Exercise 1: kernel machines (34 points). Consider the following dataset $\{(x_n, y_n)\}_{n=1}^N \subset \mathbb{R} \times \{-1,+1\}$ in dimension 1 for binary classification:

$$\{(-6,-1),\ (-5,-1),\ (-1,-1),\ (0,1),\ (1,1),\ (2,1),\ (3,1),\ (4,1)\}.$$

This dataset is linearly separable. You have to find the maximum-margin solution, i.e., the linear support vector machine.

1. (4 points) Write the general formulation as a (primal) quadratic program (QP) of the optimal SVM parameters $(w, w_0) \in \mathbb{R}$ and give the expression for the optimal margin ρ.
2. (6 points) Apply it to the problem above, listing the objective function and all the constraints.
3. (8 points) Plot the problem in the (w, w_0) space, showing contours for the objective function and lines for the constraints. Indicate which part of the space is feasible, i.e., the region of points (w, w_0) satisfying all the constraints.

Although solving a QP cannot be generally done by hand, you can solve this one using a combination of intuition, visual inspection and math.

4. (4 points) Determine the optimal solution (w^*, w_0^*).
5. (2 points) Determine the margin ρ^*.
6. (3 points) Write the discriminant $g(x)$ and determine the boundary between the classes (x such that $g(x) = 0$).
7. (1 point) Verify in your plot that the margin equals the distance from the hyperplane (boundary) to the closest point of either class.
8. (3 points) Give the support vectors and indicate their constraints in the plot.
9. (3 points) Determine a subset of the original dataset $\{x_n\}_{n=1}^N$ above (keeping the same labels) as small as possible such that its optimal SVM would be the same as for the original dataset.

In all of the above, show your work and explain the result. Just giving the final result earns no points, even if correct.

Exercise 2: graphical models (6 points). Consider the following two graphical models defined on binary random variables $X, Y, Z \in \{0, 1\}$, given by their joint distributions:

$$p(X,Y,Z) = p(Z|X,Y)\,p(Y|X)\,p(X) \qquad \text{and} \qquad p(X,Y,Z) = p(Z)\,p(Y|Z)\,p(X)$$

For each of them:

1. (4 points) Prove that $\sum_{X,Y,Z} p(X,Y,Z) = 1$.
2. (2 points) Draw the graphical model.

Exercise 3: graphical models (21 points).

Consider 3 binary random variables with joint distribution given by the table.

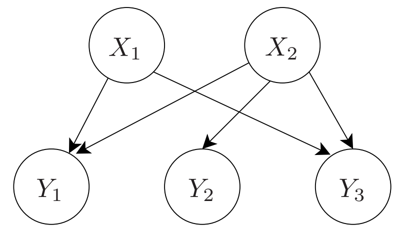
1. (14 points) Evaluate the following probabilities (notation: X means $X = 1$, $\overline{X}$ means $X = 0$): $p(X)$, $p(Y)$, $p(Z)$, $p(X,Y)$, $p(X|Y)$ $p(X|Z)$, $p(Y|Z)$, and $p(X,Y|Z)$.
2. (4 points) Show by direct evaluation that this distribution has the property that X and Y are marginally dependent, i.e., $P(X,Y) \neq p(X)\,p(Y)$ (for all values of X and Y); but that they become independent when conditioned on Z, i.e., $p(X,Y|Z) = p(X|Z)\,p(Y|Z)$ for all values of X, Y and Z.
3. (3 points) Show by direct evaluation that $p(X,Y,Z) = p(Y)\,p(Z|Y)\,p(X|Z)$ (for all values of X and Y). Draw the corresponding directed graph for this graphical model.

X	Y	Z	$p(X,Y,Z)$
0	0	0	0.126
0	0	1	0.168
0	1	0	0.162
0	1	1	0.036
1	0	0	0.014
1	0	1	0.392
1	1	0	0.018
1	1	1	0.084

Show your work in all cases.

Exercise 4: graphical models (21 points).

Consider a graphical model defined on binary random variables (where variables X_i correspond to diseases and variables Y_j to symptoms), given by the following diagram and conditional probability tables at each node.
Note: in the tables and the questions, the notation "$p(Y_3|\overline{X}_1, X_2)$" means "$p(Y_3 = 1|X_1 = 0, X_2 = 1)$", etc.



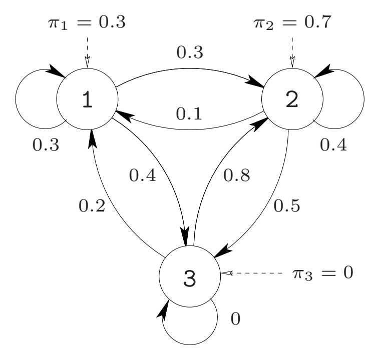
conditional probability tables at each node				
X_1 ("flu")	X_2 ("hayfever")	Y_1 ("headache")	Y_2 ("fever")	Y_3 ("fatigue")
$p(X_1) = 0.3$	$p(X_2) = 0.2$	$p(Y_1\|X_1, X_2) = 0.8$ $p(Y_1\|X_1, \overline{X}_2) = 0.9$ $p(Y_1\|\overline{X}_1, X_2) = 0.6$ $p(Y_1\|\overline{X}_1, \overline{X}_2) = 0.2$	$p(Y_2\|X_2) = 0.9$ $p(Y_2\|\overline{X}_2) = 0.2$	$p(Y_3\|X_1, X_2) = 0.9$ $p(Y_3\|X_1, \overline{X}_2) = 0.8$ $p(Y_3\|\overline{X}_1, X_2) = 0.4$ $p(Y_3\|\overline{X}_1, \overline{X}_2) = 0.5$

1. (3 points) Give the expression of the joint distribution it defines over all the variables.
2. (18 points) Calculate the value of the following probabilities:
   (a) $p(\overline{Y}_3|\overline{X}_1, X_2)$.
   (b) $p(Y_1, \overline{Y}_2|\overline{X}_1, \overline{X}_2)$.
   (c) $p(Y_3)$.
   (d) $p(Y_2|X_1)$.
   (e) $p(X_1|\overline{Y}_1, \overline{Y}_3)$.
   (f) $p(X_2|Y_1, Y_2, Y_3)$.

Show your work in all cases.

Exercise 5: discrete Markov models (11 points).

Consider the discrete Markov model given by the diagram.



1. (3 points) Give the set of states of this discrete Markov model, its transition matrix $\mathbf{A}$ and its vector of initial state probabilities $\boldsymbol{\pi}$.
2. (8 points) Compute the probability of the following sequences: `2213`, `123221` and `3123`.

Show your work in all cases.

Exercise 6: discrete Markov models (7 points). Consider a discrete Markov model with two states `a`, `b`.

1. (5 points) We have a training set consisting of the following sequences: `bbaab`, `ababa`, `babbbb`, `baabaa`. Give the maximum likelihood estimate of the parameters ($\mathbf{A}$, $\boldsymbol{\pi}$).
2. (2 points) Draw the corresponding discrete Markov model as in the previous exercise.

Show your work in all cases.