Singular Value Decomposition

When to use SVD:

- When the L2 norm of the features matrix is not invertible
- High dimensions and low number of observations (M+1 > N)
- The features matrix is ill-conditioned
- A method for ordering the dimensions along which data points exhibit the greateat variance
- Method for data reduction: once we have determined the dimensions capturing most of the variance, one can find the best approximation for the original data
- Application in NLP: Ignore variance below a threshold to reduce size of data (e.g. noise?)

Implementation Considerations:

- The rank of the design matrix (which is equal to the L2 norm of the design matrix), is less than or equal to the minimum betweem M+1 and N (e.g. the number of non-zero eigen values)
- When r = M+1 - the MLE is unique
- When r < M+1 and r < N, there will be infinitely many MLE's
- In the case where N < M+1 and r = N, result is overfitting (e = 0, qmle = inf)

Resources:

- SVD_regression.pdf
- Singular*Value*Decomposition_Tutorial.pdf (page 14)

To Read:

- Bourlard, H. and Y. Kamp (1988). Auto-association by multilayer perceptrons and singular value de- composition. Biological Cybernetics 59, 291–294.
- Resources in Singular*Value*Decomposition_Tutorial.pdf