Name: Hongda Li   Class 546 SPRING 2021   HW4A

A1: Conceptual Questions

A1.a

True. This is true because SVD is looking for a orthogonal matrix (PCA uses SVD) U, such that $U\Sigma V^T$ minimizes that reconstruction error. In this case the rank of matrix U can have the same rank as the subspace span by the columns of the data matrix X giving us zero reconstruction errors.

A1.b

True Because:

$$\begin{aligned} X^TX &= (USV^T)^T(USV) \\ &= VS^TU^TUSV \\ &= VS^TSV \end{aligned} \tag{A1.b.1}$$

Take notice that S^TS is diagonal and V is orthogonal, and X^TX is Symmetric. By properties of Hermitian Adjoint Matrices, it has orthogonal Eigen Decomposition with unique real eigenvectors. And VS^TSV matches it, therefore V is the eigen vectors of matrix X^TX.

A1.c

False. The objective should not be choosing k to minimize the Loss because if $k = n$ is always the global minimum in that case and it doesn't provide any useful interpretations on the data.

A1.d

False. Singular values Decomposition has U, V that are the eigenvectors for XX^T and X^TX, eigen values decomposition is not unque because you can multiply eigenvector by negative one (or even worse by the complex unit $exp(i\theta)$) to get another normalized eigen vector that still works. In the case of SVD remember to flip the u, v vector corssponds to the same singular value together to get different decomposition for the same matrix.

A1.e

False when the matrix is degenerate. In this case a eigenvalue can have a geometric multiplicity higher than it's algebaric multiplicity, then the rank of the matrix will be more than the number of eigenvalues it has.

A1.f

True, becaues Autoencoders with non-linear activation function can incooperate non-linear representation of data in the lower dimension.

A2: Basics of SVD and Subgradients

PCA

Unsupervised Learning with Autoencoders

K-means Clustering

ML in the Real World