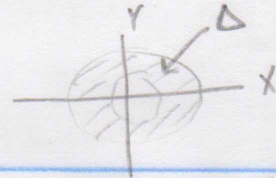
Anas Salamah
Learnability



2.3

$$\Pr[\text{error}] = \Pr\left[\bigcup_{i=1}^{2} x \notin R_i\right] \quad \text{Union Bound}$$

$$\leq \sum_{i=1}^{2} \Pr[x \notin R_i] \quad \text{Independence}$$

$$= \sum_{i=1}^{2} (1 - P(R_i))^m$$

Let us assume $P(R_i) \geq \frac{\epsilon}{2}$, then:

$$\Pr[\text{error}] \leq 2\left(1 - \frac{\epsilon}{2}\right)^m \leq 2e^{-\frac{m\epsilon}{2}} \quad (1 + x \leq e^{-x} \text{ inequality})$$

Solving for m:

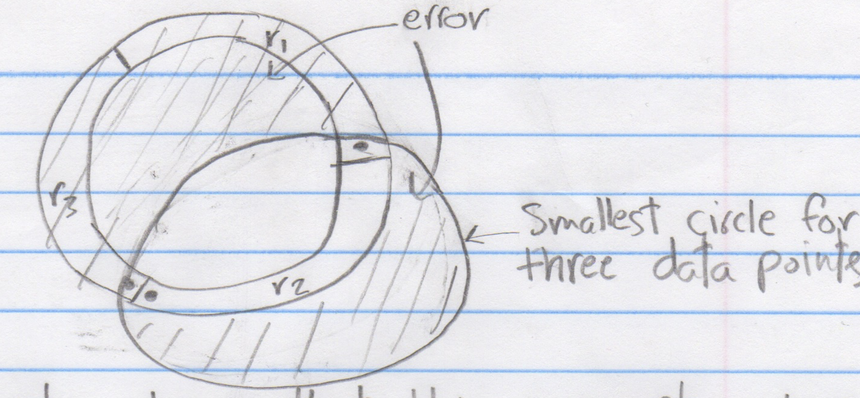
$$\Pr[R(h) \leq \epsilon] \leq 1 - \delta \quad (\text{Substituting } 2e^{\frac{-m\epsilon}{2}} \text{ for } \epsilon)$$

$$2e^{\frac{-m\epsilon}{2}} \leq \delta \quad (\text{take the log})$$

$$-\frac{m\epsilon}{2} \leq \log(\delta) \quad (\text{multiply by } -\tfrac{2}{\epsilon})$$

$$\boxed{m \geq \frac{2}{\epsilon} \log\left(\frac{2}{\delta}\right)}$$ So, Class is PAC-Learnable

$\boxed{2.4}$



This example shows that this concept class is not PAC-Learnable. The error here is clearly greater than ϵ and so, there might be data points that are missed outside of the three reigons

$\boxed{2.6}$

(a) The probability of R' missing r_j, $j \in [1,4]$ is the product of $P(x_i \notin r_j)$ or $P(x_i \in r_j \wedge \text{flipped positive})$

$$= P(x_i \notin r_j \vee (x_i \in r_j \wedge \text{flipped positive } x_i))$$
$$= P(x_i \notin r_j) + P(x_i \in r_j \wedge \text{flipped positive } x_i)$$
$$= (1 - P(x_i \in r_j)) + \eta \Pr(x_i \in r_j)$$
$$= (1-\eta)(1 - P(x_i \notin r_j)) + \eta \quad (\text{four reigons})$$
$$\leq (1-\eta)(1-\epsilon/4) + \eta$$
$$\boxed{= (1 - \tfrac{\epsilon}{4}) + \eta\tfrac{\epsilon}{4} \leq 1 - \tfrac{\epsilon}{4}(1-\eta')}$$

(b) Assume $P(R') \geq \epsilon/4$

$$\Pr[\text{error}] \leq 4(1 - \epsilon/4(1-\eta')) \leq \boxed{4e^{-m\frac{\epsilon}{4}(1-\eta')}}$$
$$\Pr[\text{error}] \leq 1 - \delta$$
$$4e^{-\frac{m\epsilon}{4}(1-\eta')} \leq \delta \quad (\text{Take the log})$$
$$-\frac{m\epsilon}{4}(1-\eta') \leq \log(\delta) \quad (\text{multiply by } \tfrac{-4}{\epsilon(1-\eta')})$$
$$\boxed{m \geq \frac{4}{\epsilon(1-\eta')} \log(\tfrac{4}{\delta})}$$

3.6 The VC-Dimension of a hypothesis set H is the size of the largest set that can by fully shattered by H:

$$VCdim(H) = max\{m : \Pi_H(m) = 2^m\}$$

Any 2K points on the real line can be shattered by K intervals. For 2K+1 points, assume that the points are alternatively labeled by +1 or -1. So, we will have K+1 positive points & need 2K+1 intervals to shatter since an interval will not have two consecutive points. As a result no 2K+1 set of points can be shattered. So, the VC dimension of subsets of the real line formed by the union of K intervals is 2K.

