Ex. 1. (3+4+2) Punkte

a. Consider the covariance matrix of data points x_i

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

a. Determine eigenvalues of C

b. compute eigenvectors of the leading two directons

c. Describe the effects that the PCA method has for the given information.

Ex. 2 (4+2+2) Punkte

x_i 1 2 3 4 5

y_i -4.2547 -2.3880 -1.0445 -1.1596 -0.8999

Assume that $y_i = \beta_1 x_i + \beta_0 + \epsilon_i$ with iid $\epsilon \sim N(0,1)$

a. Determine the linear regression vector $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1]^T$. Use the fact that for a matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

the inverse of A is given by

$$A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

b. Compute the adjusted estimator $\hat{\sigma}^2_{ad}$

c. Compute the coefficient of determination R^2. What does the value tell you? How would you interpret this? Justify your answer.

Ex. 3. A teacher is unhappy with the weather precast as she wants to organise a picnic with her class. She starts to get a feeling that the forecast is not better than just flipping a coin where head is predicting rain and tail no rain. She records the weather for 30 randomly selected days. The forecast is correct on 21 of these days. Test if the teachers hypothesis is correct at a 5% significant level. State the hypothesis and justify all your calculations and answers.

4. 7 punkte

Given are following data points

$x^{(1)} = (1,1)$, $x^{(2)} = (3,1)$, $x^{(3)} = (3,3)$, $x^{(4)} = (4,2)$

$x^{(5)} = (6,5)$, $x^{(6)} = (5,6)$, $x^{(7)} = (6,7)$, $x^{(8)} = (4,7)$

Compute 2 iterations of K-means algorithm by hand using the forgy's initialization choosing $x^{(3)}, x^{(4)}, x^{(6)}$. calculate the loss function in each iteration

Exercise 5. (1+4+2+3) points

Brend collects figures from suprice eggs, yet there is not a figure in every egg and he therefore has to open a lot of them at times to finally obtain a figure. Just to get a better idea of how often the figures occur, he writes down how many eggs he has to open to get a figure and collects that data in the table below.

Samples	1	2	3	4	5	6	7	8	9	10
number of n: of trials still sucess (ie found figure)	2	5	8	4	12	2	2	7	2	18

Assume that the samples are generated by drawing from iid random variables $X_i \sim Geom(p)$, i.e the samples are distributed according to the geometric distribution with parameter p.

a. what is the likelihood with respect to the underlying model assumption for parameter p?

b. derive the associated maximum likelihood estimator (MLE) of p?

c. Calculate the MLE of p with respect to the sample set in the table.

d. determine if the MLE of p is unbiased

6 : 3+6 Punkte

consider the filtering setting

$$x_t = A x_{t-1} + \epsilon_t$$

$$y_t = H x_t + n_t$$

where H is the observation operator, $\epsilon_t \sim N(0,B)$ and $n_t \sim N(0,R)$

a. Given that H is the identity, explain under which circumstances $K \approx 0$ and $K \approx 1$. what are the expected effects in these cases?

b. determine m_1^a and P_1^a for

$$A = \begin{bmatrix} 1 & -1 \\ 3 & 2 \end{bmatrix}, y = [0.5], H = [0, 1], R = [0.1]$$

$$B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, m_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \text{ and } P_0 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$