Multivariate Statistics – Spring 2025
Exercises

Prof. Victor Panaretos
TA: Leonardo Santoro `leonardo.santoro@epfl.ch`

Exercise 1. Recall that $\mathsf{Q} \in \mathbb{R}^{p \times p}$ is a projection matrix if it is *symmetric* ($\mathsf{Q} = \mathsf{Q}^\top$) and *idempotent* ($\mathsf{Q} = \mathsf{Q}^2$). Show that:

1. The only possible eigenvalues of a projection are 0 and 1.
2. Every vector v may be decomposed uniquely as $v = u + w$ where $u \in \text{Range}(\mathsf{Q})$ and $w \in \ker(\mathsf{Q})$.
3. If P and Q are projections onto the same subspace V, then $\mathsf{P} = \mathsf{Q}$.
4. the projection operator Q_v onto the span of a vector $v \in \mathbb{R}^p$ is $\mathsf{Q}_v = vv^\top / \|v\|^2$

Exercise 2. Suppose $\mathsf{A} = \mathsf{U}\Lambda\mathsf{V}^\top$ is the SVD of A.

1. Assuming A is invertible, find the SVD of its inverse.
2. If A is square, show that $|\det(\mathsf{A})|$ is the product of the singular values.
3. *(optional)* If A is positive definite, show that $\mathsf{U} = \mathsf{V}$.

Exercise 3. ($\star$) Recall that a matrix $\Omega \in \mathbb{R}^{p \times p}$ is said to be positive definite, in which case we write $\Omega \succ 0$, if it is symmetric and $v^\top \Omega v > 0$ for all $v \in \mathbb{R}^p$. For matrices P, Q, we say that $\mathsf{P} \succ \mathsf{Q}$ if $\mathsf{P} - \mathsf{Q} \succ 0$. Show that $\text{tr}(\mathsf{P}) > \text{tr}(\mathsf{Q})$. [*Hint: use the SVD of* $\mathsf{P} - \mathsf{Q}$*, linearity and the cyclic property of the trace.*]

Exercise 4. Let P denote an arbitrary $n \times p$ matrix.

1. Show that $\mathsf{A} = \mathsf{P}^\top \mathsf{P}$ is symmetric and nonnegative definite.
2. Show that every eigenvalue of A must be nonnegative

Denote these eigenvalues by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. with corresponding eigenvectors $u_1, u_2, \ldots, u_p$.

3. Show that the matrix $\sum_{i=1}^r (1/\lambda_i)\, u_i u_i^T$ is a generalized inverse of $\mathsf{P}^\top \mathsf{P}$, where r is the number of non-zero eigenvalues of A.

Exercise 5. The Orthogonal Procrustes problem:

$$\min_{\mathsf{X} : \mathsf{X}\mathsf{X}^\top = \mathsf{I}} \|\mathsf{A}\mathsf{X} - \mathsf{B}\|_F,$$

can geometrically be seen as seeking a transformation of points (contained in A) to other points (contained in B) that involves only rotation. Show that the solution of the Procrustes problem above can be found via the SVD of the matrix $\mathsf{A}^\top \mathsf{B}$.

Exercise 6. `[Implementation]` The Power Iteration method is an iterative algorithm used to find the dominant eigenvector and eigenvalue of a square matrix A. It consists in iteratively applying the matrix to a random vector and normalizing the result to converge towards the dominant eigenvector.

To implement the Power Iteration algorithm, start with a random vector v_0 of appropriate size and iterate the following until convergence:

$$v_{i+1} = \frac{\mathsf{A}v_i}{\|\mathsf{A}v_i\|}$$

At convergence, after say N iterations, estimate the dominant eigenvalue λ_1 by $\hat{\lambda}_1 = \frac{v_N^\top \mathsf{A} v_N}{v_N^\top v_N}$.

Test your implementation with a few sample matrices of different sizes. Compare the computed dominant eigenvector and eigenvalue with the ones obtained using the built-in function for eigenvalue decomposition. Discuss the convergence behavior and accuracy of your implementation. In particular, verify that the speed of convergence is given by the ratio λ_2/λ_1.

