Group discussion 1: Covariance, Correlation and Multivariate Normal Distribution

Ex. 1: Discuss the difference between the following terms: Variance, Covariance, Correlation.
Variance: The “average” difference in the dataset/ Describes the ‘spread’ of a datasets values.
Covariance: Variance between multiple variables. The influence variables have on other variables…
Correlation: Simply a normalized version of the covariance to the interval [-1,1]

Ex. 2: Prove the following formulas using the definition of Covariance and Correlation

1. $Cov[X, X] = Var[X]$

   $Var[X] = E[(X - E[X])^2]$

   $Cov[X, X] = E[(X - E[X])(X - E[X])] = E[(X - E[X])^2] = Var[X]$

2. $Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$

   $Var[X + Y] = E[((X + Y) - E[(X + Y)])^2]$

   $= E[(X + Y)^2] - E[(X + Y)]^2$

   $= E[(X + Y)^2] - (E[X] + E[Y])^2$

   $= E[X^2 + Y^2 + 2XY] - E[X]^2 - E[Y]^2 - 2E[X]E[Y]$

   $= E[X^2] + E[Y^2] + E[2XY] - E[X]^2 - E[Y]^2 - 2E[X]E[Y]$

   $= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 + E[2XY] - 2E[X]E[Y]$

   $= Var[X] + Var[Y] + 2(E[XY] - E[X]E[Y])$

   $= Var[X] + Var[Y] + 2Cov[X, Y]$

3. $Cov[aX + bY, Z] = aCov[X, Z] + bCov[Y, Z]$

   $Cov[aX + bY, Z] = E[((aX + bY) - E[aX + bY])(Z - E[Z])]$

   $= E[((aX + bY) - (aE[X] + bE[Y]))(Z - E[Z])]$

$= E[(aX + bY)Z - (aX + bY)E[Z] - (aE[X] + bE[Y])Z + (aE[X] + bE[Y])E[Z]]$

   $= E[aXZ + bYZ - aXE[Z] - bYE[Z] - aE[X]Z - bE[Y]Z + aE[X]E[Z] + bE[Y]E[Z]]$

   $= E[a(XZ - XE[Z] - E[X]Z + E[X]E[Z]) + b(YZ - YE[Z] - E[Y]Z + E[Y]E[Z])]$

   $= aE[XZ - XE[Z] - E[X]Z + E[X]E[Z]] + bE[YZ - YE[Z] - E[Y]Z + E[Y]E[Z]]$

   $= aE[(X - E[X])(Z - E[Z])] + bE[(Y - E[Z])(Z - E[Z])]$

   $= aCov[X, Z] + bCov[Y, Z]$

Ex. 3: Discuss the difference between the Univariate and the Multivariate Distribution. (discuss in the two cases: the variables, mean, covariance)

Multivariate distribution has multiple random variables.
The mean goes from regular mean μ to a vector μ with size D x 1.
Uni distribution is defined by variance (the spread) while multi. distribution is defined by covariance (a relationship between variables).

Ex. 4: Let X be a 2-dimensional multivariate normal random variable with zero means, i.e. $\mu = \mu_1, \mu_2)^T = (0, 0)^T$ and a symmetric 2 × 2 covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

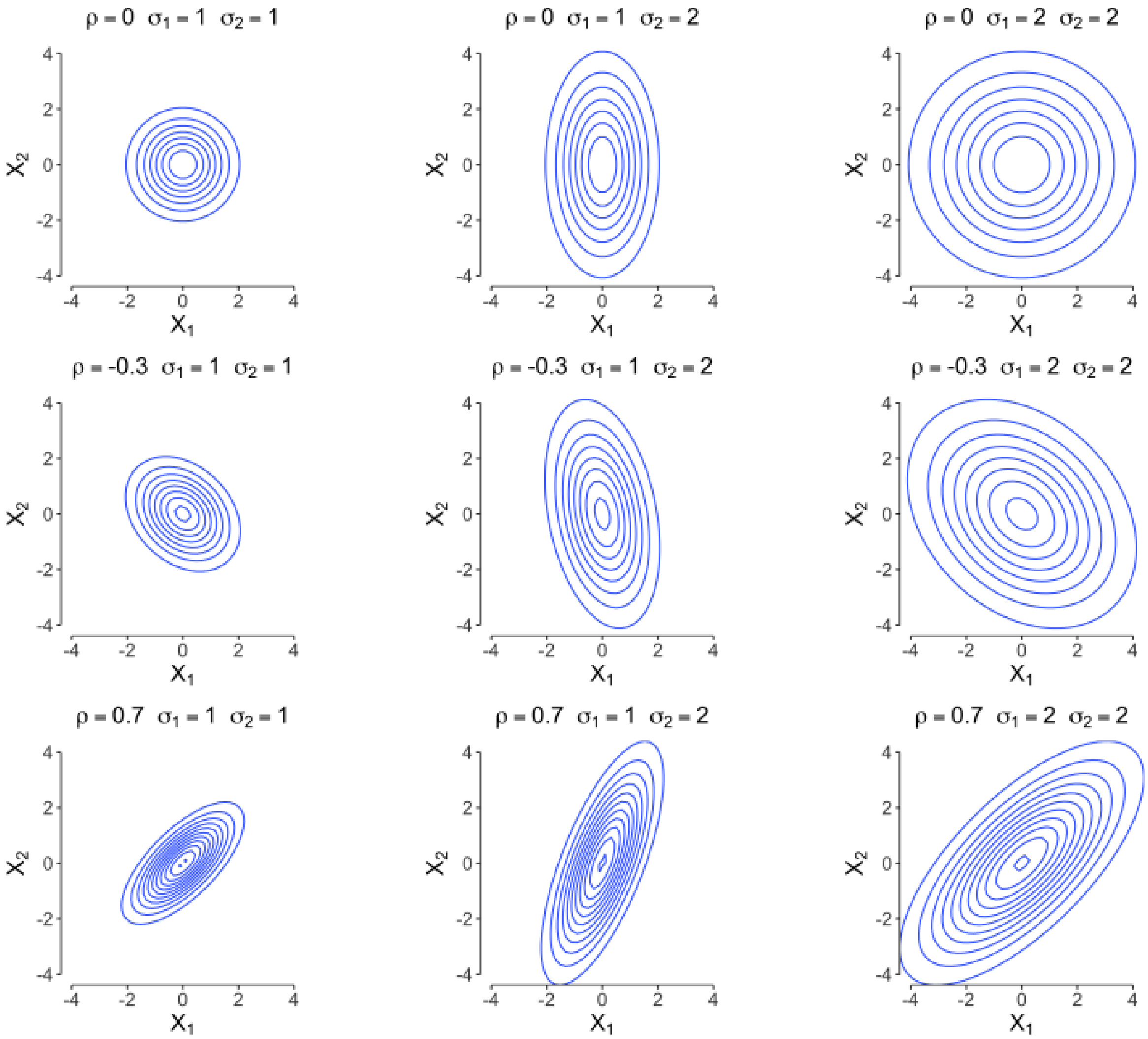
where σ_1^2 and σ_2^2 are the population variances of the random variables X_1 and X_2, respectively, and ρ is the correlation between the two.

