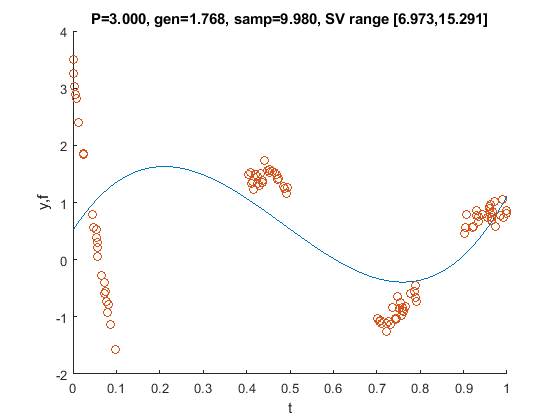
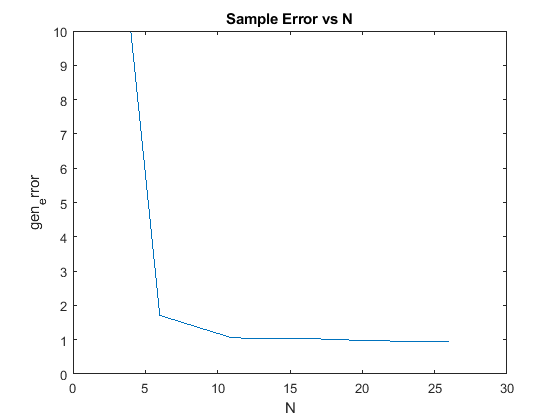
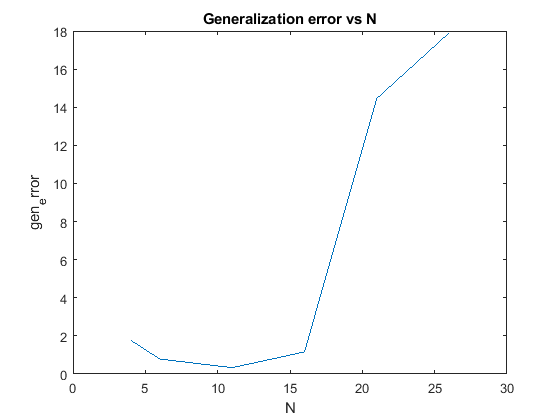
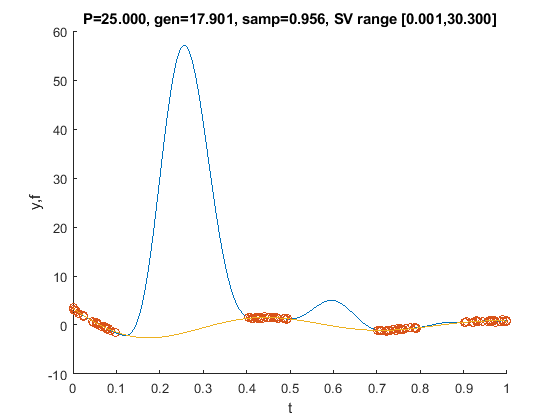
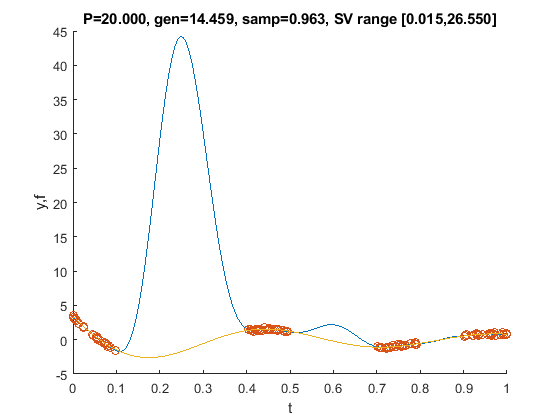
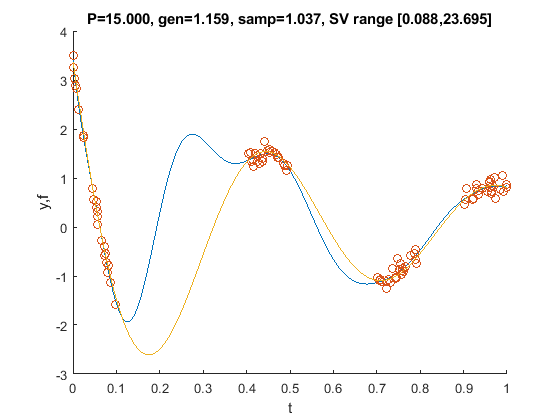
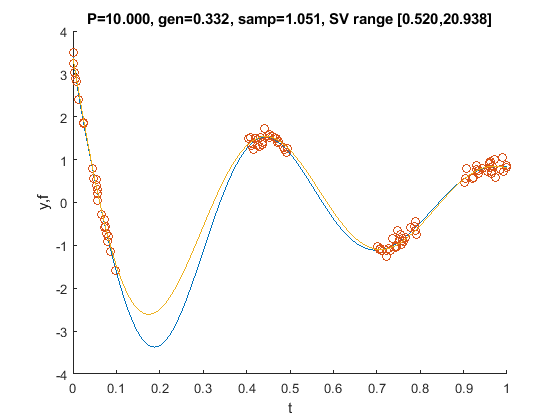
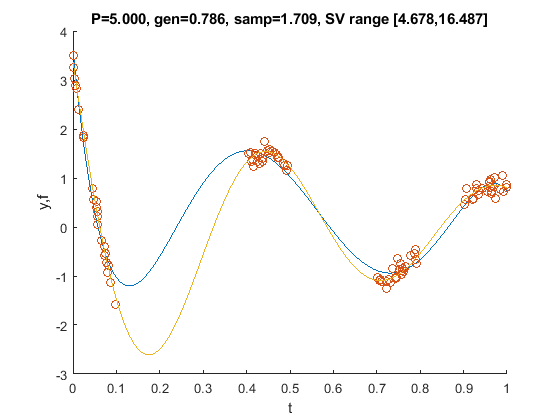
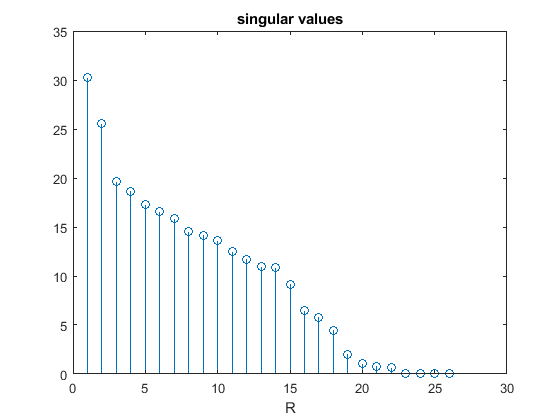
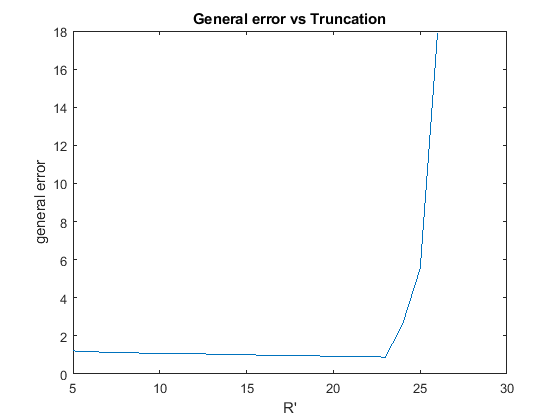
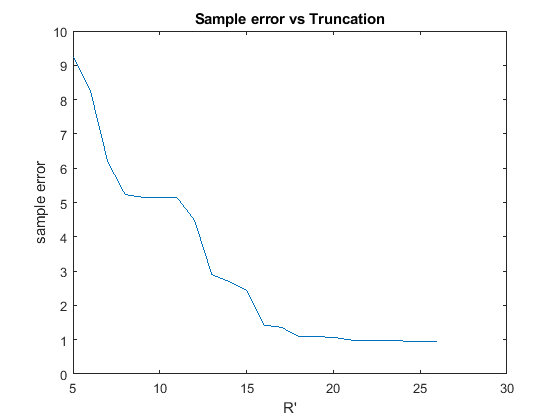
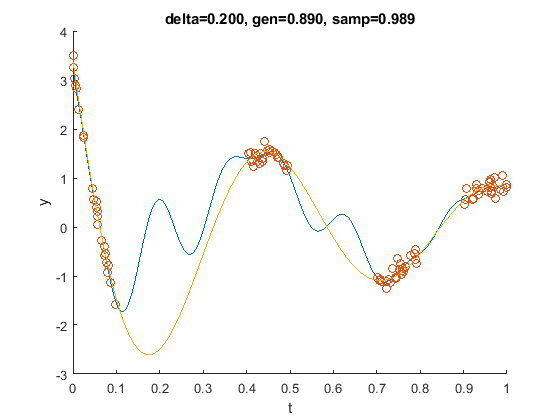
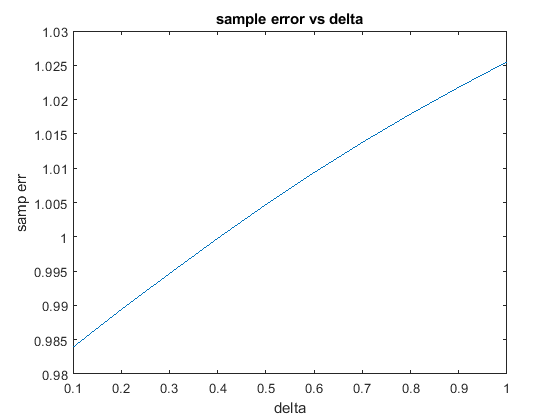
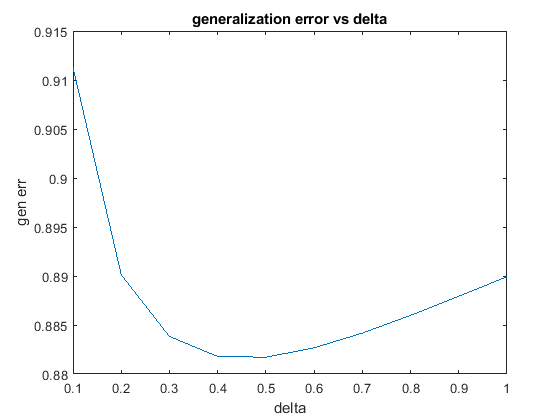
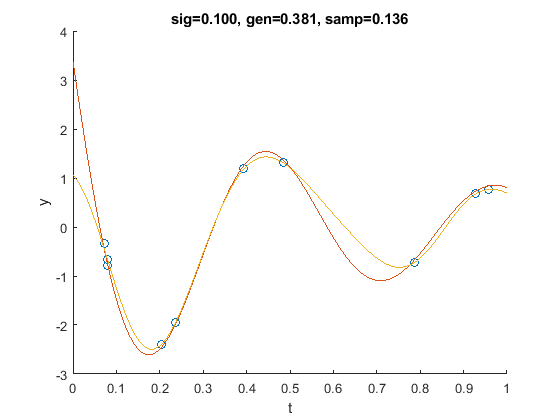
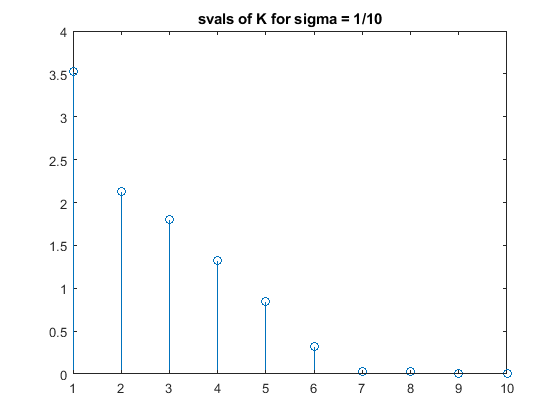
1. .
   1. 
   2. Sample and gen error in title