Exercise 7. Recall the definition of the Kroneker matrix product:

$$\mathsf{P} \otimes \mathsf{Q} = \begin{pmatrix} p_{11}\mathsf{Q} & \dots & p_{1p}\mathsf{Q} \\ \vdots & \ddots & \vdots \\ p_{n1}\mathsf{Q} & \dots & p_{np}\mathsf{Q} \end{pmatrix}.$$

Show that:

1. the Kroneker product is bilinear and associative, but not commutative
2. $(\mathsf{P} \otimes \mathsf{Q})(\mathsf{X} \otimes \mathsf{Y}) = (\mathsf{PX}) \otimes (\mathsf{QY})$.
3. $(\mathsf{P} \otimes \mathsf{Q})^{-1} = (\mathsf{P}^{-1} \otimes \mathsf{Q}^{-1})$.
4. (*optional*) $\mathrm{vec}(\mathsf{PXQ}) = (\mathsf{Q}^\top \otimes \mathsf{P})\mathrm{vec}(\mathsf{X})$.

If we further assume that P, Q are square matrices, then also show that:

5. $\mathrm{tr}(\mathsf{P} \otimes \mathsf{Q}) = \mathrm{tr}(\mathsf{P})\mathrm{tr}(\mathsf{Q})$

Exercise 8. ($\star$) Let $\mathsf{P}, \mathsf{Q} \in \mathbb{R}^{p \times p}$ be symmetric and non-negative definite matrices, i.e. $\mathsf{P}, \mathsf{Q} \succeq 0$. Show that:

1. $\mathsf{Q} - \mathsf{P} \succeq 0 \Rightarrow \mathrm{Range}(\mathsf{P}) \subset \mathrm{Range}(\mathsf{Q})$
2. $\mathrm{Range}(\mathsf{P}) \subset \mathrm{Range}(\mathsf{Q}) \Rightarrow c\mathsf{Q} - \mathsf{P} \succeq 0$ for some $c > 0$.

Exercise 9. Let $\mathsf{Q} \in \mathbb{R}^{p \times p}$. Show that $\mathsf{Q} \succeq 0$ if and only if there exist vectors $q_1, \dots, q_m \in \mathbb{R}^p$ with $m \leq p$ such that $\mathsf{Q} = \sum_{i=1}^m q_i q_i^\top$.

Exercise 10. Let $\mathsf{Q} \in \mathbb{R}^{p \times p}$ be symmetric. Show that the following are equivalent:

1. $x^\top \mathsf{Q} x \geq 0$, $\forall\, x \in \mathbb{R}^p$
2. if $\mathsf{Q}v = \lambda v$ for some $v \in \mathbb{R}^p$ and $\lambda \in \mathbb{R}$, then $\lambda \geq 0$.

Exercise 11. ($\star$) Let $v_1, \dots, v_m \in \mathbb{R}^p$ and $\mathsf{Q} \in \mathbb{R}^{p \times p}$. Show that:

$$v_1, ..., v_m \in \mathrm{Range}(\mathsf{Q}) \;\Leftrightarrow\; \mathrm{Range}\left(\sum_{i=1}^m v_i v_i^\top\right) \subset \mathrm{Range}(\mathsf{Q})$$

Exercise 12. Consider a bivariate Pareto density:

$$f(x, y) = c(x + y - 1)^{-p-2}, \text{ for } x, y > 1, \text{ and } p > 2.$$

1. Show that c is equal to $p(p+1)$.
2. Determine the marginal laws of this density and compute $\mathbb{E}X$.

3. Calculate the variance-covariance matrix Σ.

4. Consider a sample $(X_1, Y_1), \dots, (X_n, Y_n)$ of independent and identically distributed random vectors following the Pareto density with parameters p. Estimate the parameter p using the maximum likelihood method.

Exercise 13. 1. Let X be a random vector in $\mathbb{R}^p$ such that $\mathbb{E}\|X\|^2 < \infty$ and with covariance Σ. Given $\mathsf{A} \in \mathbb{R}^{p\times d}$ a real matrix, show that the covariance matrix of $\mathsf{A}X$ is $\mathsf{A}\Sigma\mathsf{A}^\top$.

