CPSC 531H Machine Learning Theory (Term 2, 2013-14)
Assignment 1

Question 1: [Expressivity of Affine Threshold Functions]

Assume that $X = \{-1, 1\}^n$. A decision list is a function whose input is a point $x \in X$, and whose output is of the form "if ℓ_1 then b_1, else if ℓ_2 then b_2, else if ℓ_3 then b_3, ..., else b_m", where each ℓ_i is a literal (either x_i or $-x_i$) and each $b_i \in \{-1, 1\}$.

(a): Show that conjunctions (like $x_1 \wedge \bar{x_2} \wedge x_3$) and disjunctions (like $x_1 \vee \bar{x_2} \vee x_3$) are special cases of decision lists.

(b): Show that decision lists are a special case of affine threshold functions. That is, any function that can be expressed as a decision list can also be expressed as an affine threshold function "$f(x) = +1$ iff $w_1 x_1 + \cdots + x_n x_n \geq w_0$", for some values $w_0, w_1, \ldots, w_n$.

Question 2: [VC-dimension of boxes]

What is the VC-dimension of axis-parallel boxes in $\mathbb{R}^3$? Such a classifier is specified by three intervals $[x_{min}, x_{max}]$, $[y_{min}, y_{max}]$, and $[z_{min}, z_{max}]$. It classifies a point (x, y, z) as positive if $x \in [x_{min}, x_{max}]$, $y \in [y_{min}, y_{max}]$, $z \in [z_{min}, z_{max}]$.

Question 3: [Growth-function bound for the inconsistent case]

Consider the empirical risk minimization algorithm. Assume that

$$m \;\geq\; \frac{K}{\epsilon^2}\Big(\log\big(\Pi_H(m)\big) + \log(1/\delta) \Big)$$

for some constant K. Prove, with probability at least $1 - \delta$, the estimation error is low, i.e.,

$$R(h_S^{ERM}) \leq \min_{h \in H} R(h) + \epsilon.$$

The proof strategy used in Lecture 3 can be modified to solve this problem.

Hint: Define the events

$$A = \{\exists h \in H \text{ with } |R(h) - \hat{R}_S(h)| > \epsilon\}$$

and

$$B = \{\exists h \in H \text{ with } |\hat{R}_S(h) - \hat{R}_{S'}(h)| \geq \epsilon/2\}$$

Whereas we previously used the Chernoff bound only once, it must now be used twice.