Handed out: April 6 (Tue)

Hints for solution: Exercise class on April 8 (Thu)

Hand in: April 14 (Wed), the latest, @ Room A348

This assignment gives you maximally 5% worth of extra points for the computer assignments and the final exam. Note: Not all exercises have equal weight!

Ex. 1 — PCA and data representation (1 out of 5 %)
Denote by $U = (\mathbf{u}_1, \ldots \mathbf{u}_m)$ the first $m \leq p$ principal component directions (weights) of the zero mean random variable $\mathbf{x} \in \mathbb{R}^p$. Assume you have n observations of $\mathbf{x}$, organized in the matrix X

$$X = (\mathbf{x}_1 \ldots \mathbf{x}_n). \tag{1}$$

1. What are the elements of the rows $\mathbf{v}_i^T$ of X?
2. Express the sample covariance matrix in terms of $\mathbf{v}_i$.
3. Let $\mathbf{z} = U^T\mathbf{x}$ and $Z = U^T X$. Write down an explicit expression for the rows of Z, and give an interpretation of the rows in terms of PCA.
4. Show that the rows of Z are orthogonal to each other.
5. Interpret orthogonality of the rows of Z.

Ex. 2 — Correlations, linear dependence, and small eigenvalues (1 out of 5 %)

1. Assume the covariance matrix C of $X = (x_1, x_2)^T$ has the form

$$C = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{2}$$

   Calculate the eigenvalues in function of ρ (by hand). What is the effect of correlation between the random variables on the eigenvalues?
2. Let $x_2 = ax_1 + n$ where n is uncorrelated with x_1, and x_1 has mean zero and variance 1. How do you have to choose the factor a and the noise n so that X has covariance matrix C ?
3. Calculate the variance of n (the noise variance) and make a scatter plot of X for $\rho = (-1, -0.25, 0, 0.5, 1)$ (either sketch by hand or make the plots with matlab/R)
4. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be the vector with the observations of X_1 and X_2, respectively. What happens to $\mathbf{v}_1$ and $\mathbf{v}_2$ as $|\rho|$ tends to one, and what happens to the conditioning number of C ?

Ex. 3 — Correlation and Projection (1 out of 5 %)
Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ be the three orthogonal, unit norm eigenvectors of the covariance matrix C, given by

$$C = \begin{pmatrix} 1 & 0 & \cos(\alpha) \\ 0 & 1 & \sin(\alpha) \\ \cos(\alpha) & \sin(\alpha) & 1 \end{pmatrix} + \frac{1}{2}\lambda_3 \begin{pmatrix} \cos^2(\alpha) & \cos(\alpha)\sin(\alpha) & -\cos(\alpha) \\ \cos(\alpha)\sin(\alpha) & \sin^2(\alpha) & -\sin(\alpha) \\ -\cos(\alpha) & -\sin(\alpha) & 1 \end{pmatrix}$$

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \\ 1 \end{pmatrix} \mathbf{u}_2 = \begin{pmatrix} -\sin(\alpha) \\ \cos(\alpha) \\ 0 \end{pmatrix} \mathbf{u}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} -\cos(\alpha) \\ -\sin(\alpha) \\ 1 \end{pmatrix}$$

with $\alpha \in [0, 2\pi]$

1. For $\lambda_3 = 0$, show that C is not invertible
2. For $\lambda_3 = 0$, verify that $\mathbf{u}_i$, $i = 1, 2, 3$ are eigenvectors and calculate the eigenvalues λ_1 and λ_2.
3. For arbitrary λ_3, recalculate C from the $\mathbf{u}_i$ and λ_i.
4. For $\lambda_3 = 0.1$, if you want to reduce the dimension of your data by 1, i.e. project your data on a two-dimensional subspace, which two principal components (PCs) would you use?
5. For $\lambda_3 = 0.1$, what is the proportion of the variance explained by your choice of PCs in the previous question?
6. Show where observations of the form $(x_1, 0, 0)^T$, $(0, x_2, 0)^T$ and $(0, 0, x_3)^T$ are projected to when using the first two PCs: Make a sketch for $\alpha = 0, \frac{\pi}{2}, \frac{\pi}{4}$ and $\frac{5\pi}{6}$.
7. How does the projection change in function of the correlation between the variables?