2. Let Σ be a real symmetric matrix. Then Σ is non-negative definite if and only if Σ is the covariance matrix of some random variable X.

Exercise 14. Suppose that $X_1, \dots, X_n$ are independent and identically distributed p-dimensional random vectors following a multivariate Gaussian distribution $\mathcal{N}(\mu, \Sigma)$. Consider the sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the sample covariance matrix

$$S = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^\top$$

1. Show that $E\left[\bar{X}\right] = \mu$.
2. Show that $E\left[S\right] = \frac{n-1}{n}\Sigma$.

Exercise 15. Let $(X^\top, Y^\top)^\top$ be a jointly Gaussian, comprised of concatenated random vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, with mean and covariance

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{XY}^\top & \Sigma_Y \end{pmatrix}.$$

Derive the conditional density $X \mid Y = y$.
Hint: It might be easier to consider the precision matrix when writing the Gaussian densities above.

Exercise 16. We consider $X \sim N_p(\mu_X, \Sigma_X)$ and $Y \mid X = x \sim N_q(\alpha + \beta x, \Sigma)$ where $\mu \in \mathbb{R}^p$, $\Sigma_X \in R^{p\times p}$, $\alpha \in \mathbb{R}^q$, $\beta \in \mathbb{R}^{q\times p}$ and $\Sigma \in \mathbb{R}^{q\times q}$.

1. Prove that $(X^\top, Y^\top)^\top \sim \mathcal{N}$, compute its mean and show that

$$\operatorname{Var}\left(X^\top, Y^\top\right)^\top = \begin{pmatrix} \Sigma_X & \Sigma_X\beta^\top \\ \beta\Sigma_X & \Sigma + \beta\Sigma_X\beta^\top \end{pmatrix}$$

Hint: Start by setting $U = Y - \alpha - \beta X$ and showing that X and U are independent, then find a matrix A and a vector c such that $(X^\top, Y^\top)^\top = \mathsf{A}(X^\top, U^\top)^\top + c$.

2. Show that the conditional distribution of $X \mid Y = y$ is Gaussian with

$$\mathbb{E}[X \mid Y = y] = \mu_X + \beta\Sigma_X(\Sigma + \beta\Sigma_X\beta^\top)^{-1}(y - \alpha - \mu_X\beta),$$
$$\operatorname{Var}[X \mid Y = y] = \Sigma_X - \Sigma_X\beta^\top(\Sigma + \beta\Sigma_X\beta^\top)^{-1}\beta\Sigma_X,$$

assuming that the matrices $\Sigma, \Sigma_X, (\Sigma + \beta\Sigma_X\beta^\top)$ are invertible.

Exercise 17. `[Implementation]` Generate synthetic 2D dataset drawing 100 samples from a Gaussian distribution with mean zero and covariance:

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

for ρ zero, positive and negative. Visualise scatterplots of your data. Verify by means of histogram plotting that any 1-d projection of your data follows approximately a Gaussian distribution. You may also perform a statistical tests verifying the normality of the projected data, as the Shapiro-Wilk test.

Exercise 18. Let X, Y be two identical, independent Gaussian random vectors in $\mathbb{R}^p$ with mean m and covariance Σ.

1. Find the distribution of $Z_1 = X - Y$ and $Z_2 = X + Y$.
2. Show that Z_1 and Z_2 are independent.

[*Hint: use the MGF*]

Exercise 19. Let $Z \sim N(0, \mathrm{I}_{k\times k})$.

1. Show that $\mathbb{E}\left[Z^\top Z\right] = k$ and $\operatorname{var}\left\{Z^\top Z\right\} = 2k$.
2. Show that the moment generating function of $Z^\top Z$ is given by

$$M_{Z^\top Z}(t) = (1-2t)^{-k/2}, \quad t < \frac{1}{2}.$$

Exercise 20.

