1 Learning Theory

1.1 Review

Def 1. ***PAC Bound*** *- Probably Approximately correct.* δ *parameter controls probability of correctness,* ϵ *parameter controls error rate.*

Different bounds for different use cases:

Prop 1. *When* $error_{train} = 0$ *Haussler's bound tells us*

$$error_{true}(h) \leq \epsilon = \frac{ln|H| + ln(1/\delta)}{m}$$

Prop 2. *In general Hoeffding's bound tells us*

$$error_{true} \leq \epsilon + error_{train} = error_{train} + \sqrt{\frac{ln(|H|) + ln(1/\delta)}{2m}}$$

What is the relationship between Haussler and Hoeffding? Which one is tighter when applicable?

Prop 3. *When the model class is infinite use VC Dimension:*

$$error_{true}(h) \leq error_{train}(h) + 8\sqrt{\frac{VC(H)(ln(m/VC(H)) + 1) + ln(8/\delta)}{2m}}$$

Is VC bound tighter or weaker than the other bounds? What's a drawback to attempting to use the VC bound?

1.2 Problems

Problem 1. *What is the VC Dimension of affine classifiers in* R^n*? Prove this*

n+1. Consider the n simplex.

Problem 2. *Consider classifier* $f(sin(\alpha x))$ *where* $f = 1$ *when* $x > 0$ *and 0 otherwise.* α *adjustble. What is the VC dimension?*

Infinite since can adjust α.

For each of the below questions, **pick the formula** you would use to estimate the number o examples you would need to learn the concept. You do not need to do any computation or plug i any numbers. Explain your answer.

1. (2 pts) Consider instances with two Boolean variables $\{X_1, X_2\}$, and responses Y are give by the XOR function. We try to learn the function $f : X \to Y$ using a *2-layer neural networ*

2. (2 pts) Consider instances with two Boolean variables $\{X_1, X_2\}$, and responses Y are give by the XOR function. We try to learn the function $f : X \to Y$ using a *depth-two decisio tree*. This tree has four leaves, all distance two from the top.

Problem 3.

i) VC since dimension infinite

ii) Haussler since classifiers finite

2 Kernelization

2.1 Problems

Problem 4. *Given kernel* $K(x,y) = (1+x\cdot y)^d$ *what is the dimensionality of the corresponding feature map?*

Polynomial in n, d.

Problem 5. *Prove a kernel* $K(x,y)$ *is symmetric.*

Write as feature map product.

Problem 6. *Prove the kernel matrix* A *where* $A_{ij} = K(x_i, x_j)$ *is positive semi-definite.*

$A = B^T B$ where B is vector of feature maps. Then