1. Show that this covariance matrix is equivalent to the covariance matrix we define in the lecture.

$$\rho_{(x_1, x_2)} = \frac{cov(x_1, x_2)}{\sigma_1\sigma_2}$$

$$cov(x_1, x_2) = \rho_{(x_1, x_2)}\sigma_1\sigma_2$$

2. Check the following plots and discuss the different cases of the variance and correlation and how it affects the plot.


ρ = 0 σ1 = 1 σ2 = 1
ρ = 0 σ1 = 1 σ2 = 2
ρ = 0 σ1 = 2 σ2 = 2
ρ = -0.3 σ1 = 1 σ2 = 1
ρ = -0.3 σ1 = 1 σ2 = 2
ρ = -0.3 σ1 = 2 σ2 = 2
ρ = 0.7 σ1 = 1 σ2 = 1
ρ = 0.7 σ1 = 1 σ2 = 2
ρ = 0.7 σ1 = 2 σ2 = 2
X1
X2

Group discussion 2: Principle component Analysis (PCA)

Ex. 5: Take turns in the group describing very very broadly what PCA is, the idea behind it and why we would want to have such method. We strongly encourage all in the group to formulate this out loud in their own words.

Looking for what "features" are "redundant" or less relevant. PCA looks for the direction in a dataset that has most information, also happens to be the direction with greatest variance.

Ex. 6: The first step in the derivation of PCA is introducing a single unit vector $\boldsymbol{u}$. If we were to draw every possible unit vector in a 2D space, what shape would appear?
Circle with radius 1

What about in 3D?
Sphere with radius 1

Ex. 7: We arrived at the expression

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

Discuss why the largest eigenvalue corresponds to the direction with the largest projected variance.
Hint: look at the definition of the projected variance - can you rearrange the above equation to make it look like the projected variance?

$\Sigma u \ = \ \lambda u$

$u^T \Sigma u \ = \ \lambda u^T u$

$\sigma^2 = \lambda$

Ex. 8: We've come to the conclusion that after having constructed the covariance matrix $\mathbf{\Sigma}$, we could simply find the eigenvectors $u_1, u_2, \ldots, u_M$ and corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_i \geq \ldots \geq \lambda_M$ λ2 ≥ ...λi ≥ ... ≥ λ_M. Discuss the following questions:

•What is M? I.e. how many eigenvectors and eigenvalues will we get?
One for each feature of our data e.i. 100 if we get 10 by 10 pictures.

•What does it mean if one of the eigenvalues $\lambda_i = 0$? Imagine a 2D-example where $\lambda_1 > 0$ and $\lambda_2 = 0$. What would be characteristic about the data $\mathbf{X}$?
If an eigenvalue is 0 the corresponding dimension in the data will always be the same (variance is equal to zero). The data will therefore only be 1D because none of the variance in the data is explained by the 2. dimension.

Ex. 9: (Optional) In the derivation of PCA, we use the following equality:

$$\frac{1}{N-1}\sum_{j=1}^{N}(\alpha_j - \mu_\alpha)^2 = \mathbf{u}^\top \boldsymbol{\Sigma}\mathbf{u}$$

Prove this equality. Hints: insert the definitions of $\alpha_j = u^T x_j$ and $\mu_\alpha = u^T \overline{x}$ and remember that

$$\boldsymbol{\Sigma} = \frac{1}{N-1}\sum_{j=1}^{N}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^\top$$

You will at some point also need $a^T b = b^T a$