1.1 设 $X=(X_1,\cdots,X_m)^\top$ 是 m 维随机变量，均值为 $E(X)\stackrel{\text{def}}{=}\mu$, 协方差矩阵为 $\operatorname{cov}(X)\stackrel{\text{def}}{=}\Sigma$。设 Σ 的特征值为 $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_m$，特征值对应的单位特征向量为 $\alpha_1,\cdots,\alpha_m$ 则 X 的第 k 个主成分是 $Y_k=\alpha_k^\top X$, 方差为 $\operatorname{var}(Y_k)=\alpha_k^\top\Sigma\alpha_k$.

证明以下性质：

$$\sum_k \rho^2(Y_k,X_i)=1$$

其中，$\rho(Y_k,X_i)=\frac{\sqrt{\lambda_k}\alpha_{ki}}{\sqrt{\sigma_{ii}}}$, $\sigma_{ii}=\operatorname{var}(X_i)$, $\alpha_{ki}=e_i^\top\alpha_k$, e_i 为基本单位向量, 其第 i 个变量为 1，其余为 0。

证： $$\rho^2(Y_k,X_i)=\frac{\operatorname{cov}^2(Y_k,X_i)}{\operatorname{Var}(Y_k)\operatorname{Var}(X_i)}=\frac{(\alpha_k^\top\Sigma e_i)^2}{(\alpha_k^\top\Sigma\alpha_k)\cdot(e_i^\top\Sigma e_i)}$$

$$=\frac{\lambda_k^2\cdot\alpha_{ki}^2}{\lambda_k\cdot\sigma_{ii}}=\frac{\lambda_k\alpha_{ki}^2}{\sigma_{ii}}$$

$$\therefore \rho(Y_k,X_i)=\frac{\sqrt{\lambda_k}\,\alpha_{ki}}{\sqrt{\sigma_{ii}}}$$

$$\sum_k\rho^2(Y_k,X_i)=\sum_k\frac{\lambda_k\cdot\alpha_{ki}^2}{\sigma_{ii}}=\sum_k\frac{tr(\alpha_k^\top\lambda_k e_i e_i^\top\alpha_k)}{\sigma_{ii}}$$

$$=\sum_k\frac{tr(\alpha_k\alpha_k^\top\lambda_k e_i e_i^\top)}{\sigma_{ii}}$$

$$=\frac{tr(\Sigma e_i e_i^\top)}{\sigma_{ii}}=\frac{e_i^\top\Sigma e_i}{\sigma_{ii}}=1$$

1.2 对以下样本数据进行主成分分析:

$$X = \begin{array}{c} \\ \\ \\ \end{array}\begin{bmatrix} 2 & 2 \\ 4 & 3 \\ 5 & 3 \\ 5 & 4 \\ 6 & 5 \\ 8 & 7 \end{bmatrix} \begin{array}{l} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array}$$

解: $\hat{\Sigma} = \frac{1}{n-1}\sum_i (X_i - \bar{X})(X_i - \bar{X})^T$ $\quad \bar{X} = \frac{1}{n}\sum X_i$

$= \begin{pmatrix} 4 & \frac{17}{5} \\ \frac{17}{5} & \frac{16}{5} \end{pmatrix}$ $\quad = (5 \quad 4)$

$S_{11}^2 = \frac{1}{n-1}\sum (X_{i1} - \bar{X}_1)^2 = 4$ $\quad \therefore \tilde{\Sigma} = \begin{pmatrix} 1 & \frac{17}{8\sqrt{5}} \\ \frac{17}{8\sqrt{5}} & 1 \end{pmatrix}$

$S_{22}^2 = \frac{1}{n-1}\sum (X_{i2} - \bar{X}_2)^2 = \frac{16}{5}$

$q_1 = (-0.71, -0.71)^T$

$q_2 = (0.71, -0.71)^T$

first component : $-0.71X_1 - 0.71X_2$

second component: $0.71X_1 - 0.71X_2$