1. $X \sim N(0, \Sigma_{p\times p})$ and Σ invertible, then $X^\top \Sigma^{-1} X \sim \chi^2_p$.
2. $X \sim N(0, \mathrm{I}_p)$ and H a projection, then $X^\top \mathrm{H} X \sim \chi^2_{\mathrm{rank(H)}}$.
3. $X \sim N(0, \Sigma_{p\times p})$, then $X^\top \Sigma^\dagger X \sim \chi^2_{\mathrm{rank}(\Sigma)}$.

Exercise 21. `[Implementation]` Generate N iid copies $X_1, \ldots, X_n$ of Gaussian random variables in $\mathbb{R}^p$ with mean μ_x and (invertible) covariance matrix Σ_X. Let $A \in \mathbb{R}^{p\times p}$ and $b \in \mathbb{R}^p$. Then generate a sequence of N random variables in $\mathbb{R}^p$ by: $Y_i = AX_i + b + Z_i$ for $i = 1, \ldots, N$, where $Z_i \sim \mathcal{N}(0, \epsilon \cdot I_d)$ is i.i.d. noise for some small $\epsilon > 0$. Find the best linear predictor of Y given each X, and validate your results. How does the best linear predictor compare to the latent true linear model?

Exercise 22. `[Implementation]` The Gaussian concentration of measure phenomenon roughly states that in "high dimensions" (p large), the realisations of $Z \sim N(0, I_{p\times p})$ highly concentrate near the surface of the sphere of radius $\sqrt{p}$. Hence, for p large, $\|(\sqrt{p})^{-1} Z\|$ should concentrate on the value 1. In this exercise we will visualise this asymptotic fact.

Generate 1000 Gaussian random variables $X_1, \ldots, X_{1000} \in \mathbb{R}^p$ for $p = 5, 20, 50, 100, 500, 1000$, with mean 0 and covariance the identity matrix. In each case, find the scaled mean of their norm:

$$(\sqrt{d})^{-1} \frac{1}{1000} \sum_{i=1}^{1000} \|X_i\|.$$

Replicate this experiment 100 times, and plot the histograms of the frequencies (or the kde) obtained for each dimension.

Exercise 23. ($\star$) Let W be a random vector. Show that the following are equivalent:

1. $\mathsf{U}W \overset{d}{=} W$ for all orthogonal U
2. $W = \xi U$ where $U \sim \mathrm{Unif}\{x : \|x\| = 1\}, \xi > 0$ is a random scalar and $U \perp \xi$
3. $v^\top W \overset{d}{=} \|v\| W_1$, for all v (where $W = (W_1, \ldots, W_p)^\top$)

Exercise 24. Recall that a random vector X is called elliptical with location μ and dispersion $\mathsf{A}\mathsf{A}^\top = \Sigma$ if it is representable as $X \overset{d}{=} \mu + \mathsf{A}W$, where W is spherical. Show that elliptical distributions vectors are closed under marginalisation and affine transformations.

Exercise 25. If X is a Gaussian $n \times p$ data matrix from $N(\mu, \Sigma)$, then $\mathsf{A}_{m\times n}\mathsf{X}\mathsf{B}_{p\times q}$ is an $m \times q$ Gaussian data matrix if and only if the following two conditions hold true:

1. $\mathsf{A}1_n = \alpha 1_m$ for some $\alpha \in \mathbb{R}$ OR $\mathsf{B}^\top \mu = 0$.

2. $\mathsf{AA}^\top = \beta \mathsf{I}_{m\times m}$ for some $\beta \in \mathbb{R}$ OR $\mathsf{B}^\top \Sigma \mathsf{B} = 0$.

Clearly, when AXB is a Gaussian data matrix, it is from a $N\left(\alpha \mathsf{B}^\top \mu, \beta \mathsf{B}^\top \Sigma \mathsf{B}\right)$.

