Lemma 1 (Sauer’s Lemma). For any $\mathcal{H}$ with finite $d = \text{VC-dim}(\mathcal{H})$ and $m$ examples,
\[
\Pi_{\mathcal{H}}(m) \leq \sum_{i=0}^{d} \binom{m}{i} = \Phi_d(m).
\]

Proposition 2. Let $0 < d/m < 1$. Then,
\[
\Phi_d(m) \leq \left(\frac{e m}{d}\right)^d.
\]

Exercise 1 (10 points). Prove the following claim using Chebyshev’s inequality:
Given a coin that succeeds with probability $p \geq \epsilon > 0$, it holds that after $m \geq 8/\epsilon$ trials the number of successes is not less than $\epsilon \cdot m/2$ with probability at least $1/2$.
(The above claim was proven in the handout “Tools for Bounding Probabilities” (Lemma 2) using a Chernoff bound. Here we want to show that we can prove the same bound using Chebyshev’s inequality.)

Exercise 2 (10 points). In class we proved Theorem 3 below.

Theorem 3. For any $h \in \mathcal{H}$, if $h$ is consistent with all $m$ examples that are sampled i.i.d. from $\mathcal{D}$, then with probability $1 - \delta$
\[
\text{error}_{\mathcal{D}}(h, c) \leq \frac{2}{m} \cdot \left(\lg \left(\Pi_{\mathcal{H}}(2m)\right) + \lg \left(\frac{2}{\delta}\right)\right),
\]
where $\text{error}_{\mathcal{D}}(h, c)$ is the generalization error of the hypothesis $h$.

We did so by bounding the probability of the event
\[
B \equiv \left[\exists h \in \mathcal{H} : (h \text{ is consistent on } S) \wedge (\text{error}_{\mathcal{D}}(h, c) > \epsilon)\right],
\]
where $S$ was a sample of size $m$. In particular we showed that
\[
\Pr(B) \leq 2 \cdot \Pi_{\mathcal{H}}(2m) \cdot 2^{-\epsilon m/2}, \tag{1}
\]
where in the end we required to bound this quantity above by $\delta$ and from where the result in Theorem 3 followed. In this exercise we want to derive a bound on the number of samples $m$ that are sufficient for training based on the VC dimension of the hypothesis class. Use Sauer’s lemma (Lemma 1) together with Proposition 2 to replace the growth function in (1) and thus give a new upper bound on the probability of the bad event $B$. As usual, require to bound the resulting quantity from above by $\delta$ and solve for $m$. What is the bound that you get?

Hint: You may first want to prove Lemma 4 (worth 5 points) so that you can get the bound on $m$.

Lemma 4. For any $x > 0$, $c > 0$ it holds $\ln x \leq \left(\ln \left(\frac{1}{c}\right) - 1\right) + c \cdot x$. 

March 29, 2019
Exercise 3 (10 points). Let FIND-S be used in order to learn a target monotone conjunction properly (i.e., it holds $H = C$). The target monotone conjunction depends on at least one variable.

Assume that at least one redundant (wrong) variable appears in the hypothesis obtained at the end of the learning process using FIND-S.

During test time the learner is normally presented with truth assignments that are drawn from the uniform distribution $U_n$; i.e., each truth assignment $x \in \{0, 1\}^n$ can be obtained with probability $Pr_{x \leftarrow U_n}(x) = 2^{-n}$.

Now assume that an adversary who has complete knowledge of the target concept $c$ and the learned hypothesis $h$ exists (i.e., the adversary has complete knowledge of which variables appear in both $c$ and $h$). During test time the adversary intercepts every truth assignment $x$ that is drawn from the uniform distribution and then makes the minimum number of changes on the instance $x$ that was drawn so that the resulting instance $x'$ is misclassified by the hypothesis $h$. Note that such a truth assignment $x'$ always exists as $h$ contains at least one redundant (wrong) variable and the bit corresponding to that extra variable in $x'$ can always be set equal to 0 while at the same time we can set all the other bits that correspond to variables that appear in $c$ equal to 1, thus resulting in $h(x')$ being negative and $c(x')$ being positive.

Show that on a truth assignment $x$ drawn from the uniform distribution $U_n$, the adversary (on average) is expected to change $\Theta(k)$ bits on $x$ so that the resulting truth assignment $x'$ is misclassified by $h$, where $k$ is the number of variables that appear in the target concept $c$.

Hint 1: Let $E = \{x \in \{0, 1\}^n \mid c(x) \neq h(x)\}$ be the symmetric difference between $h$ and $c$ and moreover let $d(x, E) = \min\{r \mid x' \text{ differs from } x \text{ in at most } r \text{ bits such that } h(x') \neq c(x')\}$. We want to show that,

\[
\sum_{x \in \{0, 1\}^n} (Pr_{x \leftarrow U_n}(x) \cdot d(x, E)) = 2^{-n} \sum_{x \in \{0, 1\}^n} d(x, E) = \Theta(k).
\]

Hint 2: For obtaining a lower bound $\Omega(k)$ it is enough if you consider only truth assignments that are falsifying both $c$ and $h$ simultaneously.