Exercise Assignment II (2015.2 - T01)

Deliver Deadline: October, 13th, 2015

Exercise 01 Prove mathematically that the Bernoulli distribution satisfies the following properties:

a) $\mathbb{E}[x] = \mu$;

b) $\text{var}[x] = \mu(1 - \mu)$

Exercise 02 In this course's website there is a geneated dataset called `data.dat`. This dataset has $m = 126$ samples and $n = 3$ variables. So assume this dataset is a matrix $\mathbf{X}$. Moreover, $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3]$ where $\mathbf{x}_i$ is the i-th column of $\mathbf{X}$. Using these informations, do what is asked in the following items.

a) Compute, through implementation or through a computer program, the covariance matrix of $\mathbf{X}$.

b) Use the covariance matrix of $\mathbf{X}$ to determine if the vector pair $(\mathbf{x}_1, \mathbf{x}_2)$ is correlated. If so, determine if it is negativelly or posivelly correlated. Do a *scatter-plot* of the vector pair $(\mathbf{x}_1, \mathbf{x}_2)$.

c) Use the covariance matrix of $\mathbf{X}$ to determine if the vector pair $(\mathbf{x}_2, \mathbf{x}_3)$ is correlated. If so, determine if it is negativelly or posivelly correlated. Do a *scatter-plot* of the vector pair $(\mathbf{x}_2, \mathbf{x}_3)$.

d) Through the covariance matrix computed in 2.a, determine $\text{var}[\mathbf{x}_1]$, $\text{var}[\mathbf{x}_2]$ and $\text{var}[\mathbf{x}_3]$.

e) Through implementation, compute the mean of $\mathbf{x}_1, \mathbf{x}_2$ and $\mathbf{x}_3$. After that, plot these vectors and pay attention to the shape of them. Which of these vectors has a Gaussian distribution. Why?

Exercise 03 Still using `data.dat`, plot the histogram of $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. After that, do a histogram density estimation (through implementation) in each vector with the following bin widths Δ and plot the results.

a) $\Delta = 0.01$;

b) $\Delta = 0.05$;

c) $\Delta = 0.25$;

d) $\Delta = 0.50$.

Exercise 04 Still using `data.dat`, plot $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. After that, Implement the K-nearest-neighbour density estimation for each vector with following values of K and plot the results.

a) $K = 1$;

b) $K = 5$;

c) $K = 10$;

d) $K = 15$.

Good luck!