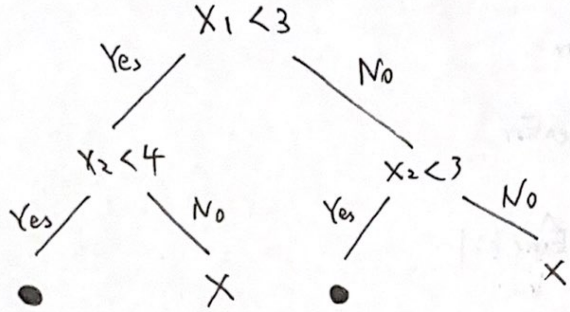
Note for homework 1 of CS 165A

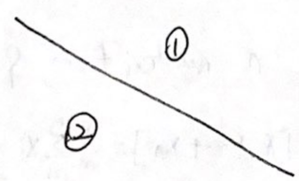
Problem 2.

(b) The answer is Yes



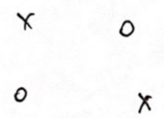
* Comparison between decision tree and linear classifier : None of them dominate each other.

Case 1.



Can be perfectly classified by linear classifier, while decision tree with arbitrary depth can not give perfect classification.

Case 2.



Can be perfectly classified by decision tree with depth 2, while linear classifier can not give perfect classification.

(c) This is a non-parametric classifier.

data ↑ complexity of the classifier ↑

(d) $|X| = 2^d$

So the answer is $2^{|X|} = 2^{2^d}$

(We say $h_1 \neq h_2$ if and only if there exists $x \in X$ such that $h_1(x) \neq h_2(x)$)

Problem 3.

(a) $Err(\hat{h}) \leq Err(h^*) + 2\,GenErr$

Proof: $Err(\hat{h}) \overset{①}{\leq} \hat{Err}(\hat{h}) + GenErr$

$\overset{②}{\leq} \hat{Err}(h^*) + GenErr$

$\overset{③}{\leq} Err(h^*) + 2\,GenErr$

①, ③ : $GenErr = \max_{h \in \mathcal{H}} | Err(h) - \hat{Err}(h) |$

$\geq \max \{ | Err(\hat{h}) - \hat{Err}(\hat{h}) |, | Err(h^*) - \hat{Err}(h^*) | \}$

② : $\hat{Err}(\hat{h}) \leq \hat{Err}(h^*)$ as $\hat{h} = \arg\min_{h \in H} \hat{Err}(h)$

(b) It suffices to prove that if we randomly pick n numbers from $\{x_1, \cdots x_m\}$, and get $\{X_{(1)}, X_{(2)}, \cdots X_{(n)}\}$, then $E\frac{1}{n}[X_{(1)} + \cdots + X_{(n)}] = \frac{1}{m}\sum_{i=1}^{m} x_i$.

Pf. $E\frac{1}{n}[X_{(1)} + \cdots + X_{(n)}] = \frac{1}{n} E(\sum_{i=1}^{m} x_i \cdot \mathbb{1}\{x_i \text{ is picked}\})$

$= \frac{1}{n} \sum_{i=1}^{m} x_i P\{x_i \text{ is picked}\}$

$= \frac{1}{n} \sum_{i=1}^{m} x_i \frac{n}{m}$

$= \frac{1}{m} \sum_{i=1}^{m} x_i$