Recitation 11. November 26

Focus: random variables, principal component analysis (PCA)

A ***random variable*** is a quantity X that takes values in $\mathbb{R}$. It can be *discrete*, meaning that it takes only countably many possible values x_i each with probability p_i, or *continuous*, in which case it is associated to a probability distribution $p(x)$ (where $p : \mathbb{R} \to \mathbb{R}$).

The ***mean*** (or ***expected value***) $E[X]$ of X is the sum $\sum_i x_i p_i$ if X is discrete and the integral $\int_{-\infty}^{\infty} xp(x)\, dx$ if X is continuous. If Y is another random variable, and $a, b \in \mathbb{R}$, then $E[aX + bY] = aE[X] + bE[Y]$ (so the mean obeys this linearity property). The ***variance*** $\Sigma = \Sigma_{XX}$ of a random variable X is $E[(X-\mu)^2] = E[(X - E[X])^2]$. The ***covariance*** Σ_{XY} of two random variables X and Y is $E[(X - E[X])(Y - E[Y])]$.

Given n random variables $X_1, \ldots, X_n$, we may assemble them into a vector $\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$, called a ***random vector***.

The ***covariance matrix*** of these random variables $X_1, \ldots, X_n$ is the matrix

$$\begin{bmatrix} \Sigma_{X_1X_1} & \cdots & \Sigma_{X_1X_n} \\ \vdots & \ddots & \vdots \\ \Sigma_{X_nX_1} & \cdots & \Sigma_{X_nX_n} \end{bmatrix} = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T],$$

where $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$ is the vector of means.

1. Sample from the numbers 1 to 1000 with equal probabilities 1/1000, and look at the last digit of the sample, squared. This square can end with $X = 0, 1, 4, 5, 6,$ or 9. What are the probabilities p_0, p_1, p_4, p_5, p_6 and p_9? Compute the mean and variance of X.

Solution:

2. Let A, H, and W denote random variables corresponding to the age, height, and weight of dogs at a local shelter, respectively. Suppose the random vector $\begin{bmatrix} A \\ H \\ W \end{bmatrix}$ takes two values, $\begin{bmatrix} 7 \\ 20 \\ 132 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ 24 \\ 120 \end{bmatrix}$ with probabilities p and $1-p$ respectively. Compute the covariance matrix of A, H, and W.

Solution:

3. Suppose now that the random variables A, H, W from above instead have the covariance matrix

$$K = \begin{bmatrix} 3 & -1 & 2 \\ -1 & 3 & -2 \\ 2 & -2 & 6 \end{bmatrix}.$$

Find three linear combinations of A, H, W which are pairwise independent random variables. What is the variance of each?

Solution:

4. Let X be a random variable. Suppose the mean $E[X] = \mu$ and the variance $\Sigma_{XX} = \sigma^2$. Compute $E[X^2]$ in terms of μ and σ.

Solution: