1. Steps:
2. Calculate the mean value of each feature
3. For all the four samples, calculate mean modified value of all its 19 features to form a new dataset A.
4. Compute covariance matrix by C=A\*A’
5. Find eigenvectors of C, then figure out the non-zero ones.

Here are three non-zero eigenvectors:

-0.0663 -0.0412 0.8625

-0.7904 0.0682 -0.3473

0.4729 0.6912 -0.2205

0.3838 -0.7182 -0.2947

1. Steps:
2. Eigen space U=A\*non-zero eigenvectors
3. Project test data onto the eigen space: omega=U’\*A
4. Normalize omega

omega =

-0.8782 -0.4410 8.3012

-10.4722 0.7295 -3.3429

6.2651 7.3907 -2.1218

5.0854 -7.6791 -2.8364

1. Transformed data=omega\*U+m

MSE (for 4 samples)=

2.94565916757332e-31

4.53465273642670e-31

4.57172507876020e-31

8.94267602986467e-32

1. MSE=

3.62680441229197

0.588160871225716

0.236958595003804

0.423432196798639

1. Distance1=12.9228

Distance2=5.6569

Distance3=16.7929

Distance4=15.7797

So sample2 is most close to the new vector.

1. Distance1=12.9228

Distance2=5.6569

Distance3=16.7929

Distance4=15.7797

The result matches. It makes sense because it means the reduced dimensions do not have much influences on final result. So it is reasonable to do such kind of dimensional reduction. We only reduced dimensions which don’t contain much useful information, so we can get same result when comparing to a new vector.