$\text{s.t.} \forall i, y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geqslant 1$$

i. Assume you have found the optimal solution to the above optimization. Prove the optimal solution of formulation A is a feasible solution for formulation B.
(Note: a feasible solution satisfies all constraints yet may not be optimal.)
ii. Assume you have found the optimal solution to formulation B. Prove that this solution is a feasible solution for formulation A.
iii. Prove that for the optimal solution in formulation A is the optimal solution for formulation B and vice versa.

Problem 2: Hard- vs. Soft-margin SVMs

a) The Perceptron algorithm does not converge on non-linearly separable data. Does the hard-margin SVM converge on non-linearly separable data? How about the soft-margin SVM? Please explain in details.

b) For the soft-margin linear SVM, we use the hyperparameter C to tune how much we penalize mis-classifications. As $C \to \infty$, does the soft-margin SVM become more similar or less similar to the hard-margin SVM? As $C \to 0^+$, what happens to the solution of the soft-margin SVM? Why?

c) Suppose you found a solution $(\hat{\mathbf{w}}, \hat{b})$ to the hard-margin SVM. The separating hyperplane is surrounded by a margin defined by hyperplanes $\{\mathbf{x} : \hat{\mathbf{w}}^\top \mathbf{x} + \hat{b} = 1\}$ and $\{\mathbf{x} : \hat{\mathbf{w}}^\top \mathbf{x} + \hat{b} = -1\}$. Prove that at least one training data point lies on each of these margin hyperplanes.