

Exercises

§8.3: Principal Component Analysis

Python: There is an optional starter script that supports Python-based problem(s) on this sheet. Use it as scaffolding: read the comments, complete the lines marked **TODO**, and compare the output with the mathematics requested in the problem(s).

Problem 1 Let $C = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$. For a unit vector $\mathbf{u} = (\cos\theta, \sin\theta)^T$, compute $\mathbf{u}^T C \mathbf{u}$. For which directions is it maximized? Identify the first principal component and its captured variance.

Problem 2 Starting from $\max_{\mathbf{u}} \mathbf{u}^T C \mathbf{u}$ subject to $\mathbf{u}^T \mathbf{u} = 1$, use a Lagrange multiplier to derive $C \mathbf{u} = \lambda \mathbf{u}$. Then show that $\mathbf{u}^T C \mathbf{u} = \lambda$ for a unit eigenvector.

Problem 3 For the covariance matrix $C = \begin{bmatrix} 5/4 & 3/4 \\ 3/4 & 1/2 \end{bmatrix}$, verify that its eigenvalues are $\frac{7 \pm 3\sqrt{5}}{8}$. Compute their sum and product and compare with $\text{tr}(C)$ and $\det(C)$.

Problem 4 The eigenvalues of a centered dataset's covariance matrix are 8, 3, 1, 0.5. Compute: (a) total variance, (b) explained-variance ratio of each component, (c) cumulative explained variance after two components, (d) mean squared reconstruction error after keeping two components.

Problem 5 Suppose the covariance matrix has eigenpairs $(\lambda_1, \mathbf{u}_1)$ and $(\lambda_2, \mathbf{u}_2)$, with $\lambda_1 > \lambda_2$ and $\mathbf{u}_1^T \mathbf{u}_2 = 0$. Let $\mathbf{u} = a\mathbf{u}_1 + b\mathbf{u}_2$ with $a^2 + b^2 = 1$. Show that $\mathbf{u}^T C \mathbf{u} = a^2 \lambda_1 + b^2 \lambda_2$. Use this to explain why $\mathbf{u}_1$ maximizes the projected variance.

Problem 6 Show that if $C = (1/P)X_c^T X_c$, then $\text{tr}(C)$ equals $\frac{1}{P} \sum_{p=1}^P \|\mathbf{q}_p\|^2$. Interpret this quantity geometrically.

Problem 7 Use Python to perform PCA on

```
x = np.array([ [0., 1.],  
               [1., 2.],  
               [2., 2.],  
               [3., 3.] ])
```

using np. Report: (a) the covariance matrix, (b) the eigenvalues in descending order, (c) the first principal direction, (d) the explained-variance ratios, (e) the one-component mean squared reconstruction error.

Problem 8 Suppose feature 1 is multiplied by 100 while feature 2 is unchanged. Explain qualitatively why the PCA direction may change dramatically. What preprocessing could be considered if the two features should be given comparable scale? Why is that preprocessing not automatically correct in every application?