Assignment 7 - General Comments

Q1: The vast majority of students got full marks. The clue is to pay attention to the dimension of each variable and whether the result is a scalar or matrix (it depends on which version that you were solving with regard to announced clarification).

Q2: Many students failed to prove that diagonal elements in matrix **C** is positive, i.e., >0, and does not include any 0. The correct answer is that since matrix **A** is of full-rank, elements in **A** must be all positive, which subsequently leads to non-zero elements along the diagonal of **C**.

We understand this may be difficult for many to prove. Therefore, after sending you the individual feedback, we discussed and decided to treat this as a bonus point (showing >0 rather than >=0) instead. We added 5 marks (0.35 after weighting) to ALL students to remove the effect of this mark on your total. For those who answered it correctly, you will get an additional bonus of 5 marks (0.35 after weighting). Those (6 of you) scored full mark will get 105, equivalent to 7.35 after weighting.

Q3:
3b: Some students plotted the ratio of variance instead of the variances. We understand the ratios are more interpretable, while the question asked the variance rather than the ratio or percentage.
3c: Many students managed to visualize the eigenvectors, but failed to give observations.
3d: Many students had troubles in calculating the mean squared error because the reconstruction of the *x_train_mnist_approx* was incorrect. An easy way is to use an existing function in the sklearn library (inverse_transform()). You may also perform the reconstruction via first performing transform() and then doing np.dot(). But please pay attention that when in PCA transformation (using transform()), the mean is subtracted (which is called centering, see line 130). Thus in reconstruction, you need to add the mean back when using np.dot() function. In the build-in inverse_transform(), the mean value is added back in the function (see line 157). This website provides some additional discussion on this.

Q4: Some students lost marks because observations on the plot were overlooked or there were mistakes in clustering.

--

This feedback was written by Yan Ge, Hardy Hardy, Fariba Yousefi and Haiping Lu.