Day 3 Notes: Parameter Estimation

Zach Neveu

September 19, 2019

- Example 1: linear and quadratic fits by hand.
- Remember: fit vector of parameters θ
- Estimating parameters useful for both ML and SP
- Quadratic fit can be expressed as: $y(x) = \sum_{k=-\infty}^{\infty} \theta_k \phi_k(x)$
- This is very similar to Fourier transform!
- Also similar to PCA
- Basically, θ is weight vector, ϕ is basis vectors

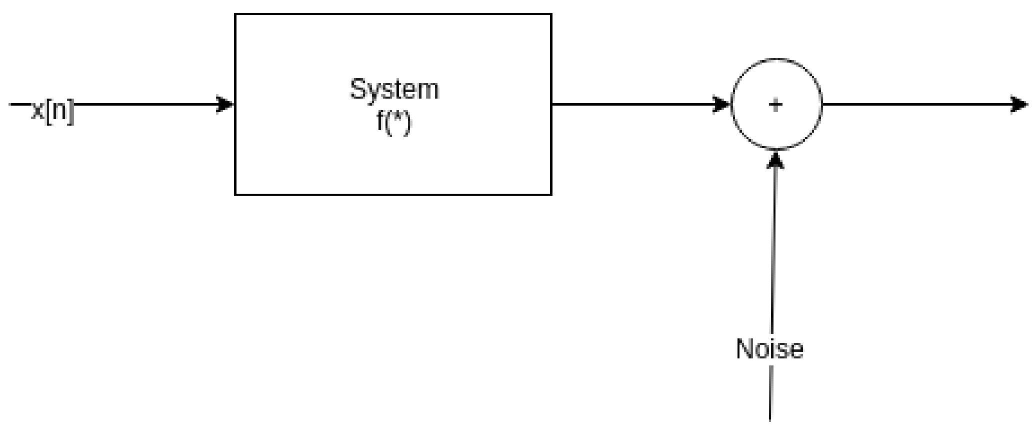

Figure 1: System with Noisy Output

- Consider noisy system in figure 1
- For this system $y = \theta^T x + \eta$
- Question: is this system LTI?
- Answer: θ must be unbiased to be LTI
- How to estimate θ?
- Loss functions! Minimize the loss to get the best fit.

1 Loss Functions

- Least Squares Loss
- Minimize squares of residuals

- Because of noise, measurement always changes. This means estimator for θ is also random, because $\hat{\theta}$ depends on y and y depends on noise η
- Bias-Variance Decomposition: expanding MSE definition shows that $MSE[x] = Var[x] + Bias[x]^2$
- Result: biased estimator can have better MSE than unbiased estimator
- norm of biased estimator always better than norm of unbiased estimator
- Ridge regression: squared loss from MSE, + squared norm of θ
- λ: regularizing term in ridge regression
- Given enough points, it is possible to find an optimal λ that is provably best
- Fun stuff: using Bayes theorem, plugging in normal dist for likelihood gets LS estimator, plugging in normal dist for prior leads to ridge regression.
- Review: ML vs MAP estimation

2 Definitions

- Bias: $Bias[\hat{\theta}] = E[\hat{\theta}] - \theta$
- Variance: $Var[\hat{\theta} = E[(\hat{\theta} - E[\hat{\theta}])^2]$
- MSE: $MSE[\hat{\theta}] = E[(\hat{\theta} - \theta)^2]$