Proof of:

THEOREM 12.2. *Given some $D \ge 1$, suppose that $\mathcal{F}$ is (N,D)-rich for all $N \ge 4D$. Then*

$$R_n(h,\mathcal{F}) \ge c(1-h)\frac{D}{nh}\left[1+\log\frac{nh^2}{D}\right] \tag{12.9}$$

for any $\sqrt{D/n} \le h < 1$, where $c > 0$ is some absolute constant.

($c = 1/72$ suffices)

Proof There exists $\bar{x}_1, \cdots, \bar{x}_N$ so that

for any $b = \{0,1\}^N_D$ there exists $f_b \in \mathcal{F}$

so $b = (f(\bar{x}_1), f(\bar{x}_2), \cdots, f(\bar{x}_N))$.

(There are $\binom{N}{b}$ such b's)

$\hat{f}_n : Z^n \to \{0,1\}$ $\quad Z^n = ((X_1, Y_1), (X_2, Y_2), \cdots (X_n, Y_n))$

P_b : dist^n on $X \times \{0,1\}$ such that if (X, Y) has dist' P_b then

X is uniformly distributed over $\bar{x}_1, \cdots, \bar{x}_N$

Given $X = \bar{x}_r$, $Y = \begin{cases} b_r & \text{w/p } \frac{1+h}{2} \\ 1-b_r & \text{"} \ \frac{1-h}{2} \end{cases}$

Note $P_b \in \mathcal{P}(h)$ for all b

so

$$L_b(f) = \frac{1-h}{2} + \frac{h}{N}\sum_{v=1}^{N} 1_{\{f(\bar{x}_v)\neq b_v\}}$$

general loss for classifier f under P_b

(f_b is optimal if P_b is the true measure)

$$L_b^* = L_b(f_b) = \frac{1-h}{2}$$

$$\sup_{P\in\mathcal{P}(h)} E^n[L(\hat{f}_n)] - L^* \geq \max_{P_b : b\in\{0,1\}^N_D} E_b^n[L_b(\hat{f}_n)] - L_b^* \geq$$

$$= \max_b h \frac{1}{N}\sum_{v=1}^{N} P_b^n\{\hat{f}_n(\bar{x}_v)\neq b_v\}$$

$$\geq h\varepsilon \text{ where } \varepsilon = \frac{1}{N}\sum_{v=1}^{N} \bar{P}^n\{\hat{f}_n(\bar{x}_v)\neq B_v\}$$

$\bar{P}^n$ - joint dist^n of B, $Z^n = ((X_1, Y_1), \ldots, (X_N, Y_N))$ such that B is uniform over $\{0,1\}^N_D$ and given $B=b$, the conditional dist^n of Z^n is P_b^n.

$$\begin{pmatrix} B = (B_1, \ldots B_N) \\ (\text{base labels for } \bar{x}_1, \ldots, \bar{x}_N) \end{pmatrix} \longrightarrow (X^n, Y^n) \longrightarrow \boxed{A} \longrightarrow \hat{f}$$

$$\hat{B} = (\hat{f}(\bar{x}_1), \ldots, \hat{f}(\bar{x}_N))$$

$$B \to \begin{matrix} X^n \\ Y^n \end{matrix} \to \boxed{A} \to \hat{f} \to \hat{B}$$

Lower bound on $I(B;\hat{B})$ (required to get bit error probability ε)

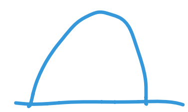
$$I(B;\hat{B}) = H(B) - H(B\mid\hat{B})$$
$$= \log\binom{N}{D} - H(\underbrace{B\oplus\hat{B}}_{\text{indicates error locations}}\mid\hat{B})$$
$$\geq \log\binom{N}{D} - H(B\oplus\hat{B})$$

$\frac{1}{N}\sum_v P((B\oplus\hat{B})_v = 1) = \varepsilon \leq \frac{1}{2}$

$\therefore H(\quad) \leq Nh(\varepsilon)$

$$\boxed{I(B;\hat{B}) \geq D\log\frac{N}{D} - Nh(\varepsilon)}$$

$$\binom{N}{D} = \frac{N}{D}\,\frac{N-1}{D-1}\cdots\frac{N-D+1}{1} \geq \left(\frac{N}{D}\right)^D$$



