CS 4780/5780 Homework 5

Problem 1: Derivation for Hard-margin Linear SVMs

a)
The hard-margin linear SVM solution gives the "best" separating hyperplane, i.e. the one that maximizes the margin to the nearest data points, whereas the perceptron gives any seperating hyperplane.
b) This is to prove the constraints are equivalent in the two formulations.
i. As in formulation A, y_i and $(w^T x_i + b)$ should have the same sign because their product is greater than 0. Since $y_i \in \{-1, +1\}$, so $y_i(w^T x_i + b)$ should have been larger or equal to 1. Therefore, an optimal solution of A is a feasible solution in B.
ii. To show that the solution is feasible with respect to A we must show that an optimal solution of B satisfies both constraints of A. To start, if $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geqslant 1$ then $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geqslant 0$ is satisfied. For the second constraint we know that since $y_i \in \{-1, +1\}$, then for all $\mathbf{x}_i$, $\left|\mathbf{w}^T\mathbf{x}_i + b\right| \geqslant 1$. Now we must show that there exists an x that gives us exact equality($\left|\mathbf{w}^T\mathbf{x}_i + b\right| = 1$). For contradiction let us suppose no such x exists. Then there is some x^* such that $\left|\mathbf{w}^T\mathbf{x}*_i + b\right|$ is minimized and $\left|\mathbf{w}^T\mathbf{x}*_i + b\right| = c$ for some $c > 1$. Note that c must be greater than 1 because of the statement we showed earlier that for all $\mathbf{x}_i$, $\left|\mathbf{w}^T\mathbf{x}_i + b\right| \geqslant 1$. But now we can divide $\mathbf{w}$ and b by c so that $\left|\mathbf{w}^T\mathbf{x}*_i + b\right| = 1$ and our objective function will decrease by a factor of $1/c^2$ which means that our solution was not optimal so we've gotten a contradiction. Therefore if the solution to B is optimal then some x must exist such that $\left|\mathbf{w}^T\mathbf{x}*_i + b\right| = 1$. Therefore, an optimal solution of B is a feasible solution for A.
iii. Both formulations have the same objective function and because we have justified above that any optimal solution for one is a feasible solution for the other. So their solutions must have the same optimal value. Otherwise, if one of them has a solution with low objective value, you use that point for the other problem and would obtain the same low object value. Therefore we show that an optimal solution for A is an optimal solution for B, and vice versa.

Problem 2: Hard- vs. Soft-margin SVMs

a)
The hard-margin SVM does not converge on non-linearly separable data but the soft-margin SVM does because the soft constraints introduce a slack variable which allows the initial constraints presented in the hard-margin version to be violated such that the optimization problem becomes solvable.
b)

$$\min_{x \in D} |\hat{\mathbf{w}}^\top \mathbf{x} + \hat{b}| = 1 \implies \exists x \text{ s.t. } \hat{\mathbf{w}}^\top \mathbf{x} + \hat{b} = 1 \text{ OR } \hat{\mathbf{w}}^\top \mathbf{x} + \hat{b} = -1$$

This only shows that at least on training datapoint lies on at least one of the two margin hyperplanes.

Now, let us proceed with the assumption that there is no point on one of the two margin hyperplanes. If this was the case, then the solution to the maximum margin optimization would yield different margin hyperplanes. This contradicts the given margin hyperplanes in the problem and so there must be at least one training datapoint on **each** hyperplane.