Hint for software Problem 9.6

Assume we have samples of a random variable X as an n-dimensional vector, and samples of a feature vector $\underline{Y}$ as an n by k matrix, where each row of the matrix contains a sample. We want to estimate the following quantities for computing a Linear Least Squares Estimator of X given $\underline{Y}$:

1. $\mathbb{E}[X]$
2. $\mathbb{E}[\underline{Y}]$
3. $\text{Var}[X]$
4. $\text{Cov}[\underline{Y}] \triangleq \mathbf{\Sigma}_{\underline{Y}}$
5. Cross-covariance $\text{Cov}[X, \underline{Y}] \triangleq \mathbf{\Sigma}_{X,\underline{Y}}$

An easy way to compute these using the built-in functions in either MATLAB or PYTHON is to use the mean and cov functions. These would easily compute the first 4 statistics, but the cross-covariance does not have a built-in function.

We can solve this problem by defining an augmented data set [X , $\underline{Y}$] which is now an n by (k+1) matrix. When take the average over the rows of that augmented data, we get

mean([X, $\underline{Y}$]) = $[\mu_X , \mu_{\underline{Y}}]$

When we take the covariance of the augmented data set, we get

$$\text{cov}([X, \underline{Y}]) = \begin{bmatrix} \mathbf{\Sigma}_X & \mathbf{\Sigma}_{X,\underline{Y}} \\ \mathbf{\Sigma}_{X,\underline{Y}}^T & \mathbf{\Sigma}_{\underline{Y}} \end{bmatrix}$$

Where $\mathbf{\Sigma}_X$ is a scalar, $\mathbf{\Sigma}_{X,\underline{Y}}$ is a 1 by k vector, and $\mathbf{\Sigma}_{\underline{Y}}$ is a x by x matrix. This augmentation trick allows you to compute the cross-covariance using the regular covariance command.