Recitation 6

These problems will cover Markov's inequality, covariance matrices, PCA and convergence.

1. (Random Vector)

   A random vector $\tilde{x}$ with zero mean has a covariance matrix $\Sigma_{\tilde{x}}$ with the following eigendecomposition

$$\Sigma_{\tilde{x}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}. \tag{1}$$

   (a) What is the variance of each of the entries of the random vector $\tilde{x}_1$, $\tilde{x}_2$ and $\tilde{x}_3$?

   (b) Is it possible to find a unit-norm vector $\vec{u}$ such that the inner product between $\tilde{x}$ and $\vec{u}$ (i.e. the amplitude of the projection of $\tilde{x}$ onto that direction) has variance greater than 1?

   (c) Find three constants a_1, a_2 and a_3, such that at least one of them is nonzero and $\mathrm{P}(a_1\tilde{x}_1 + a_2\tilde{x}_2 + a_3\tilde{x}_3 = 0) = 1$. Justify your answer mathematically, and interpret it geometrically.

2. (Not centering) To analyze what happens if we apply PCA without centering, let $\tilde{x}$ be a d-dimensional vector with mean $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma_{\tilde{x}}$ equal to the identity matrix. If we compute the eigendecomposition of the matrix $\mathrm{E}(\tilde{x}\tilde{x}^T)$ what is the value of the largest eigenvalue? What is the direction of the corresponding eigenvector?

3. (Markov's inequality)

   (a) For a discrete random variable $\tilde{x}$ which takes integer values between 1 and n, prove

$$\sum_{x=1}^{n} x p_{\tilde{x}}(x) \geq a \sum_{x \geq a} p_{\tilde{x}}(x). \tag{2}$$

   (b) Use the inequality in Eq. (2) to prove Markov's inequality for random variables that take integer values between 1 and n.

   (c) Suppose we know that the number of items produced in a factory during a week is a random variable with mean 500. What can be said about the probability that this week's production will be at least 1000?

4. (Sample median as an estimator of the median)

   Prove that the sample median is a consistent estimator of the median of an iid sequence of random variables.