

# HOMEWORK 7

>>NAME HERE<<

>>ID HERE<<

**Instructions:** Although this is a programming homework, you only need to hand in a pdf answer file. There is no need to submit the latex source or any code. You can choose any programming language, as long as you implement the algorithm from scratch.

Use this latex file as a template to develop your homework. Submit your homework on time as a single pdf file to Canvas. Please check Piazza for updates about the homework.

## 1 Directed Graphical Model [30 points]

Consider the directed graphical model (aka Bayesian network) in Figure ??.

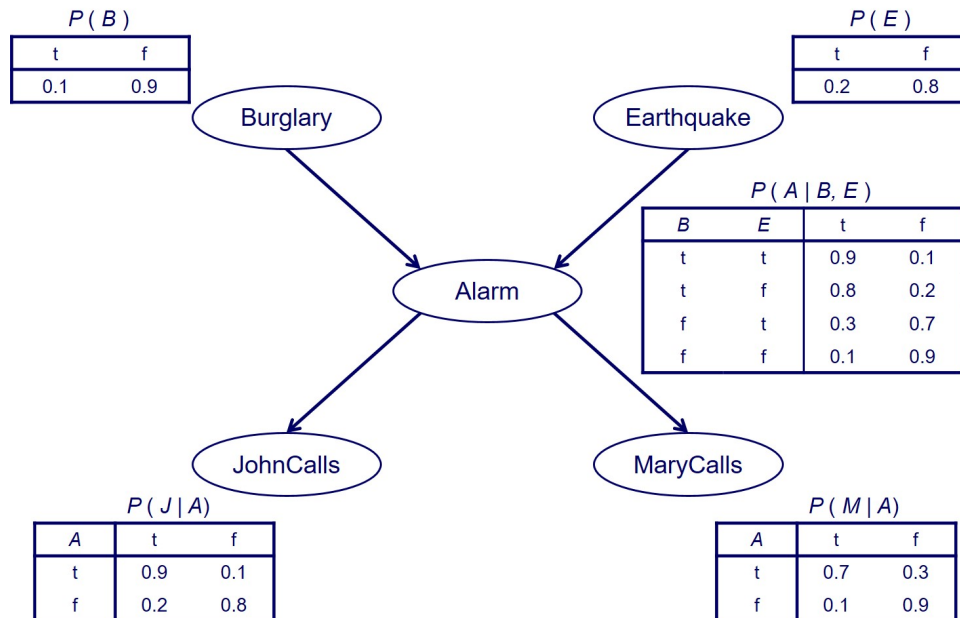


Figure 1: A Bayesian Network example.

Compute  $P(B = t | E = f, J = t, M = t)$  and  $P(B = t | E = t, J = t, M = t)$ . These are the conditional probabilities of a burglar in your house (yikes!) when both of your neighbors John and Mary call you and say they hear an alarm in your house, but without or with an earthquake also going on in that area (what a busy day), respectively.

## 2 Chow-Liu Algorithm [40 pts]

Suppose we wish to construct a directed graphical model for 3 features  $X$ ,  $Y$ , and  $Z$  using the Chow-Liu algorithm. We are given data from 100 independent experiments where each feature is binary and takes value  $T$  or  $F$ . Below is a table summarizing the observations of the experiment:

$X$	$Y$	$Z$	Count
T	T	T	36
T	T	F	4
T	F	T	2
T	F	F	8
F	T	T	9
F	T	F	1
F	F	T	8
F	F	F	32

1. Compute the mutual information  $I(X, Y)$  based on the frequencies observed in the data.
2. Compute the mutual information  $I(X, Z)$  based on the frequencies observed in the data.
3. Compute the mutual information  $I(Z, Y)$  based on the frequencies observed in the data.
4. Which undirected edges will be selected by the Chow-Liu algorithm as the maximum spanning tree?
5. Root your tree at node  $X$ , assign directions to the selected edges.

### 3 Kernel SVM [30 points]

Consider the following kernel function defined over  $z, z' \in Z$ :

$$k(z, z') = \begin{cases} 1 & \text{if } z = z', \\ 0 & \text{otherwise.} \end{cases}$$

1. Prove that for any integer  $m > 0$ , any  $z_1, \dots, z_m \in Z$ , the  $m \times m$  kernel matrix  $K = [K_{ij}]$  is positive semi-definite, where  $K_{ij} = k(z_i, z_j)$  for  $i, j = 1 \dots m$ . (Let us assume that for  $i \neq j$ , we have  $z_i \neq z_j$ .) Hint: An  $m \times m$  matrix  $K$  is positive semi-definite if  $\forall v \in \mathbb{R}^m, v^\top K v \geq 0$ .
2. Given a training set  $(z_1, y_1), \dots, (z_n, y_n)$  with binary labels, the dual SVM problem with the above kernel  $k$  will have parameters  $\alpha_1, \dots, \alpha_n, b \in \mathbb{R}$ . (Assume that for  $i \neq j$ , we have  $z_i \neq z_j$ .) The predictor for input  $z$  takes the form

$$f(z) = \sum_{i=1}^n \alpha_i y_i k(z_i, z) + b.$$

Recall the label prediction is  $\text{sgn}(f(z))$ . Prove that there exists  $\alpha_1, \dots, \alpha_n, b$  such that  $f$  correctly separates the training set. In other words,  $k$  induces a feature space rich enough such that in it any training set can be linearly separated.

3. How does that  $f$  predict input  $z$  that is not in the training set?

Comment: One useful property of kernel functions is that the input space  $Z$  does not need to be a vector space; in other words,  $z$  does not need to be a feature vector. For all we know,  $Z$  can be turkeys in the world. As long as we can compute  $k(z, z')$ , kernel SVM works on turkeys.

### 4 Extra Credit: Kernel functions over discrete space [10 points]

Kernel functions can be defined over objects as diverse as graphs, sets, strings, and text documents. Consider, for instance, a fixed set  $D$  and define a nonvectorial space consisting of all possible subsets of this set  $D$ . If  $A_1$  and  $A_2$  are two such subsets then one simple choice of kernel would be

$$k(A_1, A_2) = 2^{|A_1 \cap A_2|}$$

where  $A_1 \cap A_2$  denotes the intersection of sets  $A_1$  and  $A_2$ , and  $|A|$  denotes the size of  $A$ . Show that this is a valid kernel function, by showing that it corresponds to an inner product in a feature space.

[Solution goes here.](#)

## 5 Extra Credit: Support Vector Machines [10 points]

Given data  $\{(x_i, y_i), 1 \leq i \leq n\}$ , the (hard margin) SVM objective is

$$\begin{aligned} \min_{w, b} \quad & \frac{1}{2} \|w\|_2^2 \\ \text{s.t.} \quad & y_i(w^\top x_i + b) \geq 1 (\forall i). \end{aligned}$$

The dual is

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^\top x_j \\ \text{s.t.} \quad & \alpha_i \geq 0 (\forall i), \quad \sum_{i=1}^n \alpha_i y_i = 0. \end{aligned}$$

Suppose the optimal solution for the dual is  $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)$ , and the optimal solution for the primal is  $(w^*, b^*)$ . Show that the margin

$$\gamma = \min_i \frac{y_i((w^*)^\top x_i + b^*)}{\|w^*\|_2}$$

satisfies

$$\frac{1}{\gamma^2} = \sum_{i=1}^n \alpha_i^*.$$

Hint: use the KKT conditions.

[Solution goes here.](#)