Introduction to Machine Learning
Homework 4 (50pt)

See Canvas

1 Support Vector Machine (10pt)

[Support Vector Machine.] Consider the following graph and we use a linear SVM for classification with a decision boundary of the form

$$\omega_1 x_1 + \omega_2 x_2 + \omega_3 = 0$$

with $\omega_1^2 + \omega_2^2 + \omega_3^2 = 1$. The data points are $(-4, 0)$, $(0, 0)$, $(0, -4)$, $(0, 2)$, $(1, 1)$ and $(2, 0)$.

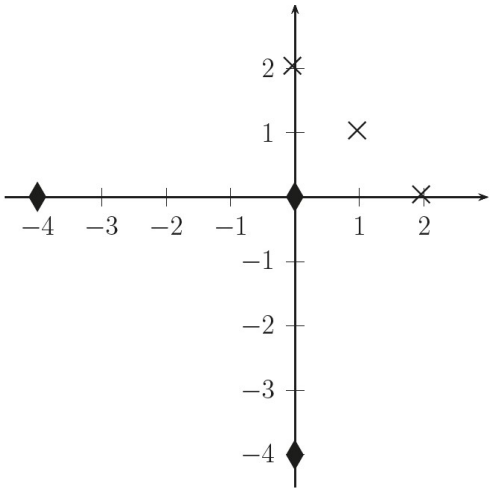


Figure 1: SVM

- **Question 1.** (3 pt) What are the values of $\omega_1, \omega_2, \omega_3$ given by SVM?
- **Question 2.** (3 pt) Which points are the support vectors?
- **Question 3.** (4 pt) Precisely write the expression for the dual problem (assuming Linear SVMs). Let α_i be the Lagrangian multipliers associated with the six data points.

2 Clustering (10pt)

Recall the loss function for k-means clustering with k clusters, sample points $x_1, ..., x_n$, and centers $\mu_1, ..., \mu_k$,

$$L = \sum_{j=1}^{k} \sum_{x_i \in S_j} (x_i - \mu_j)^2$$

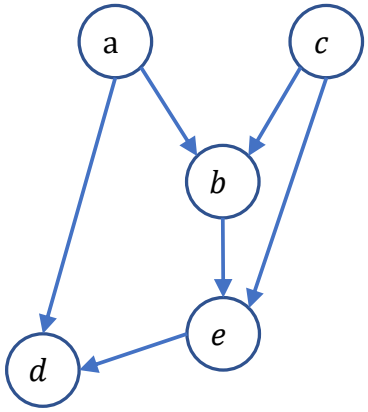
where S_j is the set of data associated with center μ_j.

- **Question 1.** (5 pt) Given the current clusters $S_1, .., S_k$, what is the update formula for computing center μ_1 in the K-means algorithm?
- **Question 2.** (5 pt) Now let us consider a new method to recompute the centers by minimizing L with batch gradient decent, fixing the current clusters $S_1, .., S_k$. Given an update formula for μ_1 with a learning rate α.

3 Bayesian Network (20pt)

3.1 Problem 1 (10pt)

Consider the following Bayesian network where all the variables are binary.



- **Question 1.** (2pt) What are the factors according to the given graph?
- **Question 2.** (3pt) Please use d-separation to discuss if the following independence can be guaranteed. (a) $a \perp e|b$ (b) $a \perp e|b, c$ (c) $a \perp e|b, c, d$
- **Question 3.** (5pt) Show the process of variable elimination for inferring $\Pr[d]$ along the order (b, c, a, e). Please include the factors that are newly generated. Please estimate the number of operations involved in the process.

3.2 Problem 2 (10pt)

[Bayesian Network.] We will use EM-algorithm to learn a Bayesian Network consisting of distributions: $\Pr[A]$, $\Pr[B|A]$, $\Pr[C|A]$ over three binary variables A, B and C.

- **Question 1.**(3pt) Please draw the Bayesian network.
- **Question 2.** (4pt) At the E-step, suppose we have

$$\begin{aligned}\Pr[A=1] &= 0.9,\\ \Pr[B=1|A=1] &= 0.1,\\ \Pr[B=1|A=0] &= 0.6,\\ \Pr[C=1|A=1] &= 0.7,\\ \Pr[C=0|A=0] &= 0.3,\end{aligned}$$

  and the current dataset is given in the following table. Please give the complete weighted dataset after the E-step.

A	B	C
0	1	?
0	1	1
?	0	1
1	1	?
1	0	?
0	0	0
1	1	1

- **Question 3.** (3pt) Please show the computation of the next M-step after the E-step in the last question.

4 Learning Theory (10pt)

In a two-dimension space, consider the problem of using axis-parallel rectangles $H = \{(a \leq x \leq b) \wedge (c \leq y \leq d) | a, b, c, d \in \mathcal{R}\}$ to binary classify the points $X = \{(x, y) | x, y \in \mathcal{R}\}$: a point is positive iff it falls in the rectangle.

- **Question 1.** (5pt) Compute VC(H).
- **Question 2.** (5pt) Find a number of training examples drawn randomly to assure that for any target in H, any consistent learner using H will, with probability at least 95%, output a hypothesis with error at most 0.15. (using the upper bound for VC dimension)

What to Turn in

- A report with your answer.