Homework 9

Due Wednesday, November 30

Problem 1. Consider the statistical query model. When learning a concept class defined over a continous domain, a slight extension of this model is necessary. A concept class $\mathcal{C}$ is *learnable in the statistical query model with unlabelled examples* if there exists an algorithm for PAC-learning $\mathcal{C}$ which, given inputs ϵ, δ and access to a *statistics oracle* $STAT(c, D)$ and an *unlabelled example oracle* $UEX(D)$, produces a hypothesis $h \in \mathcal{H}$ which has error$(h) \leq \epsilon$ with probability at least $1-\delta$. The unlabelled example oracle $UEX(D)$ simply draws instances $x \in X$ according to D, and returns x.

Let $\mathcal{C}_I = \{[a, b] : 0 \leq a \leq b \leq 1\}$ be the class of subintervals of the unit interval $[0, 1]$.

(a) Argue that $\mathcal{C}_I$ is probably *not* learnable in the statistical query model.
Hint: Consider a hard distribution over the instance space $[0, 1]$.

(b) Argue that $\mathcal{C}_I$ is efficiently learnable in the statistical query model with unlabelled examples.

(c) Argue that efficient learnability in the statistical query model with unlabelled examples implies both efficient PAC-learnability and efficient PAC-learnability in the presence of classification noise.