$\mathcal{F}$
$f^* \in$
$\mathcal{F}$
$y =$
$f^*(x)$
$x \in$
$\mathcal{X}$
$y =$
$1-$
x

$\mathcal{F}$
$\mathcal{F} <$
∞
$y =$
$f^*(x)$
$f^* \in$
$\mathcal{F}$
$S =$
$\{(x_i, y_i)\}_{i=1}^n$
D

$$f_S^{ERM} :=_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \ell(f(x_i), y_i).$$

$n >$
$\log(\mathcal{F}/\delta)/\epsilon$
$>$
$1-$
δ
$\{(x_i, y_i)\}_{i=1}^n$
f_S^{ERM}
ϵ
$f^* \in$
$\mathcal{F}$
$R(f^*) =$
0

$$\mathcal{F}_{bad} := \{f \in \mathcal{F} : R(f)\epsilon\}.$$

$f \in$
$\mathcal{F}_{bad}$
$f' \in$
$\mathcal{F}_{bad}$
$_{S \in D^n}[f' is an ERM] =$
$P_{S \sim D^n}[f^* has zero empirical risk]$
$=$
$P_{S \sim D^n}[f'(x_i) =$
$f^*(x_i) for all i \in$
$[n]]$
$(i.i.d. assumption) =$
$\prod_{i=1}^n P_{(x_i, y_i) \sim D}[f'(x_i) =$
$f^*(x_i)](1-$
$\epsilon)^n$
$R(f')\epsilon$
$1-$
xe^{-x}

$$P_{S \in D^n}[f' is an ERM] e^{-n\epsilon}.$$

$f \in$
$\mathcal{F}_{bad}$

$$_{S \sim D^n}\left[\bigcup_{f \in \mathcal{F}_{bad}} \{f is an ERM\}\right] \sum_{f \in \mathcal{F}_{bad}} P(\{f is an ERM\}) \mathcal{F}_{bad} e^{-n\epsilon} \mathcal{F} e^{-n\epsilon}.$$

$$n > \epsilon^{-1}(\log \mathcal{F} + \log(1/\delta)) = \frac{\log(\mathcal{F}/\delta)}{\epsilon}$$

$$P\Big[\bigcup_{f \in \mathcal{F}_{bad}} \{f is an ERM\}\Big] < \delta,$$

$>$
$1-$
δ
$\mathcal{F}_{bad}$
$R(f_S^{ERM}) <$
ϵ
ϵ^2
ag-
o-
nis-
tic
learn-
ing
f
$\mathcal{F}$
d
W
$\mathcal{F} =$