Exercise 26. Let X be a Gaussian data matrix from $N(\mu, \Sigma)$. Then

$$\mathsf{AXB} \perp \mathsf{CXD} \Longleftrightarrow \mathsf{AC}^\top = 0 \text{ or } \mathsf{B}^\top \Sigma \mathsf{B} = 0$$

Hint: Consider vec(X) the np vector obtained by stacking the columns of X on top of one another:

$$\text{vec}(\mathsf{X}) \sim N(\mu, \Sigma \otimes I) \tag{1}$$

where $\otimes$ denotes the Kronecker multiplication, and use the fact that

$$\text{vec}(\mathsf{AXB}) = (\mathsf{B}^\top \otimes \mathsf{A})\text{vec}(\mathsf{X}). \tag{2}$$

Exercise 27. Show the following:

1. $\mathsf{W} \sim W_p\left(\Sigma_{p\times p}, n\right) \Longleftrightarrow \mathsf{W} \overset{d}{=} \sum_{i=1}^n \mathsf{W}_i, \quad \mathsf{W}_i \overset{iid}{\sim} W_p\left(\Sigma_{p\times p}, 1\right).$
2. $\mathsf{W} \sim W_p\left(\Sigma_{p\times p}, n\right) \Longrightarrow \mathbb{E}[\mathsf{W}] = n\Sigma.$
3. The lower triangular part of $\mathsf{W} \sim W_p\left(\Sigma_{p\times p}, n\right)$ has density if and only if $n \geq p$.
4. $\mathsf{W} \sim W_p\left(\Sigma_{p\times p}, n\right) \Longrightarrow \theta^\top \mathsf{W}\theta / \theta^\top \Sigma\theta \sim \chi^2_n, \forall \theta \notin \ker(\Sigma)$

Exercise 28.

1. If $X \sim N\left(\mu, \Sigma_{p\times p}\right)$ independently of $\mathsf{W} \sim W_p(\Sigma, n)$ with Σ non-singular and $n \geq p$, then
$$n(X-\mu)^\top \mathsf{W}^{-1}(X-\mu)^\top \sim T^2(p, n).$$
2. Let $\bar{X}$ and $\widehat{\Sigma}$ be the sample mean and covariance of a $N(\mu, \Sigma)$ iid sample. If $n \geq p$ and Σ is non-singular, then
$$(n-1)(\bar{X}-\mu)^\top \widehat{\Sigma}^{-1}(\bar{X}-\mu) \sim T^2(p, n-1)$$

Exercise 29. Show by counterexample that separate weak convergence of each coordinate does not imply weak convergence of the random vector.

Exercise 30. Let $\mathsf{W} \sim W(\Sigma, 1)$. Recall that $\text{cov}\left\{w_{ij}, w_{kl}\right\} = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}$. Show that:

1. Σ diagonal $\Longrightarrow$ Wishart entries uncorrelated w/ variances $\sigma_{ii}\sigma_{jj}$.
2. $\Sigma \succ 0 \Longrightarrow \text{cov}\left[\left\{w_{ij}\right\}_{i\leq j}\right] \succ 0$ (lower triangular part of W).

Exercise 31. `[Implementation]` In this exercise we use hotellisation to construct confidence intervals for a multivariate Gaussian. Let $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(\mu, \Sigma)$ in $\mathbb{R}^d$. We know that $n(\overline{X}-\mu)\widehat{\Sigma}^{-1}(\overline{X}-\mu)$ is distributed as Hotelling's T^2 with parameters $d, n-1$. An ellipsoidal confidence set with coverage probability $1-\alpha$ consists of:

$$\{m \,:\, n(\overline{X}-m)^\top \widehat{\Sigma}^{-1}(\overline{X}-m) < T^2_{d,n-1}(1-\alpha)\}$$

Generate and visualise the confidence ellipsoids for the mean on the basis of $n = 100$ Gaussian samples in $\mathbb{R}^2$, i.e. $d = 2$. *Note: this may require a little bit of thinking...*

Exercise 32. If $X_1, \ldots, X_n \overset{\text{IID}}{\sim} N(\mu, \Sigma)$ with $n > p$ and $\Sigma \succ 0$, then the unique MLE of μ is $\bar{X}$. Prove it is also the minimum variance unbiased estimators of μ. [*Hint: use projections.*]

Exercise 33. Show that the MLE is parametrisation equivariant. That is, for any transformation g, $g(\hat{\theta})$ is MLE of $g(\theta)$ where $\hat{\theta}$ is MLE of θ. Furthermore, show that if g is 1-1, then uniqueness is also inherited when present.

