1. **The sample complexity of confidence.** In lecture we saw that in order to learn a class of VC-dimension $d$ with confidence $1 - \delta$ and error bounded by $\epsilon$, it is sufficient to find a representation $h$ that is consistent with $O\left(\frac{1}{\epsilon} (d \log \frac{1}{\epsilon} + \log \frac{1}{\delta})\right)$ examples. We also saw that $\Omega(d/\epsilon)$ examples were necessary to guarantee an error rate of $\epsilon$ with confidence $1 - \delta_0$ for some constant $\delta_0 > 0$. In this problem we will extend this lower bound to include a dependence on $\delta$ when we wish to learn to arbitrarily high confidence $1 - \delta$.

   (a) Show that for any class of VC-dimension at least 2, $\Omega\left(\frac{1}{\epsilon^3} \log \frac{1}{\delta}\right)$ examples are necessary for an algorithm to identify a representation that has error at most $\epsilon$ with probability at least $1 - \delta$.

   (b) Now, using the lower bound we proved in lecture, conclude that for any algorithm to find a representation with error at most $\epsilon$ with probability $1 - \delta$ when the data is labeled by a class of VC-dimension $d$ requires $\Omega\left(\frac{1}{\epsilon^3} (d + \log \frac{1}{\delta})\right)$ examples (as $d \to \infty, \epsilon \to 0, \delta \to 0$).

   **Remark.** Hanneke recently improved the analysis of the upper bound, to obtain that $O\left(\frac{1}{\epsilon^2} (d + \log \frac{1}{\delta})\right)$ examples suffice to find a representation in a class of VC-dimension $d$ that is correct with probability $1 - \epsilon$ with confidence $1 - \delta$. Thus, the sample complexity of learning a class of VC-dimension $d$ is asymptotically $\Theta\left(\frac{1}{\epsilon^2} (d + \log \frac{1}{\delta})\right)$ examples.

2. **VC-dimension bounds for nonzero error.** We can derive analogous sample complexity bounds in cases where no representation $h$ from the class $\mathcal{H}$ fits the data perfectly, perhaps due to noise in the data. Suppose a class $\mathcal{H}$ of representations of VC-dimension $d$ is given. Show that if we take a sample of

   $$m = O\left(\frac{1}{\epsilon^2} (d \log \frac{1}{\epsilon} + \log \frac{1}{\delta})\right)$$

   examples from a distribution $D$, then with probability $1 - \delta$ ($\delta < 1/2$) every representation $h \in \mathcal{H}$ has an empirical error $\sum_{j=1}^{m} \frac{1}{m} I[h(x(j)) \neq b(j)]$ within an additive $\epsilon$ ($\epsilon < 1/2$) of its true error $Pr_D[h(x) \neq b]$. (Here $I[\varphi]$ denotes the 0/1 valued indicator function that is 1 iff $\varphi$ is true.)

   **(Hint.** You can reuse the lemmas stated in lecture! Try to follow the analogous argument except that instead of choosing a set of $m$ out of $2m$ elements uniformly at random, try considering a pair of $m$-element samples in which you independently swap the $i$th element in the two samples with probability $1/2$.)
