Crib 3

03 Bias Variance

by Alvin Wan . alvinwan.com/cs189/fa17

1 Bias Variance Decomposition

1. setup: i.i.d. $\{x_i\}_{i=1}^n, x_i \in \mathbb{R}^d$, possibly with $y = Xw + \epsilon$ for $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$
2. mean-squared error (MSE) decomposes in **bias**, **variance**, and **irreducible error**
3. $E[\|\hat{y} - y\|_2^2] = \underbrace{E[\|E[\hat{y}] - y\|_2^2]}_{\text{bias}} + \underbrace{E[\|\hat{y} - E[\hat{y}]\|_2^2]}_{\text{variance}} + \underbrace{\text{var}(\epsilon)}_{\text{irreducible error}}$
4. **bias**: MSE between estimator mean and true mean
5. **variance**: MSE between estimator and mean estimator
6. **irreducible error**: what it sounds like (shows up, if we model with noise)
7. Add regularization term: increases bias, decreases variance
8. Add new feature, more expressive model: decreases bias, increases variance
9. Add more data: variance decreases

2 Probability Review

Confused by definitions? See my 70 materials or my booklet for more probability review.

1. $x, \mu \in \mathbb{R}^d$, $E[(x-\mu)(x-\mu)^T] = cov(x)$, $E[(x-\mu)^T(x-\mu)] = var(x)$
2. $E[ax + b] = aE[x] + b$ for constants a, b
3. $Tr(E[x]) = E[Tr(x)]$ for any x, scalar or r.v.
4. $Tr(ABC) = Tr(BCA) = Tr(CAB)$
5. $Tr(A) = \sum_i \lambda_i(A)$ where $\lambda_i(A)$ is the ith eigenvalue of A
6. $det(A) = \Pi_i \lambda_i(A)$