Problem Set (CS-729)

Note: Provide legible and short proofs/derivations/answers.

1. Consider the finite function class learning bound derived in the lectures, where we assumed that the loss is bounded. Now additionally suppose we assume that the variance of loss evaluated at a random data point is known. Since we are given this additional information about the unknown distribution, we expect to get tighter bounds. Using lemma 1 in [1], derive a tigher learning bound for this case.

2. Consider the class of linear binary classifiers i.e., $\left\{f \mid \exists w \ni f(x) = sign(w^\top x) \ \forall\ x \in \mathcal{X}\right\}$. One way to describe an "ideal" w is by saying $P[Yw^\top X > \Delta] \geq 1-\delta$, where $\Delta > 0, \delta \in (0,1)$ are given[1]. Assuming that the true means and support for $X/Y = -1$ and $X/Y = 1$ are known, derive a system of regular inequalities[2], involving only w, Δ, δ and the means and the support information, which when satisfied will imply that the above probability condition is satisfied. In other words, if one writes an algorithm to solve this system of inequalities, then one will find an "ideal" w.

3. Exercise 2.9 in [2].

4. In the lecture we bounded data in a sphere and then showed that Rademacher complexity of $\mathcal{F}_W \equiv \{f | \exists w \in \mathbb{R}^n, \|w\| \leq W \ni f(x) = w^\top x \ \forall x \in \mathcal{X}\}$ is upper bounded. Now since any bounded data can even more tightely packed by an ellipsoid, lets assume that you are given a $\Sigma \succ 0$ such that $x^\top \Sigma^{-1} x \leq 1 \ \forall\ x \in \mathcal{X}$. Now, modify the function class $\mathcal{F}$ minimally such that the Rademacher complexity

[1] We expect to have small δ and large Δ

[2] Not involving any probabilities etc.

with the ellipsoid assumption can be upper bounded by $\frac{W}{\sqrt{m}}$. *Hint*: You simply need to change the Euclidean norm to something else!

5. Exercise 3.4, 3.5, 3.22 in [2].

References

[1] A. BenTal, S. Bhadra, C. Bhattacharyya, and J. SakethaNath. Chance constrained uncertain classification via robust optimization. *Mathematical Programming, Series B*, 127:145–173, 2011.

[2] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of Machine Learning*. MIT Press, 1 edition, 2012.