

Exercise Sheet 3 (theory part)

Exercise 1: Principal Component Analysis (15 + 15 P)

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

$$\arg \max_{\mathbf{u}} \frac{1}{N} \sum_{i=1}^N [\mathbf{u}^\top (\mathbf{x}_i - \mathbf{m})]^2 \quad \text{with} \quad \|\mathbf{u}\|^2 = 1 \quad \text{and} \quad \mathbf{m} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i.$$

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

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

where $\Sigma = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^\top$ is the covariance matrix.

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

$$\Sigma \mathbf{u} = \lambda \mathbf{u}$$

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

Exercise 2: Bounds on Eigenvalues (10 + 10 P)

Let λ_1 denote the largest eigenvalue of the matrix Σ . The eigenvalue λ_1 measures the variance of the data when projected on the first principal component. We study how the latter can be bounded with the diagonal elements of the matrix Σ .

(a) Show that $\sum_{i=1}^d \Sigma_{ii}$ is an upper bound to the eigenvalue λ_1 .

(b) Show that $\max_{i=1}^d \Sigma_{ii}$ is a lower bound to the eigenvalue λ_1 .

Exercise 3: Iterative PCA (5 + 10 + 5 P)

When performing principal component analysis, computing the full eigendecomposition of the covariance matrix Σ is typically slow, and we are often only interested in the first principal components. An efficient procedure to find the first principal component is *power iteration*. It starts with a random unit vector $\mathbf{w}^{(0)} \in \mathbb{R}^d$, and iteratively applies the parameter update

$$\mathbf{w}^{(t+1)} = \Sigma \mathbf{w}^{(t)} / \|\Sigma \mathbf{w}^{(t)}\|$$

until some convergence criterion is met. Here, we would like to show the exponential convergence of power iteration. For this, we look at the error terms

$$\mathcal{E}_k(\mathbf{w}) = \left| \frac{\mathbf{w}^\top \mathbf{u}_k}{\mathbf{w}^\top \mathbf{u}_1} \right| \quad \text{with} \quad k = 2, \dots, d,$$

will demonstrate that they all converge to zero as \mathbf{w} approaches the eigenvector \mathbf{u}_1 and becomes orthogonal to other eigenvectors. To demonstrate this, we proceed in three steps:

(a) Show that $\mathcal{E}_k(\mathbf{w}^{(t+1)}) = \left| \frac{\mathbf{w}^{(t)\top} \Sigma \mathbf{u}_k}{\mathbf{w}^{(t)\top} \Sigma \mathbf{u}_1} \right|$.

(b) Starting from the result above, show that $\mathcal{E}_k(\mathbf{w}^{(t+1)}) = |\lambda_k / \lambda_1| \cdot \mathcal{E}_k(\mathbf{w}^{(t)})$ (Hint: to show this, it is useful to rewrite the covariance matrix in terms of eigenvalues and eigenvectors, i.e. $\Sigma = \sum_{i=1}^d \mathbf{u}_i \mathbf{u}_i^\top \lambda_i$.)

(c) Starting from the result above, show that $\mathcal{E}_k(\mathbf{w}^{(T)}) = |\lambda_k / \lambda_1|^T \cdot \mathcal{E}_k(\mathbf{w}^{(0)})$, i.e. the convergence of the algorithm is exponential with the number of time steps T .