General notes

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

1 Regression (6 pts)

The following table shows the dollar/euro exchange rate (t) for the last business day of each of the past 7 months (x):

x	5	6	7	8	9	10	11
t	1.43	1.45	1.41	1.42	1.33	1.38	1.34

Consider the following regression models

a) $y = w_0 + w_1 x$,

b) $y = w_0 + w_1 x + w_2 x^2$, and

c) $y = w_0 + w_1 \sin\left(\frac{\pi}{6}x\right) + w_2 \sin\left(\frac{\pi}{3}x\right) + w_3 \sin\left(\frac{\pi}{2}x\right) + w_4 \sin\left(\pi x\right)$.

Predict the exchange rate of the last business day of the year (i.e., $x = 12$) with each of the models using linear regression.
Which of the models fits the data best? Justify your answer.

2 Conditional independence (6 pts)

Many machine learning techniques are based on modeling assumptions to overcome data sparseness and to render algorithms tractable. One of such assumptions is the *conditional independence* of certain random variables.

We are given three binary random variables (x, y, and z) which may or may not depend on each other in some way. The following table provides the results of a number of drawings as expressed by the count c of co-occurring events:

x	y	z	c
0	0	0	1
0	0	1	4
0	1	0	6
0	1	1	2
1	0	0	3
1	0	1	2
1	1	0	3
1	1	1	1

Assuming the counts reflect the real probability distribution of co-occurring events, determine whether or not

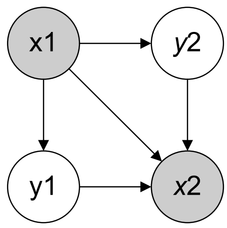
a) $x \perp y|z$

b) $x \perp y|\neg z$

c) $\neg x \perp \neg y|z$

3 Bayesian networks (6 pts)

We are given the following Bayesian network:



a) Formally express the probability $p(x_1, x_2)$.

b) Formally express the probability $p(x_1, x_2, y_1|y_2)$.

c) Show that y_1 is conditionally independent of y_2 given x_1.

Hint: Use the "magic" formula.

4 Naïve Bayes (6 pts)

A typical OCR (optical character recognition) task is the classification of digits given graphical features. The following table provides counts $n_{\{r,a,s,y\}}$ of (binary) features (roundness r, angularity a, straightness s, symmetricity y) co-occurring with a class c (0 to 9) as well as total counts n for each digit as extracted from some training data:

$n_r(c)$	$n_a(c)$	$n_s(c)$	$n_y(c)$	$n(c)$	c
9	0	2	9	13	0
0	0	9	9	17	1
3	3	2	4	8	2
6	2	0	8	15	3
2	9	5	0	11	4
3	8	2	4	9	5
7	3	1	2	14	6
1	3	6	1	12	7
8	3	0	9	18	8
7	3	1	2	11	9

From scanning the zip code on an envelope, sequences of feature values were derived:

$r_1^5 = (1, 1, 0, 0, 1)$

$a_1^5 = (1, 0, 1, 1, 0)$

$s_1^5 = (1, 0, 1, 0, 0)$

$y_1^5 = (1, 1, 0, 0, 1)$

What is the most likely zip code given these features? Use naïve Bayes classification and additive smoothing with $\alpha = 0.001$.

5 Decision trees (6 pts)

A simple app for stock brokers is to propose whether stocks should rather be bought ($c = 1$) or sold ($c = -1$). In order to make this decision, certain market features (gold price g, oil price o) are available for a number of stock types (s = BP [British Petroleum]; s = JPM [JPMorgan Chase]), as well as the number of occurrences (n) and the optimal sales strategy (c) for respective historic situations:

g	o	s	n	c
low	low	JPM	3	1
low	high	JPM	9	1
high	low	BP	5	1
high	low	BP	3	-1
high	low	BP	1	1
high	high	JPM	3	-1

a) Calculate $H(c)$.

b) Which total information gain $I = I(c; g) + I(c; o) + I(c; s)$ do you expect:

   i) $I = 0\text{bit}$

   ii) $0\text{bit} < I < H(c)$

   iii) $I = H(c)$

   iv) $H(c) < I < 1\text{bit}$

   v) $I = 1\text{bit}$

   vi) $I > 1\text{bit}$

   Justify your answer.

c) Which question (g, o, s) will be asked at the root node of a decision tree built by maximizing information gain? Justify your answer.