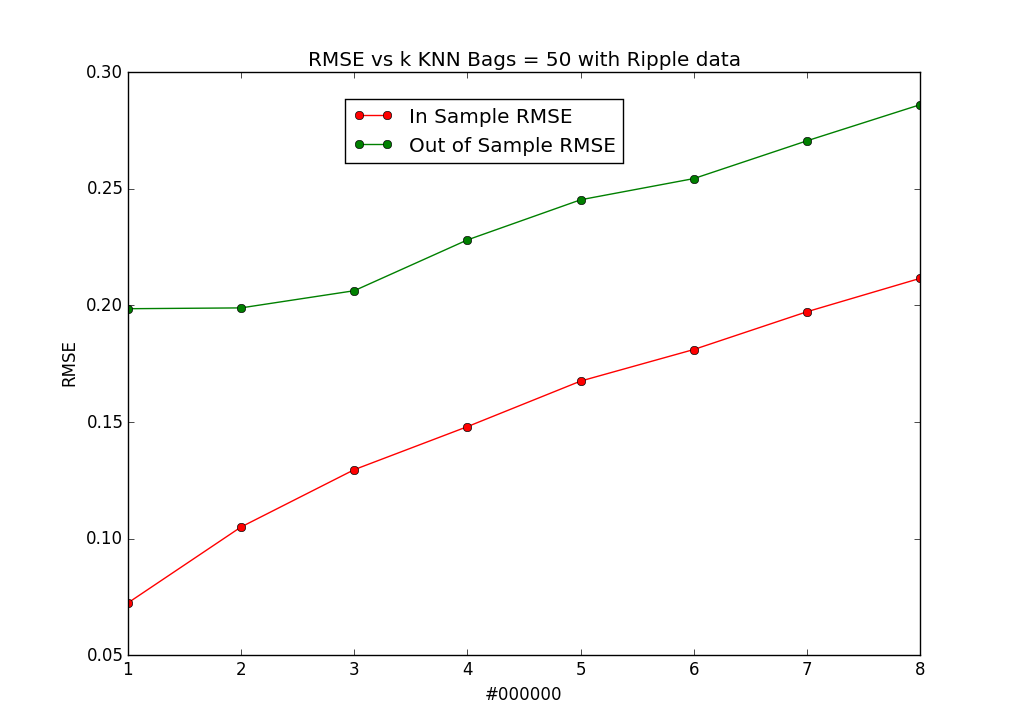
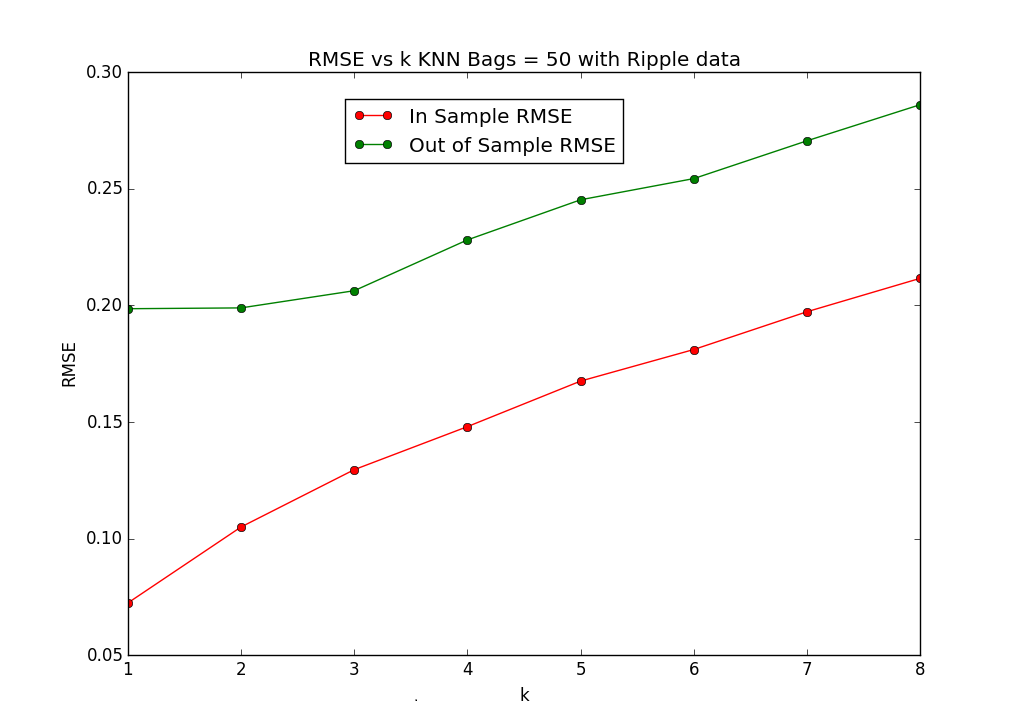
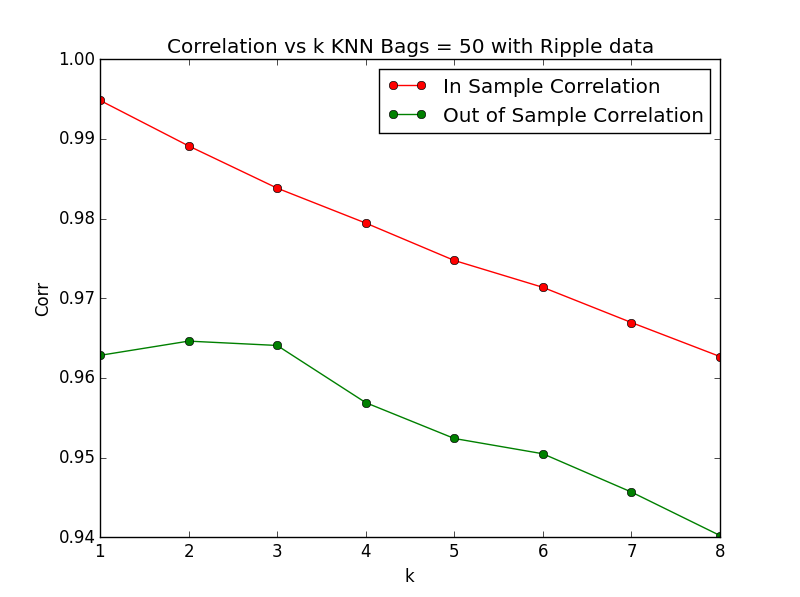
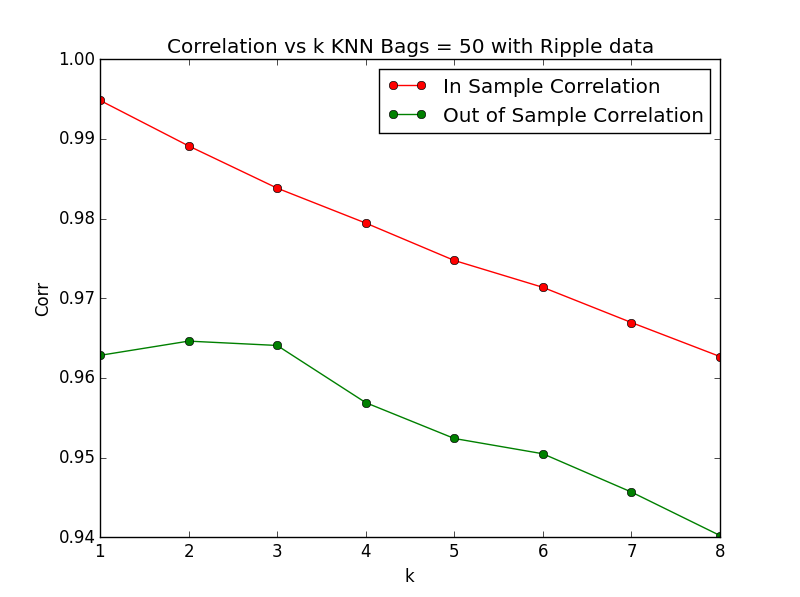
Can bagging reduce or eliminate overfitting with respect to K for the ripple dataset?