Exercise 34. `[Implementation]` Consider testing for the difference in the mean of two gaussian samples with known, identical covariance matrix. Using Monte Carlo, verify the confidence level of the hypothesis test constructed.

Exercise 35. `[Implementation]` Given a set of N bivariate sample pairs $(X_i, Y_i), i = 1, \ldots, N$, the sample correlation coefficient r is given by

$$r = \frac{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^{N} \left(Y_i - \bar{Y}\right)^2}}.$$

Fisher's z-transformation of r is defined as $z = \operatorname{artanh}(r)$. If (X, Y) has a bivariate normal distribution with correlation ρ and the pairs (X_i, Y_i) are independent and identically distributed, then z is approximately normally distributed with mean $\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$, and standard deviation $\frac{1}{\sqrt{N-3}}$. Use this transformation, and its inverse $r = \tanh(z)$ to construct a large-sample confidence interval for r.

Exercise 36. ($\star$) Establish the matrix determinant and Sherman-Morrison formulas:

1. If $\Sigma \succ 0$, then $\left|\Sigma + uu^\top\right| = |\Sigma|\left(1 + u^\top \Sigma^{-1} u\right)$
2. If $\Sigma \succ 0$, then $\left(\Sigma + uu^\top\right)^{-1} = \Sigma^{-1} - \frac{1}{1+u^\top\Sigma^{-1}u}\Sigma^{-1}uu^\top\Sigma^{-1}$

Exercise 37. In the context of partial correlation tests (slides 162-163) verify that $\rho_{\epsilon_X,\epsilon_Y}$ is the sample correlation between the residuals obtained when regressing X on Z and those when regressing Y on Z. Use this to establish the distribution under the null.

Exercise 38. Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \Sigma_{p\times p})$ with $\Sigma \succ 0$ and $n > p$. Consider the hypothesis pair,

$$\begin{cases} H_0: & \Sigma = \lambda I \text{ for some } \lambda > 0, \\ H_1: & \Sigma \neq \lambda I \text{ for all } \lambda > 0. \end{cases}$$

Show that the LRT rejects H_0 for large values of

$$(\gamma(\hat{\Sigma})/\alpha(\hat{\Sigma}))^n$$

$\alpha(\hat{\Sigma})$ and $\gamma(\hat{\Sigma})$ are the arithmetic and geometric means of the eigenvalues of $\hat{\Sigma}$.

Exercise 39. `[Implementation +` *theory* `]` Generate N IID gaussian vectors in $\mathbb{R}^p$, i.e. $X_1, \ldots, X_N \overset{IID}{\sim} \mathcal{N}(0, \Sigma)$ and express them in the coordinate system given by the (orthonormal) eigenvectors of Σ. Verify that the variance explained by each coordinate is *decreasing*. Explain this by the Karhunen-Loeve theorem *and* by the Optimal Linear Dimension Reduction Theorem.

Exercise 40. Show how to perform PCA (i.e. compute the principal component scores) of a centered sample $(X_1, \ldots, X_n)$ in terms of the SVD of the corresponding data matrix.

Exercise 41. `[Implementation]` Choose one of the following two datasets and analyses.

1. We consider the Wisconsin breast cancer dataset, which consists of 596 samples and 30 features. The features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image and are linked with a label, malignant or benign. To import the dataset, import the `sklearn` module, and run `sklearn.datasets.load_breast_cancer()`. Reduce the dimensionality of the data, and visualise the effect of the first two principal components. Visualise the first two principal component scores, comparing with the labels (i.e. malignant or benign). Use your favourite classifier (for instance nearest centroid) to assess the value of the first two principal component scores in predicting the label of the cell nuclei, and validate your results using cross-validation.

2. We consider the digits dataset, containing handwritten digits from 0 to 9. The dataset is comprised of 1797 samples of 64 features. We would like to group images such that the handwritten digits on the image are the same. To import the dataset, import the `sklearn` module, and run `sklearn.datasets.load_digits()`. Reduce the dimensionality of the data, and visualise the effect of the first two principal components. Visualise the first two principal component scores, coloring the scatterplot according to the label (i.e. the digit). Use your favourite classifier (for instance nearest centroid) to assess the value of the first two principal component scores in predicting correct digit, and validate your results using cross-validation.

