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 $X_1, \dots, X_n$, 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 $X_1, \dots, X_n$ 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: If $n = 10k$, then the last digit of n^2 will be 0. If $n = 10k + 1$ or $n = 10k + 9$, then the last digit of n^2 will be 1. If $n = 10k + 2$ or $n = 10k + 8$, then the last digit of n^2 will be 4. If $n = 10k + 3$ or $n = 10k + 7$, then the last digit of n^2 will be 9. If $n = 10k + 4$ or $n = 10k + 6$, then the last digit of n^2 will be 6. If $n = 10k + 5$, then the last digit of n^2 will be 5. Thus,

$$p_0 = \frac{1}{10}; p_1 = \frac{1}{5}; p_4 = \frac{1}{5}; p_5 = \frac{1}{10}; p_6 = \frac{1}{5}; p_9 = \frac{1}{5}.$$

We therefore see that the mean is

$$0 \cdot \frac{1}{10} + 1 \cdot \frac{1}{5} + 4 \cdot \frac{1}{5} + 5 \cdot \frac{1}{10} + 6 \cdot \frac{1}{5} + 9 \cdot \frac{1}{5} = \frac{9}{2},$$

and the variance

$$E[(X - \frac{9}{2})^2] = \left(0^2 \cdot \frac{1}{10} + 1^2 \cdot \frac{1}{5} + 4^2 \cdot \frac{1}{5} + 5^2 \cdot \frac{1}{10} + 6^2 \cdot \frac{1}{5} + 9^2 \cdot \frac{1}{5}\right) - \left(\frac{9}{2}\right)^2 = \frac{293}{10} - \frac{81}{4} = \frac{181}{20}.$$

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: Let $\mu_A = 7p + 4(1-p) = 3p + 4$, $\mu_H = 20p + 24(1-p) = 24 - 4p$, and $\mu_W = 132p + 120(1-p) = 12p + 120$, the means of the random variables. Then,

$$\Sigma_{AA} = E[(A - \mu_A)^2] = (49p + 16(1-p)) - (3p+4)^2 = (33p+16) - (9p^2 + 24p + 16) = -9p^2 + 9p = 9p(1-p)$$

Similarly,

$$\Sigma_{HH} = (400p + 576(1-p)) - (24 - 4p)^2 = (-176p + 576) - (576 - 192p + 16p^2) = 16p(1-p)$$

and

$$\Sigma_{WW} = (120 - 132)^2 p(1-p) = 144p(1-p).$$

Now,

$$\Sigma_{AH} = E[(A - \mu_A)(H - \mu_H)] = E[AH] - \mu_H E[A] - \mu_A E[H] + \mu_A \mu_H = (7-4)(20-24)p(1-p) = -12p(1-p)$$

and then similarly

$$\Sigma_{AW} = (7-4)(132-120)p(1-p) = 36p(1-p)$$

and

$$\Sigma_{HW} = (20-24)(132-120)p(1-p) = -48p(1-p).$$

Thus the covariance matrix is

$$p(1-p)\begin{bmatrix} 9 & -12 & 36 \\ -12 & 16 & -48 \\ 36 & -48 & 144 \end{bmatrix}.$$

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: We begin by diagonalizing K. Its characteristic polynomial is

$$p_K(\lambda) = (3-\lambda)((3-\lambda)(6-\lambda) - 4) + ((-1)(6-\lambda) + 4) + 2(2 - 2(3-\lambda)) = (2-\lambda)^2(8-\lambda),$$

so the eigenvalues of K are 2 (with multiplicity 2) and 8. We now find a basis of eigenvectors: We have that

$$K - 8I = \begin{bmatrix} -5 & -1 & 2 \\ -1 & -5 & -2 \\ 2 & -2 & -2 \end{bmatrix},$$

from which we can spot $\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ as a vector spanning its null space. Thus, $\frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ is an eigenvector (of norm 1) of K corresponding to eigenvalue 8.

Similarly, we have that

$$K - 2I = \begin{bmatrix} 1 & -1 & 2 \\ -1 & 1 & -2 \\ 2 & -2 & 4 \end{bmatrix},$$

from which we can spot $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ as vectors spanning its null space; moreover, these vectors are orthogonal. (In this case, it was fairly easy to find a pair of orthogonal vectors spanning the null space by inspection, but in general you can always row reduce to find a basis for the null space and then apply Gram-Schmidt.) Thus, we have that $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix}$ and $\frac{1}{\sqrt{3}}\begin{bmatrix}-1\\1\\1\end{bmatrix}$ form an orthonormal basis for the eigenspace for eigenvalue 2.

We therefore have that

$$K = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}^T$$

$$= \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

Therefore, the random vector

$$\begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} A \\ H \\ W \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}}A - \frac{1}{\sqrt{6}}H + \frac{2}{\sqrt{6}}W \\ \frac{1}{\sqrt{2}}A + \frac{1}{\sqrt{2}}H \\ -\frac{1}{\sqrt{3}}A + \frac{1}{\sqrt{3}}H + \frac{1}{\sqrt{3}}W \end{bmatrix}$$

consists of random variables which are pairwise independent (see p.94 of the lecture notes, for instance). Their variances are, respectively, 8, 2 and 2.

This process is known as principal component analysis. Note that because the covariance matrix in #2 has rank 2, it has 0 as an eigenvalue. Therefore, by a similar analysis we find that there must be a linear combination in that case of A, H, W which has variance 0, i.e. it is a constant.

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: We have by definition that

$$\Sigma_{XX} = E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 E[1] = E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2,$$

so $E[X^2] = \sigma^2 + \mu^2$.