As we see from the RMSE vs k graph above, bagging greatly reduces overfitting with respect to K for the ripple dataset, as bagging will compensate for overfitting for low k.  It appears that bagging is completely eliminated.

As k moves to less than 3, the out of sample RMSE actually slightly decreases instead of increasing when bagging wasn’t used



Summary results:

Number of bags: 50

k = 1 inRMSE= 0.0723566318356 inCorr= 0.99483902148 outRMSE= 0.198546942473 outCorr= 0.96283099935

k = 2 inRMSE= 0.104943173811 inCorr= 0.989123307945 outRMSE= 0.198896104334 outCorr= 0.964618664897

k = 3 inRMSE= 0.129525583607 inCorr= 0.983819689854 outRMSE= 0.20626428804 outCorr= 0.964073082818

k = 4 inRMSE= 0.147984039857 inCorr= 0.97943572872 outRMSE= 0.228058971662 outCorr= 0.956898094932

k = 5 inRMSE= 0.1675179645 inCorr= 0.974744184974 outRMSE= 0.245263380373 outCorr= 0.952398331189

k = 6 inRMSE= 0.181055268527 inCorr= 0.971370427606 outRMSE= 0.254354282721 outCorr= 0.950478959485

k = 7 inRMSE= 0.197262416073 inCorr= 0.966951719036 outRMSE= 0.270560862423 outCorr= 0.945671668702

k = 8 inRMSE= 0.211585234043 inCorr= 0.96268432519 outRMSE= 0.286046442116 outCorr= 0.94021617475