   3. Least squares begins to fall apart somewhere between N=10 and N=15 as that is when the generalization error begins increasing. The sample error goes down monotonically because least squares is still minimizing the error between the approximation function and the sample values, but at some point the coefficients get too large and the approximation strays far from the real function in an attempt to perfectly match erroneous sample values, so the generalization error becomes very large since the functions begin to mismatch.
   4. Pick R’ = 22, Generalization Error = 0.9071, Sample Error = 0.9852
   5. comment
   6. Smallest reasonable singular value is .68, first unacceptable singular value is .09. 0.68^2 = 0.46, reasonable delta would be 0.2 to leave that one undampened and the other dampened. Delta of 0.2 gives this plot (errors labeled):as delta is swept (from 0.1 to 1 in this case), the generalization error reaches a minimum and then climbs back up while the sample error slowly increases, as shown in the plots below.
2. .
   1. This delta was a good choice since it dampens the smallest singular values of the K matrix: Singular value #7 is 0.0310, so a delta value of 0.04 dampens this and the values below it well.
   2. The data point density clearly affects our choice of sigma. In this case, 1/5 has the best generalization error. This is because as sigma decreases, the kernel functions become more localized and the gaps are not filled in. The more points you have, the lower sigma should be (assuming the points are evenly spread out). If we had hundreds of samples, maybe sigma of 1/50 or 1/100 would not be so wrong. 