

Question 1.

The least squares problem requires the concept of Singular Value Decomposition (SVD). We analyse different solutions. We might get a unique solution, the closest solution, or a set of solutions, from which a smart solution can be used. We discuss the pseudo inverse and how it would help in getting the closest inverse to A , in the case A is not square. After getting the pseudo inverse, we analyse the stability of the pseudo inverse, when a perturbation is introduced while observing y . Applying pseudo inverse to this y will get us a least squares estimate. This is compared to the case when the y is clean, without any noise. The difference between the clean reconstruction and the least squares (with noise) reconstruction (both done using pseudo inverse) is found to be dependent on the singular values of A (which can be obtained through the SVD). We see that the reconstruction can be bad if the smallest singular value is too small. And if the error is additive Gaussian white noise, the average reconstruction error blows up if the smallest singular value is too small. To avoid this, we truncate the terms of the SVD and thus eliminate the possibility of the error term blowing up. The reconstruction error now (between the truncated reconstruction & clean reconstruction) depends on the truncation limit that we have taken (taking this limit depends on the problem we are solving and is mainly intelligent guesswork).

Another type of stable reconstruction is the Tikhonov regularization procedure. This can be interpreted using optimization and it can be computed without the direct computation of the SVD. A new inverse is introduced and we get the noise to be upper bounded. Since the SVD is not explicitly calculated here, computations are cheap even in the case of bigger matrices. In the case of Kernel regression, we proceed by calculating the kernel matrix and using it to develop an inverse (with a delta value dependent on the eigen values of the kernel matrix). This gives us an estimate of the function and we compare it with the original function and estimate errors. We also looked at the different types of kernels used commonly in machine learning. These include polynomial kernels, radial basis kernels and sigmoid kernels. All these methods will work efficiently if the matrix dimensions are considerably small. If not, it would be best to simplify matrices using Cholesky decomposition, QR decomposition, etc. We delve into these topics in the coming classes.