Problem 1 – VC Dimension

Let $F$ be the class of rectangle classifiers in $\mathbb{R}^2$. I.e., $F = \{ f(x) = \pm(\frac{1}{2} - 1_{[a,b] \times [c,d]}(x)), a < b, c < d \}$. 

Prove that this class has VC dimension at least 5.

[Optional, harder: Prove (or disprove) that the VC dimension of $F$ is equal to 5.]

[Optional, not graded. Problem 2 – Linear SVM ]

1. Prove that for the linear SVM in the linearly separable case, the solution satisfies 
   \[ ||w||^2 = \sum_i \alpha_i \]
   (Hint: a simple solution exists, but you need a fact from the notes.)
2. Show that the value of $b$ is theoretically the same no matter what particular support vector is used for its calculation.
3. Define $\beta_i = \alpha_i y^i \in \mathbb{R}$ for $i = 1 : N$. Reformulate the SVM dual problem in terms of variables $\beta_{1:N}$. What will the expression of $w$ be, as a function of $\beta_{1:N}$?

Problem 3 – Leave one out CV and support vectors

Assume the data set $D$ contains $N$ samples. You perform leave-one-out cross-validation, i.e., for $i = 1 : N$ you compute a linear support vector machine classifier $f_{-i}$ on $N - 1$ points, leaving out $(x^i, y^i)$.

a. Assume that the original data set is linearly separable. Prove that each of the $N$ support vector problems is also linearly separable.

b. Is it possible that $f_{-i}(x) \equiv f_{-j}(x)$ for $i \neq j$ two points in the training set $D$? Give a short motivation or proof.

c. Denote by $\hat{L}_{01}^{loo}$ the error rate in leave-one-out CV, i.e.
   \[ \hat{L}_{01}^{loo} = \frac{|\{i, f_{-i}(x^i) \neq y^i\}|}{N} \]
   Prove that $\hat{L}_{01}^{loo} \leq \frac{\#\text{support vectors of } f}{N}$, where $f$ is the linear support vector classifier trained on all the data.
Problem 4 – Regularization is monotonic on the penalty

Let \( J_\lambda(w) = \hat{L}_h(w) + \frac{\lambda}{2} ||w||^2 \) be the linear support vector formulation from Lecture 3 (for simplicity we take \( b = 0 \), i.e. \( f(x) = w^T x \)). Let \( \lambda_1 > \lambda_2 > 0 \) and denote \( w_{1,2} = \arg\min_w J_{\lambda_{1,2}} \) the optimal solutions for \( \lambda_1 \), respectively \( \lambda_2 \), and assume further that \( J_{\lambda_{1,2}} \) have unique global minima.

a. Prove that \( ||w_1|| < ||w_2|| \) whenever \( w_{1,2} \neq 0 \).

b. Prove also that \( \hat{L}_h(w_1) > \hat{L}_h(w_2) \).

In other words, imposing more regularization reduces the regularized quantity, and increases the un-regularized one.