$\boxed{3.5}$ We are going to use the Growth Function. If G is a function taking values $\{-1,+1\}$, then

$$\mathbb{R}_m(G) \leq \sqrt{\frac{2 \ln \Pi_G(m)}{m}}$$

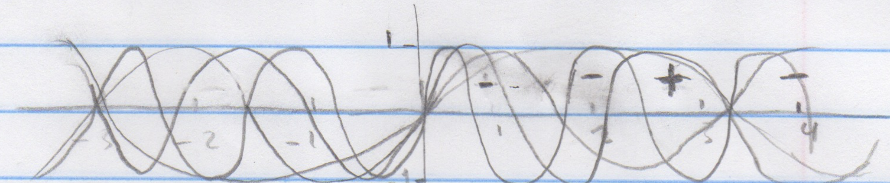
where $\Pi_G(m)$ is Growth Function. To connect that to VC dimensions and Radamacher complexity, we use Sauer's Lemma:

$$\Pi_H(m) \leq \sum_{i=0}^{d} \binom{m}{i} \equiv \phi_d(m)$$

Where $\phi_d(m) \leq \underset{m}{E}[\phi(m)] + \sqrt{\frac{\ln \frac{1}{\delta}}{2m}}$. The above contradicts Professor Jesetoo's bound on Radamacher Complexity

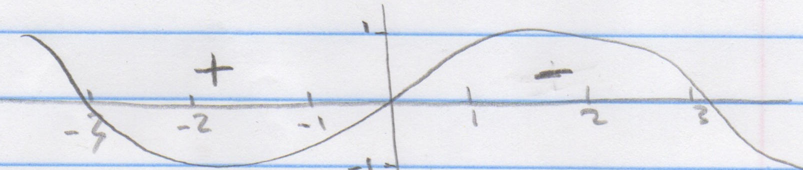
$\boxed{3.12}$

(a)



Let $X \in \mathbb{R}$ be fixed then, the four points labeled $- - + -$ cannot be shattered

(b) By making the frequency of the sine function greater, any two points of opposite signes can be seperated by a sine cycle



$\boxed{3.19}$ (a)

$$\text{error}(f_0) = Pr[f_0(S) \neq X] \quad \text{(Prediction Rule)}$$
$$= Pr[f_0(S) = X_A \wedge X = X_B] + Pr[f_0(S) = X_B \wedge X = X_A]$$
$$= Pr[N(S) < \tfrac{m}{2} \mid X = X_B] Pr[X = X_B] +$$
$$Pr[N(S) \geq \tfrac{m}{2} \mid X = X_A] Pr[X = X_A]$$
$$= \tfrac{1}{2} Pr[N(S) < \tfrac{m}{2} \mid X = X_B] + \tfrac{1}{2} Pr\, N(S) \geq \tfrac{m}{2} \mid X = X_A]$$
$$\geq \tfrac{1}{2} Pr[N(S) \geq \tfrac{m}{2} \mid X = X_A]$$

(b)

Using the Binomial distribution,

$$\text{error}(f_0) \geq \tfrac{1}{2} Pr\left[N \geq \frac{m\epsilon/2}{1/4(1-\epsilon^2)m}\right] = \tfrac{1}{2} Pr\left[N \geq \frac{\sqrt{m\epsilon}}{\sqrt{1-\epsilon^2}}\right]$$

Using Tail bound,

$$\text{error}(f_0) \geq \tfrac{1}{4}\left(1 - \sqrt{1 - e^{-u^2}}\right).$$ This agrees with 3.57

(c)

Since $Pr[N(S) \geq \frac{m}{2} \mid X = X_A] \geq Pr[N(S) \geq \frac{m+1}{2} \mid X = X_A]$

So, $\text{error}(f_0) \geq \frac{1}{2} Pr[N(S) \geq \frac{m+1}{2} \mid X = X_A]$

(d)

if error $\geq \delta$, then $\frac{1}{4}\left[1 - \left[1 - e^{-\frac{2(m/2)\epsilon^2}{1-\epsilon^2}}\right]^{\frac{1}{2}}\right] < \delta$

$$e^{\frac{-2(m/2)\epsilon^2}{1-\epsilon^2}} < 1 - (1-4\delta)^2 = 4\delta(2-4\delta) = 8\delta(1-2\delta)$$
$$m > 2\left[\frac{1-\epsilon^2}{2\epsilon^2} \log \frac{1}{8\delta(1-2\delta)}\right]$$

The lower Bound changes based on $\frac{1}{\epsilon^2}$

3.19

(e) $\text{error}(f) = \sum Pr[S \wedge X_B] + \sum Pr[S \wedge X_A]$

If $N(S) \geq m/2$, $Pr[S | X_A]$ and if $N(S) < \frac{m}{2}$, $Pr[S | X_A] \geq Pr[S | \bar{X}_B]$

$$\text{error}(f) \geq \sum Pr[S \wedge X_B] + \sum Pr[S \wedge X_A]$$

$$\text{error}(f) \geq \frac{1}{2} \sum Pr[S | X_B] + \frac{1}{2} \sum Pr[S | X_A] +$$

$$\frac{1}{2} \sum Pr[S | X_B] + \frac{1}{2} \sum Pr[S | X_A]$$

$$= \frac{1}{2} \sum Pr[S | X_B] + \frac{1}{2} \sum Pr[S | X_A]$$

$$= \text{error}(f_0)$$