Exercise 42. Prove the eigenvalue perturbation bound:

$$\max_j \left|\hat{\lambda}_i - \lambda_j\right| \leq \|\hat{\Sigma} - \Sigma\|_{\mathbb{R}^{p\times p}}$$

Exercise 43. Let $X_1, \ldots, X_n$ be iid random vectors in $\mathbb{R}^p$. Show that the best approximating k-hyperplane to the points $\{X_1, \ldots, X_n\}$ is given by $\bar{X} + \mathcal{R}\left(\hat{\Sigma}_k\right)$, where $\sum_k = \sum_{i=1}^k \hat{\lambda}_i \hat{u}_i \hat{u}_i^\top$ is the rank-k spectral truncation of $\hat{\Sigma}$.

Exercise 44. ($\star$) Let $X_1, \ldots, X_n$ be iid Gaussian vectors, with covariance Σ and empirical covariance $\hat{\Sigma}$ Then it is known that $\sqrt{n}(\hat{\Sigma} - \Sigma) \xrightarrow{d} Z$ for some mean-zero gaussian matrix Z. Define

$$W_{ij} = \mathbb{E}\left[\langle Zu_i, u_i\rangle \langle Zu_j, u_j\rangle\right]$$

where $\{(\lambda_i, u_i)\}_{i=1}^p$ is the spectrum of Σ. Show that when $X_1, \ldots, X_n$ are Gaussian, W is diagonal with $W_{ii} = 2\lambda_i^2$. *Hint: derive the covariance of* $\mathrm{vec}(Z)$ *using slide 129 and the Gaussian sampling theorem. The use properties of coviariance matrices, inner products, vectorisation and kronecker products ...*

Exercise 45. [`Implementation`] Generate N samples from a lower dimensional signal + isotropic noise model (as in slide 201). Select the number of components for performing PCA by percentage of explained variance, by setting a bound on the conditioning number and by multiple testing of sphericity.

Exercise 46. Verify the (two) missing steps in the proof of the theorem on Asymptotic Law of Wishart Spectrum, in slide 193.

Exercise 47. Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ be random vectors with strictly positive covariances Σ_X and Σ_Y respectively, with absolutely continuous law. Show that $\mathbb{P}(\hat{\Sigma}_X \succ 0, \hat{\Sigma}_Y \succ 0) = 1$ for all n sufficiently large.

Exercise 48. Show that model selection in the low rank plus noise model takes the form $\hat{\lambda}_k >$ threshold(n) which resemble a (sample-size dependent) condition number criterion.

Exercise 49. In the framework of slide 211, show that when $p = 1$ the only non-trivial canonical correlation vector is the standardised least square estimator of the regression coefficient vector.

Exercise 50. Use the spectrum to show that $\Sigma \mapsto \Sigma^2$, $\Sigma \mapsto \Sigma^{1/2}$ and $\Sigma \mapsto \Sigma^{-1}$ are C^1 at $\Sigma \succ 0$.

Exercise 51. Consider a stationary Markov Chain with jointly gaussian, independent increments $\xi_1, \xi_2, \ldots$ (see slide 218). Show that there exist constants σ, ρ and a sequence of random variables $\epsilon_1, \epsilon_2, \ldots$ such that:

1. $\xi_1 \sim \mathcal{N}(0, \sigma^2/(1-\rho^2))$
2. $\xi_{k+1} = \rho\xi_k + \epsilon_k$
3. $\rho = \mathrm{corr}\{\xi_k, \xi_{k+1}\} < 1$
4. $\xi_k \perp \epsilon_k$

5. $\epsilon_k \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$

Hint: use the regression representation of conditional independence.

Exercise 52. Compute the log-likelihood of a stationary Gaussian AR(1) based on a single realisation via the Markov factorisation.

Exercise 53. `[Implementation]` Implement a gaussian AR(1) model and estimate its parameters (σ, ξ) on the basis of a sample path of $N = 100$ steps. Visualise the path, and forecast by Monte Carlo (MC) the following 5 steps.