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 Conditional probabilities and Bayes' rule

(15 pts)

A brand-new technology is to tell words or concepts a human is thinking about by applying machine learning algorithms to brain image data. In the present example, an output feature of the fMRI image analyzer is the hexadecimal $x \in \{0, \ldots, \mathrm{F}\}$. The specific task being investigated is to classify the decimal digits $c \in \{0, \ldots, 9\}$ from brain activity. Due to the poor image analysis model in use, unique features can be associated with multiple classes as the following training data shows. The following count table displays how often unique features are associated with certain classes in the training data:

	c										
x	0	1	2	3	4	5	6	7	8	9	Σ
0	2	1	0	0	2	1	1	8	1	2	18
1	2	2	2	0	9	2	2	8	2	1	30
2	2	0	1	2	1	2	2	2	1	9	22
3	8	1	2	1	1	8	2	2	1	2	28
4	9	0	2	8	1	2	2	2	1	8	35
5	1	2	1	2	2	1	9	1	1	2	22
6	1	1	8	0	2	1	1	2	2	2	20
7	1	2	1	0	1	1	2	8	2	1	19
8	2	9	1	1	0	0	2	1	2	1	19
9	1	9	0	2	2	2	2	2	1	2	23
A	2	1	1	0	2	2	1	1	9	2	21
B	0	2	2	1	2	2	0	2	2	9	22
C	0	1	1	8	1	1	2	1	1	1	17
D	1	1	1	1	1	2	8	1	2	1	19
E	2	2	2	0	0	2	1	9	2	0	20
F	1	1	2	8	2	2	1	0	1	1	19
Σ	35	35	27	34	29	31	38	50	31	44	354

In a field test, the following feature sequence is observed:

$$x_1^8 = (1, \mathrm{E}, \mathrm{A}, 2, 4, 7, \mathrm{B}, \mathrm{F}).$$

a) What is the most likely digit sequence $c_{1(a)}^8$ associated with x_1^8?

The brain image digit classifier is supposed to be used as input interface for the 1986 IBM PC version of the game Tetris. As can be seen in the below screenshot, this is the association between digits and actions:

c	action
7	left
9	right
8	rotate
1	draw next
6	speed up
4	drop

b) Considering the Tetris scenario, what is the most likely digit sequence $c_{1(b)}^{8}$ associated with x_1^8?

From a vast number of games, statistics of game actions are available (n is the number of tokens observed per action):

$n/1000$	action
512	left
516	right
233	rotate
5	draw next
11	speed up
95	drop

c) Considering the Tetris scenario and the game action statistics, what is the most likely digit sequence $c_{1(c)}^{8}$ associated with x_1^8?

2 Regression (15 pts)

The following table shows the average quarterly minimum temperature in Stuttgart, t, for the past 4 years, x (for the sake of simplicity, we use $x = 1$ for "Q1 2008", $x = 2$ for "Q2 2008", and so on):

	2008				2009				2010				2011			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
t	4	6	17	13	1	6	17	15	2	5	17	13	2	7	16	15

Consider the following regression functions

a) $y_a = w_0 + w_1 x$,

b) polynomial regression y_b with $D = 2$,

c) polynomial regression y_c with $D = 15$,

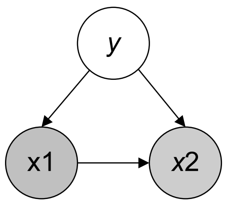
d) $y_d = w_0 + w_1 \sin\left(\frac{\pi}{2}x\right) + w_2 \sin\left(\frac{\pi}{2}\left(x + \frac{1}{4}\right)\right) + w_3 \sin\left(\frac{\pi}{2}\left(x + \frac{1}{2}\right)\right) + w_4 \sin\left(\frac{\pi}{2}\left(x + \frac{3}{4}\right)\right)$.

e) $y_e = w_0 + w_1 \sin(\pi x) + w_2 \sin\left(\pi\left(x + \frac{1}{4}\right)\right) + w_3 \sin\left(\pi\left(x + \frac{1}{2}\right)\right) + w_4 \sin\left(\pi\left(x + \frac{3}{4}\right)\right)$.

Train linear regression models for each of the functions $y_a, \ldots, y_e$ and calculate the models' empirical risks $R_a, \ldots, R_e$.
Predict the average minimum temperatur for Q1 2013 with each of the models.
Which of the models fits the data best? Justify your answer.

3 Bayesian networks (6 pts)

We are given the following Bayesian network:



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

$$
\begin{aligned}
p(y) &= p(y|x_1, x_2) && (1)\\
p(y|x_1, x_2) &= p(y, x_1, x_2) && (2)\\
p(x_1, x_2) &= p(x_1|y)p(x_2|x_1, y) && (3)\\
\min_y p(y|x_1, x_2) &= \min_y p(y, x_1, x_2) && (4)\\
\arg\min_y p(y|x_1, x_2) &= \arg\min_y p(y, x_1, x_2) && (5)\\
\arg\min_y(\log(p(y)) + \log(x_1|y)) &= \arg\min_y(\log(p(y)) + 2\log(x_1|y)) && (6)
\end{aligned}
$$

Justify your answers.

4 Naïve Bayes (20 pts)

A popular application of document classification is e-mail spam filtering. One filter algorithm is based on whether or not the following word types appear in an e-mail:

i	type
1	cash
2	click
3	free
4	go
5	make
6	one
7	out
8	Stuttgart
9	weight
10	winner

To train a classifier for the spam filter, $n(1)$ e-mails have been manually marked as spam ($s = 1$) and $n(0)$ as no spam ($s = 0$). The following table shows in how many of these e-mails the above word types appeared:

s	$n(s)$	$n_1(s)$	$n_2(s)$	$n_3(s)$	$n_4(s)$	$n_5(s)$	$n_6(s)$	$n_7(s)$	$n_8(s)$	$n_9(s)$	$n_{10}(s)$
0	520	0	10	14	230	177	167	84	65	4	0
1	131	56	69	34	87	87	56	50	0	32	48

Build a naïve Bayes classifier. In doing so, use a smoothing approach that replaces zero counts by a constant $\alpha = 1$.

a) Classify the following e-mails:

   i. *Dear students: Feel free to make up for the class next time you are in Stuttgart. Yours, David.*

   ii. *Need \$\$\$? Get cash for your gold teeth. Now!*

   iii. *Payday has come. Just click on the top banner to become rich, famous, and omnipotent. Stop thinking, just CLICK!*

   iv. *Guys, I always told you that the winner class in naïve Bayes classification depends on the smoothing weight as well as on the priors. Keep this in mind for the exam. Good luck, David.*

b) How do classification results in a) change when using $\alpha = 100$?