Upper bound on $I(B;\hat{B})$

$$I(B;\hat{B}) \leq I(B;(X^n,Y^n))$$
$$= \underline{I(B;X^n)} + I(B;Y^n\mid X^n)$$

where samples are taken is independent of B

$$= I(B;Y^n\mid X^n)$$
$$\leq I(B;Y^n\mid X^n) + I(X^n;Y^n)$$
$$= I((X^n,B);Y^n)$$

coordinates of Y^n are conditionally independent given X^n, B.

Suppose $(X^n,B) = (x^n,b)$

what's conditional dist^n of Y_i.

If $x_i = \bar{x}_v$ the $Y_i = b_v$ w/p $\frac{1+h}{2}$

$1-b_v$ " $\frac{1-h}{2}$

$$\underline{\underline{I(B;\hat{B})}} \le I((X^n,B);Y^n)$$

dist^{n} on Y^n

$$\le D(P_{Y^n|X^n,B} \,\|\, Q \,|\, P_{X^n,B})$$

$$\le \underline{\frac{nD}{N}\left[\frac{2h^2}{1-h}\right]}$$

Let Q be product dist^{n} such that under Q, $Y_1,\dots,Y_n$ are iid $Be\left(\frac{1-h}{2}\right)$

$$D(P_{Y^n|x^n,b} \,\|\, Q) \le \frac{2mh^2}{1-h}$$

$m =$ # of samples such that label is noisy version of 1

For $1\le i\le n$ - there are two cases for each i.

<u>Case 1</u> $b_v = 0$ where v is such that $x_i = \bar{x}_v$

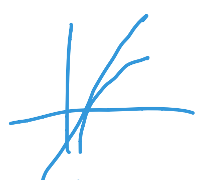
Then $Y_i \sim Be\left(\frac{1-h}{2}\right)$ $\qquad D\left(\text{dist. of } Y \,\|\, Be\left(\frac{1-h}{2}\right)\right) = 0$

<u>Case 2</u> $b_v = 1$ where v is s...

Then $Y_i \sim Be\left(\frac{1+h}{2}\right)$ $\qquad D\left(\text{dist of } Y \,\|\, Be\left(\frac{1-h}{2}\right)\right)$

$$= D\left(Be\left(\tfrac{1+h}{2}\right) \,\|\, Be\left(\tfrac{1-h}{2}\right)\right)$$

$$= \frac{1+h}{2}\ln\frac{\frac{1+h}{2}}{\frac{1-h}{2}} - \frac{1-h}{2}\ln\frac{\frac{1+h}{2}}{\frac{1-h}{2}}$$



$\ln t \le t-1$

$$= h\ln\frac{1+h}{1-h} \le h\left[\frac{1+h}{1-h} - 1\right]$$

$$= h\left(\frac{2h}{1-h}\right)$$

$$= \frac{2h^2}{1-h}$$

Combining upper + lower bounds on $I(B;\hat{B})$

$$D\log\frac{N}{D} - N h(\varepsilon) \le \frac{2h^2}{1-h}\left(\frac{nD}{N}\right)$$

$$\left(R_N \ge n h(\varepsilon) \right)$$

$$\boxed{h(\varepsilon) \ge \frac{D}{N}\log\frac{N}{D} - \frac{nD}{N^2}\left(\frac{\cancel{2}h^2}{1-h}\right)}$$

select N to make this ~~term~~ $1/2$ the first term

$$h(\varepsilon) \ge \frac{D}{2N}\ln\frac{N}{D} \qquad N = \left\lfloor \frac{9nh^2}{(1-h)\left(1+\log\frac{nh^2}{D}\right)} \right\rfloor$$

$$\therefore \quad \varepsilon \ge \frac{D}{8N}$$

$$R \ge h\varepsilon \ge \frac{hD}{8N} \ge \frac{(1-h)\left(1+\log\left(\frac{nh^2}{D}\right)\right)}{72\, n h^2}$$