## 2.

### i.

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}(X)(Y-\mathbb{E}(Y)] \\
&= \mathbb{E}[XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) - \mathbb{E}(X)\mathbb{E}(Y)] \\
&= \mathbb{E}(XY) - \mathbb{E}[X\mathbb{E}(Y)] - \mathbb{E}[Y\mathbb{E}(X)] - \mathbb{E}[\mathbb{E}(X)\mathbb{E}(Y)] \text{ (linearity of expectation)} \\
&= \mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) \ (\mathbb{E}(C) = C, C \in \mathbb{R}) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(XY) \text{ (def.)} \\
&= 0
\end{aligned}
$$

□

### ii.

$$
\begin{aligned}
C' &= \mathbb{E}\left[(\tilde{x}_i - \mathbb{E}(\tilde{x}_i))\,(\tilde{x}_i - \mathbb{E}(\tilde{x}_i)^\top\right] \\
&= \mathbb{E}\left[(\tilde{x}_i)(\tilde{x}_i)^\top\right] \text{ (given } \mathbb{E}(\tilde{x}_i) = 0, \forall i) \\
&= \mathbb{E}\left[\Lambda^{-1/2}U^\top(x_i-\mu)\left(\Lambda^{-1/2}U^\top(x_i-\mu)\right)^\top\right] \\
&= \mathbb{E}\left[\Lambda^{-1/2}U^\top(x_i-\mu)(x_i-\mu)^\top U\left(\Lambda^{-1/2}\right)^\top\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\Lambda^{-1/2}U^\top(x_i-\mu)(x_i-\mu)^\top U\Lambda^{-1/2} \text{ (definition of } \mathbb{E}(\cdot)) \\
&= \Lambda^{-1/2}U^\top\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^\top U\Lambda^{-1/2} \\
&= \Lambda^{-1/2}U^\top CU\Lambda^{-1/2} \text{ (definition of } C) \\
&= \Lambda^{-1/2}U^\top U\Lambda U^\top U\Lambda^{-1/2} \text{ (singular value decomposition of } C) \\
&= \Lambda^{-1/2}I\Lambda I\Lambda^{-1/2} \text{ (associativity)} \\
&= \Lambda^{-1/2}\Lambda\Lambda^{-1/2} \\
&= I
\end{aligned}
$$

□

### iii.

In this question, the dimension with the largest variance is selected in PCA to construct new coordinates to perform data whitening. Since the obtained covariance matrix is the identity matrix, it means that the whitened data is only relevant to itself and not to other data vectors. In summary, by finding principal components, PCA identifies the direction in which variables are irrelevant in the data space, thereby realizing the decorrelation process. The obtained covariance matrix reflects the degree of correlation of data variables after PCA is executed. The identity matrix means that the principal components are found and the correlations are perfectly separated.