Exercise1: ‘Kernel properties’ (10 points) You want to center the inner-product kernel matrix K (euclidean inner product) of a data set afterwards (you forgot to do it on the original data and do not have it anymore, or your data are from a structured domain). You can use some random or real world data (make sure they are not already mean-centered) to generate a kernel matrix e.g. using a standard inner-product kernel.

1. Provide some python code to illustrate the relation of a kernel on centered data and the eigen-decomposition of uncentered data. How do some results of the eigendecomposition change / or not change?

   (In Code File)

2. Prove using the eigenvalue equation, that a matrix K can be changed as if the original data had been mean centered

   The eigenvalue equation for a centered kernel matrix (K') can be derived from the eigendecomposition of the centered kernel matrix (K_c). Given a kernel matrix (K) and its centered version (K_c), the relationship between their eigenvalues can be proven.

   Let's denote the centered kernel matrix as (K_c) and the original kernel matrix as (K).

   The eigenvalue equation for (K_c) is ($K_c v = \lambda v$), where (v) is the eigenvector and (λ) is the eigenvalue.

   The centered kernel matrix (K_c) can be expressed as:

   $$[K_c = \left(I - \frac{1}{n}11^T\right) K \left(I - \frac{1}{n}11^T\right)]$$

   Where:
   - (I) is the identity matrix.
   - ($\frac{1}{n}11^T$) is a matrix where each entry is (▯{1}{n}) ($i.e.,$ (n) is the number of rows/columns in (K)).

   Given the eigenvalue equation for (K_c), it can be expanded as:

   $$[K_c v = \left(I - \frac{1}{n}11^T\right) K \left(I - \frac{1}{n}11^T\right) v = \lambda v]$$

   Now, substitute ($K = \left(I - \frac{1}{n}11^T\right)^{-1} K_c \left(I - \frac{1}{n}11^T\right)^{-1}$) into the equation:

   $$[\left(I - \frac{1}{n}11^T\right) \cdot \left(I - \frac{1}{n}11^T\right)^{-1} K_c \cdot \left(I - \frac{1}{n}11^T\right) \cdot \left(I - \frac{1}{n}11^T\right)^{-1} v = \lambda v]$$

   Simplify the equation:

   $$[\, K\, v \,=\, \lambda v \,]$$

This proves that the eigenvalue equation for (K) is the same as that of (K_c).

3. (Extra) what does a centre of the data mean in the context of structured data (e.g. a similarity matrix based on graphs)

   The concept of centring the data can be a bit different compared to centring traditional tabular data. The notion of a centre in this context generally refers to making the data "zero mean" or adjusting the similarity matrix in a way that reflects a reference point. For a similarity matrix based on graphs or any structured data, centring typically involves adjusting the similarities or distances between elements with respect to a certain reference or "centre."

Exercise2: 'Frobenius norm of a matrix' (10 points) In machine learning we play a lot with matrices (e.g. kernels, adjacency matrices, graphlaplacian, . . .). Norms, often the Frobenius norm, can be used to score matrices and are common in the optimization of an objective function. The original formulation of the Frobenius norm requires to access all matrix elements in a non-vectorial style (loops) : to show

1. Provide some python code to calculate the Frobenius norm of a matrix A and evaluate its relation to the sum of the (squared) singular values of A

   (In Code File)

2. Prove that the Frobenius norm can indeed be calculated by means of a singular value decomposition, explain the steps.

   The Frobenius norm is defined as the square root of the sum of squares of all the elements of the matrix.

   Given a matrix (A) with SVD as $(A = U\Sigma V^T)$, where (U) is an $(m \times m)$ orthogonal matrix, (Σ) is an $(m \times n)$ rectangular diagonal matrix with singular values on the diagonal, and (V^T) is an $(n \times n)$ orthogonal matrix.

   The steps to prove that the Frobenius norm can be calculated using the SVD:

   1. Compute SVD: Obtain the singular value decomposition of matrix (A) such that $(A = U\Sigma V^T)$.

   2. Calculate Frobenius Norm Using SVD: The Frobenius norm of matrix (A) is given by:

$$[|A|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2}]$$

   Where (σ_i) represents the $(i) - th$ singular value of matrix (Σ).

