General notes

When calculating results using Octave, please provide code snippets by writing them on your worksheet. This is crucial for me to understand how you came up with your solution.

1 Regression (20 pts)

A start-up company on Wall Street specializes on precious metal trading. Its main business to date is silver trading for which the company has developed a highly precise regressor to predict future prices.

To increase its profitability, the firm considers expanding the portfolio with the trading of platinum. For this goal, a regressor is to be developed that is able to predict the price of platinum. The following table shows historic prices for platinum and silver averaged per year:

	price/(USD/oz)	
year	platinum	silver
2003	691	4.87
2004	845	6.67
2005	896	7.31
2006	1142	11.54
2007	1303	13.38
2008	1573	14.98
2009	1203	14.67
2010	1608	20.19
2011	1721	35.11
2012	1551	31.14

The company's software developers consider regression functions based on two features, time ($x = \text{year} - 2000$) and silver price ($s = \text{silver price/(USD/oz)}$):

a) $y_a = w_0 + w_1 x$,

b) $y_b = w_0 + w_1 s$,

c) $y_c = w_0 + w_1 x + w_2 s$,

d) $y_d = w_0 + w_1 x + w_2 e^x + w_3 s$,

e) $y_e = w_0 + w_1 x + w_2 s + w_3 s^2$,

f) polynomial regression w.r.t. x with $D = 5$ (y_f),

g) polynomial regression w.r.t. s with $D = 4$ (y_g),

Given the above data,

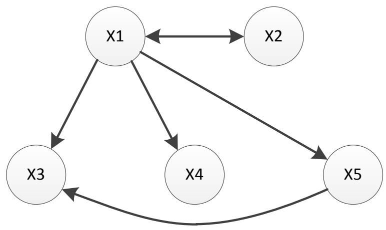
i) train linear regression models for each of the functions $y_a, \dots, y_g$ and calculate the models' empirical risks $R_a, \dots, R_g$.

ii) Predict the average platinum price for 2013 and an estimated silver price of 28.42 USD/oz with each of the models.

iii) Which of the models fits the data best? Justify your answer.

2 Bayesian Networks (10 pts)

We are given the following Bayesian network:



Which of the following equations are correct w.r.t. this network? Which ones are wrong?

$$p(x_3|x_1, x_5) = p(x_3|x_1) \quad (1)$$
$$p(x_5|x_1, x_3) = p(x_5|x_1) \quad (2)$$
$$\sum_{i \in \{1,\ldots,5\}} p(x_i) = 1 \quad (3)$$
$$\max_{x_1} p(x_3|x_1, x_5) = \max_{x_1} \frac{p(x_1, x_5|x_3)p(x_3)}{p(x_1, x_5)} \quad (4)$$
$$\arg\max_{x_1} p(x_3|x_1, x_5) = \arg\max_{x_1} \frac{p(x_1, x_5|x_3)}{p(x_1, x_5)} \quad (5)$$

Justify your answers.

3 Naïve Bayes (20 pts)

An EEG neuroheadset is to be used to detect when a patient is blinking with the eye. To train an eye aperture classifier, four electrodes are mounted to a subject's head to measure brain activity in certain regions. Ten records are taken whereby for each of the electrodes x_1 to x_4, it is measured whether the electrode is excited by volume conduction or not. Furthermore, it is manually determined whether the eye was open or closed (y). The data is given in the following table:

i	x_1	x_2	x_3	x_4	y
1	0	0	0	1	0
2	1	0	1	0	0
3	1	1	0	1	1
4	0	0	1	0	0
5	0	1	1	1	0
6	1	0	1	1	1
7	1	0	1	0	0
8	1	0	1	1	1
9	0	1	1	1	1
10	0	1	1	0	0

Build a naïve Bayes classifier.

a) Classify the following test samples:

$$\begin{aligned} x_{1,1}^{10} &= (0,0,1,0,0,1,1,1,1,1) \\ x_{2,1}^{10} &= (0,0,0,1,0,0,0,0,1,1) \\ x_{3,1}^{10} &= (0,0,1,0,0,0,0,1,1,1) \\ x_{4,1}^{10} &= (0,1,0,1,1,0,1,1,1,0) \end{aligned}$$

In doing so, use a smoothing approach that replaces zero probabilities in $p(x|y)$ by a constant $\alpha = 0.1$.

b) The test samples come along with manually annotated eye open/closed tags:

$$y_1^{10} = (0,0,0,1,0,0,1,1,1,0)$$

Which classification accuracy does the classifier achieve on the data?

c) How is the classification accuracy affected when using a smoothing parameter of $\alpha = 0.5$?

4 Hidden Markov Models (20 pts)

The setup of the eye aperture classifier from Section 3 is to be simplified by reducing the number of necessary electrodes to one (x_4). To compensate for the missing input features, temporal information of the eye aperture is to be exploited by using a hidden Markov model. For the sake of simplicity, the following tables provide the necessary probabilities at a glance:

x_4	y	$p(y)$	$p(y\|x_4)$	$p(x_4\|y)$
0	0	0.6	1	$0.\bar{6}$
0	1	0.4	0	0
1	0	0.6	$0.\bar{3}$	$0.\bar{3}$
1	1	0.4	$0.\bar{6}$	1

y_i	y_{i-1}	$p(y_i\|y_{i-1})$
0	0	0.4
0	1	0.75
1	0	0.6
1	1	0.25

a) Determine the output sequence corresponding to the test feature sequence $x_{4,1}^{5} = (0, 1, 0, 1, 1)$ using a uni-gram HMM.

b) Which classification accuracy do you achieve comparing to $y_1^5 = (0, 0, 0, 1, 0)$?

c) Repeat steps a) and b) using a bi-gram HMM.

Hint: Do not apply smoothing.