STAT 391

2/27/25

Lecture 16

Ragression: Double descent Non-linear f Classification HWF TBRosted [Hw8 Not graded] Poll TB Posted





Double Descent

STAT 535+LPL2019

Beyond the Bias-Variance trade-off

Marina Meila University of Washington





Classical regime p < N</p>

- Modern/Deep Learning/High dimensional regime N > n
 - Think N fixed, p increases, gamma=p/N
 - Training error = 0 (interpolation)
 - Test error decreases with p (or gamma)



Double descent curves for the generalization error

- Random Fourier Features (RFF)
- ReLU 2 layer networks (with random first layer weights)
- Random Forests, 12-Adaboost
- Linear regression
- With and without noise



- Model y = <phi(x), beta >
- Large N (cover a compact data domain)
- Features random
- Min-norm solution beta*



+ RFF









Doubled bescardt for LP
True:
$$y = \sqrt{1}x + \varepsilon$$

 $\sqrt{1}x \in \mathbb{R}^{2}$
 $\sqrt{1}x \in \mathbb{R}^{2}$
 $\sqrt{1}x = \sqrt{1}(\sqrt{1})^{2} = 1:11y$
 $\sqrt{1} = \sqrt{1}(\sqrt{1})^{2}$
 $\sqrt{1} = \sqrt{1}(\sqrt{1})^{2} = 1:11y$
 $\sqrt{1} = \sqrt{1}(\sqrt{1})^{2}$
 $\sqrt{1} = \sqrt{1}($