The proof essentially relies on the fact that the SVD breaks down the matrix into its singular values, and the Frobenius norm utilizes these singular values to compute the norm of the original matrix by summing their squares. This relationship is a consequence of the Pythagorean theorem when considering the singular values as lengths in a higher-dimensional space.

3. (Extra) When could it be more useful to calculate the Frobenius norm by a SVD instead of using the double loop (besides of matrix /vector parallelizations)?

   1. Numerical Stability: The SVD provides a more stable and accurate way to compute the Frobenius norm, especially for matrices that are ill-conditioned or nearly singular.

   2. Dimensionality Reduction and Rank Estimation: SVD enables identifying the effective rank of a matrix by looking at the number of significant singular values. This can aid in dimensionality reduction or understanding the intrinsic dimensionality of the data.

   3. Algorithmic Requirements: Certain algorithms, especially those related to matrix factorizations, optimization, or solving linear systems, might involve SVD as a key component. In such cases, calculating the Frobenius norm through the SVD can be a natural choice, as it leverages an already performed or needed SVD.

Exercise3: 'Correction of non-psd similarities' (10 points) Prove or disprove the following statement. Adding an offset to the diagonal of a symmetric matrix changes the eigenvalues of the matrix by this value (hint: use eigenvalue equation)

1. Provide some python code to illustrate that adding an offset indeed shifts the eigenvalues

   (In Code File)

2. Prove the statement

   statement: "Adding an offset to the diagonal of a symmetric matrix changes the eigenvalues of the matrix by this value."

   Given a symmetric matrix (A) and its eigenvalues (λ), the eigenvalue equation is ($A v = \lambda v$), where (v) is the eigenvector corresponding to eigenvalue (λ).

   Now, if an offset (α) is added to the diagonal of matrix (A) to create a new matrix (B), then ($B = A + \alpha I$), where (I) is the identity matrix.

   The eigenvalue equation for matrix $(B) is (Bv = (\lambda + \alpha)v)$.

   This demonstrates that each eigenvalue of matrix (B) is equal to the corresponding eigenvalue of matrix (A) plus the added offset (α). In other words, adding an offset to the diagonal of a symmetric matrix indeed changes the eigenvalues of the matrix by this value.

This relationship is established based on the properties of the eigenvalue equation and the behaviour of diagonal elements in a matrix.

3. (Extra) What are the pros, cons, challenges of doing this in a practical setting?
   Applying an offset to the diagonal of a symmetric matrix and observing its impact on the eigenvalues holds various practical implications, accompanied by pros, cons, and challenges:

   Pros:

   1. Eigenvalue Manipulation: It allows direct manipulation of eigenvalues by simply adding an offset to the diagonal.
   2. Adjusting Parameters : It might help adjust or fine-tune problem parameters.

   Cons:

   1. Alteration of Matrix: Changing eigenvalues affects various characteristics of the matrix, which might disrupt certain relationships or properties within the matrix.

   2. Complexity in Result: Manipulating eigenvalues directly can make result interpretation more complex.

   Challenges:

   1. Effects on Stability: Modifications in eigenvalues can affect the stability of solutions in problems that rely on eigenvalue characteristics.

   2. Theoretical Interpretation: The impact of changing eigenvalues might not always be straightforward or intuitive.

Exercise4: 'Krein spaces' (10 points)

Prove or disprove the following statement. If a matrix A has positive and negative eigenvalues, then we can decompose the matrix into two submatrices with A+ +A− = A where A+ contains only the positive eigenvalues and corresponding eigenvectors and A− contains the negative eigenvalues and corresponding eigenvectors. Prove this result by means of the eigenvalue equation. Use the matrix provided in the moodle course (krein-data).

1. Provide some python code to illustrate that the claimed statement indeed holds for indefinite (non-psd) kernel matrices (use the eigen-decomposition)

   (In Code File)

2. Prove the statement.

   prove the statement using the eigenvalue equation:

   Statement: If a matrix (A) has both positive and negative eigenvalues, then we can decompose the matrix into two submatrices (A^{+}) and (A^{-}) such that$(A^{+} + A^{-} = A)$, where

(A^+) contains only the positive eigenvalues and corresponding eigenvectors, and (A^-) contains the negative eigenvalues and corresponding eigenvectors.

