Prevent overfitting

Sunday, August 25, 2024 9:47 AM

How to prevent overfitting

① Limit complexity of $\mathcal{H}$, eg. VC dimension($\mathcal{H}$) is finite
"classical"

Thm Binary classification under statistical learning setup,
then if $f_n \in \mathrm{ERM}_{\mathcal{H}}(S_n)$, then

$$\lim_{n\to\infty} R(f_n) - \hat{R}_n(f_n) = 0 \quad \text{if } \mathrm{VCdim}(\mathcal{H}) < \infty$$

$R(f_n)$: true risk; $\hat{R}_n(f_n)$: empirical risk

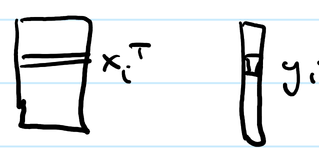
$\underset{f\in\mathcal{H}}{\operatorname{argmin}}\ \hat{R}_n(f)$

② Regularize ERM
"classical"

$$f_n \in \underset{f\in\mathcal{H}}{\operatorname{argmin}}\ \hat{R}_n(f) + \mathrm{Regularizer}(f)$$

ultra-classic
ex: "ridge regression"
"Tikhonov regularization"
(regularized) pseudoinverse

$f(x) = \langle w, x\rangle = w^T x = x^T w$ (w: parameters)



ERM is $\min_w \|X\cdot w - y\|^2$

Tikhonov is $\min_w \|X\cdot w - y\|^2 + \lambda\cdot\|w\|^2 \quad \lambda > 0$

adds bias, reduces variance

③ modern magic (poorly understood)

- dropout
- random initialization
- early stopping
- random training (SGD)

theories exist but not all satisfactory