Exercise 2: Principal Component Analysis (10 + 10 P)

We consider a dataset $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^d$. Principal component analysis searches for a unit vector $\boldsymbol{u} \in \mathbb{R}^d$ such that projecting the data on that vector produces a distribution with maximum variance. Such vector can be found by solving the optimization problem:

$$\arg\max_{\boldsymbol{u}} \frac{1}{N}\sum_{k=1}^{N}\Big[\boldsymbol{u}^\top \boldsymbol{x}_k - \frac{1}{N}\Big(\sum_{l=1}^{N}\boldsymbol{u}^\top \boldsymbol{x}_l\Big)\Big]^2 \quad \text{with} \quad \|\boldsymbol{u}\|^2 = 1$$

(a) *Show* that the problem above can be rewritten as

$$\arg\max_{\boldsymbol{u}} \boldsymbol{u}^\top \boldsymbol{S}\boldsymbol{u} \quad \text{with} \quad \|\boldsymbol{u}\|^2 = 1$$

where $\boldsymbol{S} = \sum_{k=1}^{N}(\boldsymbol{x}_k - \boldsymbol{m})(\boldsymbol{x}_k - \boldsymbol{m})^\top$ is the scatter matrix, and $\boldsymbol{m} = \frac{1}{N}\sum_{k=1}^{N}\boldsymbol{x}_k$ is the empirical mean.

$$\frac{1}{N}\sum_{k=1}^{N}\Big(u^T x_k - \frac{1}{N}\cdot\sum_{l=1}^{N}u^T x_l\Big)^2 =$$

$$= \frac{1}{N}\sum_{k=1}^{N}\Big(u^T x_k u^T x_k - \frac{1}{N}\sum_{l=1}^{N}u^T x_k u^T x_l - \frac{1}{N}\sum_{l=1}^{N}u^T x_l u^T x_k + \frac{1}{N^2}\sum_{l=1}^{N}u^T x_l \cdot \sum_{l=1}^{N}u^T x_l\Big)$$

$$= \frac{1}{N}\sum_{k=1}^{N}\Big(\underbrace{u^T x_k x_k^T u}_{\substack{\text{this is}\\ \text{a scalar so}\\ u^T x_k = x_k^T u}} - u^T x_k \cdot \underbrace{\frac{1}{N}\sum_{l=1}^{N}x_l^T}_{m^T} u - u^T \cdot \underbrace{\frac{1}{N}\sum_{l=1}^{N}x_l}_{m} \cdot x_k^T u + u^T \cdot \underbrace{\frac{1}{N}\sum_{l=1}^{N}x_l}_{m} \cdot \underbrace{\frac{1}{N}\sum_{l=1}^{N}x_l^T}_{m^T} \cdot u\Big)$$

$$= \frac{1}{N}\sum_{k=1}^{N}u^T\big(x_k x_k^T - x_k m^T - m x_k^T + m m^T\big)u$$

$$= \frac{1}{N}\sum_{k=1}^{N}u^T(x_k - m)(x_k - m)^T u$$

$$= u^T S u$$

(b) *Show* using the method of Lagrange multipliers that the problem above can be reformulated as solving the eigenvalue problem

$$\boldsymbol{S}\boldsymbol{u} = \lambda \boldsymbol{u}$$

and retaining the eigenvector $\boldsymbol{u}$ associated to the highest eigenvalue λ.

Problem: $\arg\max_u (u^T S u)$ s.t. $\|u\|^2 = 1$

Let us construct the Lagrangian: $\mathcal{L}(u,\lambda) = u^T S u + \lambda(1 - \|u\|^2)$

- $u^T S u = (u_1 \ldots u_d)\begin{pmatrix} s_{11} & \cdots & s_{1d} \\ \vdots & & \vdots \\ s_{d1} & \cdots & s_{dd}\end{pmatrix}\begin{pmatrix} u_1 \\ \vdots \\ u_d\end{pmatrix} = (u_1 \ldots u_d)\begin{pmatrix} \sum_{i=1}^{d} s_{1i}u_i \\ \vdots \\ \sum_{i=1}^{d} s_{di}u_i\end{pmatrix} = $
$= \sum_{j=1}^{d}\sum_{i=1}^{d} u_j s_{ji} u_i$
- $\|u\|^2 = u^T u = \sum_{j=1}^{d} u_j^2$

$$\nabla_u \mathcal{L}(u,\lambda) = \nabla_u \left(\sum_{j=1}^{d} \sum_{i=1}^{d} u_j s_{ji} u_i \right) + \lambda \nabla_u \left(1 - \sum_{j=1}^{d} u_j^2 \right)$$

$$= \sum_{j=1}^{d} \sum_{i=1}^{d} \sum_{k=1}^{d} \frac{\partial}{\partial u_k} (u_j s_{ji} u_i) \hat{u}_k - \lambda \sum_{j=1}^{d} \sum_{k=1}^{d} \frac{\partial u_j^2}{\partial u_k} \hat{u}_k$$

$$= \sum_{j=1}^{d} \sum_{i=1}^{d} \sum_{k=1}^{d} \left(\delta_{jk} s_{ji} u_i + u_j s_{ji} \delta_{ik} \right) \hat{u}_k - 2\lambda \underbrace{\sum_{j=1}^{d} u_j \hat{u}_j}_{u \text{ vector}}$$

$$= \sum_{j=1}^{d} \sum_{i=1}^{d} \left(s_{ji} u_i \hat{u}_j + u_j s_{ji} \hat{u}_i \right) - 2\lambda u$$

$$= \sum_{j=1}^{d} \sum_{i=1}^{d} \left(s_{ji} u_i \hat{u}_j + u_i s_{ij} \hat{u}_j \right) - 2\lambda u$$

S is symmetric

$$= 2 \sum_{j=1}^{d} \sum_{i=1}^{d} s_{ji} u_i \hat{u}_j - 2\lambda u$$

$$= 2 \begin{pmatrix} \sum_{i=1}^{d} s_{1i} u_i \\ \vdots \\ \sum_{i=1}^{d} s_{di} u_i \end{pmatrix} - 2\lambda u = 2Su - 2\lambda u$$

Solve for $\nabla_u \mathcal{L}(u,\lambda) = 0$:

$$Su = \lambda u$$

$$\nabla_\lambda \mathcal{L}(u,\lambda) = 0 \iff \|u\|^2 = 1$$

$$\underset{\|u\|^2=1}{\operatorname{argmax}} (u^T S u) = \underset{\|u\|^2=1}{\operatorname{argmax}} (u^T \lambda u) = \underset{\|u\|^2=1}{\operatorname{argmax}} (\lambda \|u\|^2) = \lambda$$

To maximize $u^T S u$, we must take the greatest eigenvalue λ of S and its corresponding eigenvector $u \in \mathbb{R}^d$ with $\|u\|^2 = 1$