

Exercises - Solutions

§8.3: Principal Component Analysis

Problem 1 Let $\mathbf{u} = (\cos\theta, \sin\theta)^T$. Then $\mathbf{u}^T C \mathbf{u} = 4\cos^2\theta + \sin^2\theta$. Since $\sin^2\theta = 1 - \cos^2\theta$: $\mathbf{u}^T C \mathbf{u} = 1 + 3\cos^2\theta$. This is maximized when $\cos^2\theta = 1$. Thus $\theta = 0$ or π modulo 2π . The principal direction is therefore $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ up to sign. Captured variance: $\lambda_1 = 4$.

Problem 2 Define $L = \mathbf{u}^T C \mathbf{u} - \lambda(\mathbf{u}^T \mathbf{u} - 1)$. Differentiate with respect to $\mathbf{u}$: $\nabla_{\mathbf{u}} L = 2C\mathbf{u} - 2\lambda\mathbf{u} = \mathbf{0}$. Therefore $C\mathbf{u} = \lambda\mathbf{u}$. If $\|\mathbf{u}\| = 1$: $\mathbf{u}^T C \mathbf{u} = \mathbf{u}^T (\lambda\mathbf{u}) = \lambda\|\mathbf{u}\|^2 = \lambda$.

Problem 3 Characteristic polynomial: $\det(C - \lambda I) = \left(\frac{5}{4} - \lambda\right)\left(\frac{1}{2} - \lambda\right) - \frac{9}{16}$. Expanding: $\lambda^2 - \frac{7}{4}\lambda + \frac{1}{16}$. Quadratic formula: $\lambda = \frac{\frac{7}{4} \pm \sqrt{\frac{49}{16} - \frac{1}{4}}}{2} = \frac{\frac{7}{4} \pm \frac{3\sqrt{5}}{4}}{2} = \frac{7 \pm 3\sqrt{5}}{8}$. Sum: $\lambda_1 + \lambda_2 = \frac{14}{8} = \frac{7}{4} = \text{tr}(C)$. Product: $\lambda_1 \lambda_2 = \frac{49 - 45}{64} = \frac{1}{16} = \det(C)$.

Problem 4 Eigenvalues: 8, 3, 1, 0.5. (a) Total variance: $8 + 3 + 1 + 0.5 = 12.5$. (b) Explained-variance ratios: 0.64, 0.24, 0.08, 0.04. (c) After two components: $\frac{8+3}{12.5} = 0.88$. So the first two components explain 88% of total variance. (d) Mean squared reconstruction error after keeping two components: $1 + 0.5 = 1.5$.

Problem 5 Since $C\mathbf{u}_1 = \lambda_1\mathbf{u}_1$ and $C\mathbf{u}_2 = \lambda_2\mathbf{u}_2$: $C\mathbf{u} = a\lambda_1\mathbf{u}_1 + b\lambda_2\mathbf{u}_2$. Then $\mathbf{u}^T C \mathbf{u} = (a\mathbf{u}_1 + b\mathbf{u}_2)^T (a\lambda_1\mathbf{u}_1 + b\lambda_2\mathbf{u}_2)$. Orthogonality removes the cross terms: $\mathbf{u}^T C \mathbf{u} = a^2\lambda_1 + b^2\lambda_2$. Since $a^2 + b^2 = 1$ and $\lambda_1 > \lambda_2$: the expression is largest when $a^2 = 1$ and $b^2 = 0$. Therefore $\mathbf{u} = \pm\mathbf{u}_1$ maximizes projected variance.

Problem 6 Since $C = (1/P)X_c^T X_c$: $\text{tr}(C) = \frac{1}{P} \text{tr}(X_c^T X_c)$. The trace of $X_c^T X_c$ is the sum of squares of all entries of X_c . Therefore $\text{tr}(C) = \frac{1}{P} \sum_{p=1}^P \|\mathbf{q}_p\|^2$. Thus $\text{tr}(C) = \text{mean squared distance of centered observations from the origin}$. Since the centered origin is the data mean, this is the total feature variance summed across all directions.

Problem 7

```
import numpy as np

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

mu = np.mean(X, axis=0)
Xc = X - mu

C = Xc.T @ Xc / len(X)

eigenvalues, eigenvectors = np.linalg.eigh(C)

order = np.argsort(eigenvalues)[-1:]
```

```

eigenvalues = eigenvalues[order]
eigenvectors = eigenvectors[:, order]

explained = eigenvalues / np.sum(eigenvalues)

u1 = eigenvectors[:, 0]

z1 = Xc @ u1
Xc_hat = np.outer(z1, u1)

mse = np.mean(
    np.sum((Xc - Xc_hat)**2, axis=1)
)

print(C)
print(eigenvalues)
print(eigenvectors)
print(explained)
print(mse)

```

(a) $C = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 0.50 \end{bmatrix}$. (b) $\lambda \approx (1.71352549, 0.03647451)$. (c) One valid first principal direction is

$\mathbf{u}_1 \approx (0.85065081, 0.52573111)^T$. Its negative is equally correct. (d) $(0.97915742, 0.02084258)$. (e) One-component mean squared reconstruction error: 0.03647451. This equals the omitted eigenvalue.

Problem 8 Multiplying feature 1 by 100 multiplies its variance by $100^2 = 10,000$. It also rescales its covariance terms. Thus feature 1 may dominate the covariance matrix and pull the first principal direction strongly toward its axis. If the two feature scales are arbitrary and should be treated comparably, one could standardize each feature to mean 0 and standard deviation 1. But standardization changes distances, variances, and therefore the PCA geometry. If the original units and variances are scientifically meaningful, erasing those scale differences may be inappropriate.