Proof:

1. Start with the eigenvalue equation for (A): $(Av = \lambda v), where\ (v)$ is an eigenvector, and (λ) is the corresponding eigenvalue.

2. For the positive eigenvalues of (A), we have $(Av^+ = \lambda^+ v^+)$, where (v^+) is the eigenvector corresponding to a positive eigenvalue (λ^+).

3. For the negative eigenvalues of (A), we have $(Av^- = \lambda^- v^-)$, where (v^-) is the eigenvector corresponding to a negative eigenvalue (λ^-).

4. Now, let's define two diagonal matrices (Λ^+) and (Λ^-) with the positive and negative eigenvalues, respectively, on their diagonals:

$$[\Lambda^+ = \begin{bmatrix} \lambda_1^+ & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2^+ & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3^+ & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_k^+ \end{bmatrix} [\Lambda^- = \begin{bmatrix} \lambda_1^- & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2^- & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3^- & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_l^- \end{bmatrix}]$$

Here, (k) is the number of positive eigenvalues, and (l) is the number of negative eigenvalues.

5. Create matrices (V^+) and (V^-) by stacking the corresponding eigenvectors $(v^+)$and (v^-) as columns:

$$[V^- = \begin{bmatrix} | & | & | & \cdots & | \\ v_1^- & v_2^- & v_3^- & \cdots & v_l^- \\ | & | & | & \cdots & | \end{bmatrix} [V^+ = \begin{bmatrix} | & | & | & \cdots & | \\ v_1^+ & v_2^+ & v_3^+ & \cdots & v_k^+ \\ | & | & | & \cdots & | \end{bmatrix}]$$

6. Now, we can express (A) as the sum of (A^+) and (A^-):

[
$$A = V^+\Lambda^+(V^+)^T + V^-\Lambda^-(V^-)^T = A^+ + A^-$$
]

Where(A^+) contains the positive eigenvalues and their corresponding eigenvectors, and (A^-) contains the negative eigenvalues and their corresponding eigenvectors.

This proves the statement that a matrix (A) with both positive and negative eigenvalues can be decomposed into submatrices (A^+) and (A^-) as described.

3. (Extra) Demonstrate that ||A||F = ||A+||F + ||A−||F (explain or prove why)

The Frobenius norm of a matrix (A) is defined as the square root of the sum of squares of all its elements:

$$[\ |\mathrm{A}|_{\mathrm{F}} = \sqrt{\sum_{\mathrm{i}=1}^{\mathrm{m}} \sum_{\mathrm{j}=1}^{\mathrm{n}} |\mathrm{a}_{\mathrm{ij}}|^{2}}\ ]$$

Given the decomposition of (A) into submatrices (A^{+}) and (A^{-}) containing the positive and negative eigenvalues and their corresponding eigenvectors, as demonstrated earlier:

$$[A = A^{+} + A^{-}]$$

The Frobenius norm is a unitarily invariant norm, meaning it is invariant under unitary transformations. The Frobenius norm of a matrix is equal to the square root of the sum of the squared singular values. Therefore, for the positive and negative submatrices, we can express the Frobenius norms in terms of their singular values:

$$[|A|_{F} = \sqrt{\sum_{i=1}^{m} \sigma_{i}^{2}}]$$

$$[|A^{+}|_{F} = \sqrt{\sum_{i=1}^{k} \sigma_{i}^{2}}]$$

$$[|A^{-}|_{F} = \sqrt{\sum_{i=1}^{l} \sigma_{i}^{2}}]$$

Where (k) is the number of positive eigenvalues, (l) is the number of negative eigenvalues, and (σ_{i}) represents the singular values of the corresponding submatrices.

Since the Frobenius norm of a matrix is the square root of the sum of the squared singular values, it follows that:

$$[|A|_{F} = |A^{+}|_{F} + |A^{-}|_{F}]$$

This is due to the orthogonality of the submatrices corresponding to the positive and negative eigenvalues. Therefore, the Frobenius norm of the entire matrix (A) is indeed the sum of the Frobenius norms of its positive and negative submatrices.