Section 6.

1. ① Normalize the mean and variance of X.

② Calculate covariance of X, cov(X).

③ Calculate the eigenvalue and eigenvectors of cov(X).

④ Arrange the eigenvectors according to the eigenvalues from large to small and pick the (D−1) largest ones, make them as matrix P.

⑤ $Y = PX$ is the result.

2. $\tan(\theta) = \frac{-0.71}{-0.71} = 1 \quad \theta = 45°$

3. $V = \begin{bmatrix} -0.71 & -0.71 \\ 0.71 & -0.71 \end{bmatrix} \quad a = \begin{bmatrix} 316 \\ 1683 \end{bmatrix}$

$-0.71 \times 316 - 0.71 \times 1683 + 0.71 \times 316 - 0.71 \times 1683 = -1419.29 - 970.57$

$= -2389.86$

$-0.71 \times 316 - 0.71 \times 1683 = -1419.29$

$Loss = 1 - \frac{1419.29}{2389.86} \approx 0.406 = 40.6\%$

4. No. PCA may result in the loss of valid information, and thus decrease the accuracy.

5. Yes. PCA calculate the eigenvalues and eigenvectors of the data matrix, and find a diagonal representation. When calculating again, the eigenvalues and eigenvalues of the diagonal matrix is the same with the original one.