Ex. 4 — PCA and linear regression (2 out of 5 %)
Assume $y_k = \mathbf{x}_k^T \beta + \epsilon_k$, for $k = 1, \ldots, n$ where the ϵ_k are iid Gaussian with mean zero and variance σ^2, and $\mathbf{x}_k \in \mathbb{R}^p$.

1. The observed data is $(y_k, \mathbf{x}_k)$, $k = 1 \ldots n$, and the goal is to estimate β from that data. The argument $\hat{\beta}$ which minimizes the residual sums of squares $J(\beta)$ is an estimate for β,

$$J(\beta) = \frac{1}{n} \sum_{k=1}^{n} (y_k - \mathbf{x}_k^T \beta)^2. \tag{3}$$

Show that $\hat{\beta}$ is

$$\hat{\beta} = (XX^T)^{-1} X \mathbf{y} \tag{4}$$

where $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, and $\mathbf{y} = (y_1, \ldots, y_n)^T$. Express $\hat{\beta}$ also in terms of the sample covariance matrix $\hat{C}_x$ of $\mathbf{x}$ (Hint: $\hat{C}_x = 1/nXX^T$)

2. What does $\hat{\beta}$ become as $n \to \infty$?
3. Show that the expectation and variance of $\hat{\beta}$ given X, i.e. $E(\hat{\beta}|X)$ and $V(\hat{\beta}|X)$ are

$$\begin{aligned} E(\hat{\beta}|X) &= \beta \quad (5) \\ V(\hat{\beta}|X) &= \frac{\sigma^2}{n}\hat{C}_x^{-1}. \quad (6) \end{aligned}$$

4. The mean squared error (MSE) of $\hat{\beta}$ given X is defined as $E(||\beta - \hat{\beta}||^2|X)$. It equals

$$MSE = \mathrm{Tr}\, V(\hat{\beta}|X) + ||\beta - E(\hat{\beta}|X)||^2 \quad (7)$$

   Find an expression for the MSE in terms of the eigenvalues of $\hat{C}_x$. What kind of data leads to a large MSE? (Hint: what kind of data gives small eigenvalues?)
5. We are now introducing PCA regression. Let $U = (\mathbf{u}_1, \dots \mathbf{u}_m)$ contain the first $m \leq p$ principal component directions (weights) that are obtained from X. Assume further that the variances d_i of the principal components satisfy $d_1 \geq d_2 \geq \dots \geq d_m$. In PCA-regression, one looks for β in the form of

$$\beta = U_m \gamma. \quad (8)$$

   Give an expression for the residual sums of squares in Equation (3) in function of γ. Denote this cost function by $J_{pc}(\gamma)$. Show that $J_{pc}(\gamma)$ has the same form as $J(\beta)$ but that the principal components take the place of the inputs $\mathbf{x}_k$.
6. Show that $\hat{\gamma}$, the γ which minimizes $J_{pc}(\gamma)$, is given by

$$\hat{\gamma} = D_m^{-1} U_m^T \frac{1}{n} X\mathbf{y}, \quad (9)$$

   where $D_m = \mathrm{diag}(d_1, \dots, d_m)$.
7. Let $\hat{\beta}_{pc} = U_m \hat{\gamma}$, i.e.

$$\hat{\beta}_{pc} = U_m D_m^{-1} U_m^T \frac{1}{n} X\mathbf{y}, \quad (10)$$

   Show that

$$\begin{aligned} E(\hat{\beta}_{pc}|X) &= U_m U_m^T \beta \quad (11) \\ V(\hat{\beta}_{pc}|X) &= \frac{\sigma^2}{n} U_m D_m^{-1} U_m^T. \quad (12) \end{aligned}$$

8. What is the MSE (defined in Eq (7)) of $\hat{\beta}_{pc}$? Discuss the effect of the parameter m (the number of principal components) on the MSE.