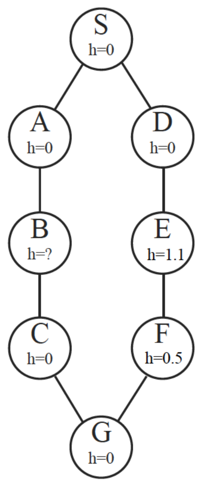
CSE 415 Winter 2018 Review Session Worksheet

Your Name: ____________________________

1 A* search

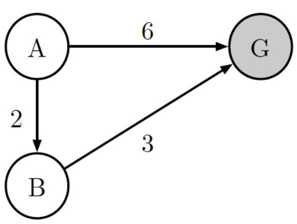
Consider the following graph, where S is the start and G is the goal state. Each edge has a cost of 1.



1. Say h(F) = 0.5 is only given (ignore the other h values for now). What are the ranges of h(D) for admissibility and consistency?
2. Now consider all the values and assume ties are broken alphabetically. For what range of values of h(B) will (a) B be expanded before E expanded before F, (b) E be expanded before B expanded before F?

2 A* search

Consider the directed graph below with A: start, G: goal.

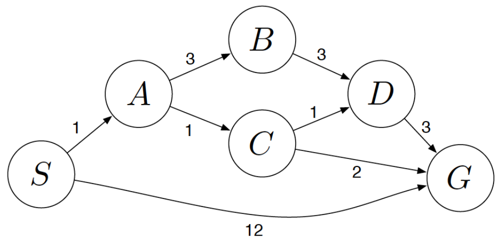


	$h(A)$	$h(B)$	$h(G)$
I	4	1	0
II	5	4	0
III	4	3	0
IV	5	2	0

1. Check for admissibility and consistency of each heuristic I, II, III and IV
2. For one heuristic to dominate another, **all** of its values must be greater than or equal to the corresponding values of the other heuristic. If not, the two heuristics have no dominance relationship. What is the dominance relationship of the given heuristics?

3 A* search

Consider the directed graph below with S: start, G: goal.



State	h_1	h_2
S	5	4
A	3	2
B	6	6
C	2	1
D	3	3
G	0	0

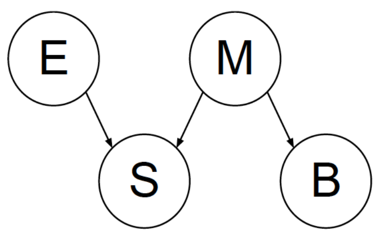
1. Find the optimal solution path for the graph given any consistent heuristic.
2. Check for admissibility and consistency of h1 and h2.

4 Joint Distribution and Naive Bayes

Consider the following tables:

$P(E)$	
$+e$	0.4
$-e$	0.6

$P(S\|E,M)$			
$+e$	$+m$	$+s$	1.0
$+e$	$+m$	$-s$	0.0
$+e$	$-m$	$+s$	0.8
$+e$	$-m$	$-s$	0.2
$-e$	$+m$	$+s$	0.3
$-e$	$+m$	$-s$	0.7
$-e$	$-m$	$+s$	0.1
$-e$	$-m$	$-s$	0.9



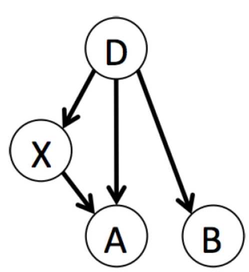
$P(M)$	
$+m$	0.1
$-m$	0.9

$P(B\|M)$		
$+m$	$+b$	1.0
$+m$	$-b$	0.0
$-m$	$+b$	0.1
$-m$	$-b$	0.9

1. Calculate P(e, s, m, b) and P($+m| + s, +e, +b$)
2. Calculate P($+b$), P($+m| + b$) and P($+e| + m$)

5 Bayes Net

$P(A\|D,X)$			
$+d$	$+x$	$+a$	0.9
$+d$	$+x$	$-a$	0.1
$+d$	$-x$	$+a$	0.8
$+d$	$-x$	$-a$	0.2
$-d$	$+x$	$+a$	0.6
$-d$	$+x$	$-a$	0.4
$-d$	$-x$	$+a$	0.1
$-d$	$-x$	$-a$	0.9



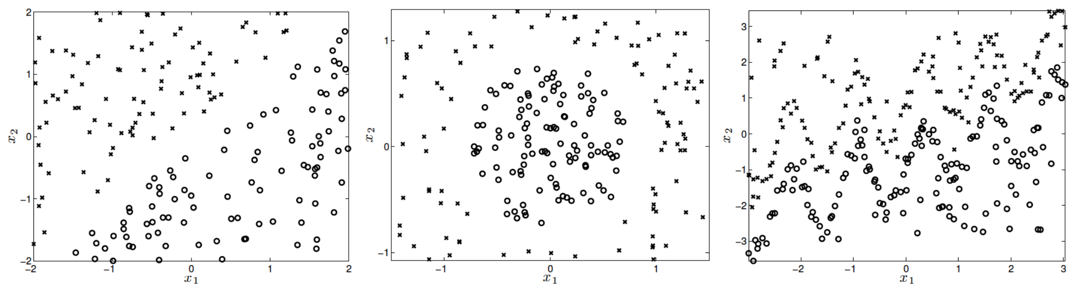
$P(D)$	
$+d$	0.1
$-d$	0.9

$P(X\|D)$		
$+d$	$+x$	0.7
$+d$	$-x$	0.3
$-d$	$+x$	0.8
$-d$	$-x$	0.2

$P(B\|D)$		
$+d$	$+b$	0.7
$+d$	$-b$	0.3
$-d$	$+b$	0.5
$-d$	$-b$	0.5

1. Find P($-d, +a$), P($+d| + a$) and P($+d| + b$). Is B independent of A?

6 Perceptrons with kernels



(A) (B) (C)

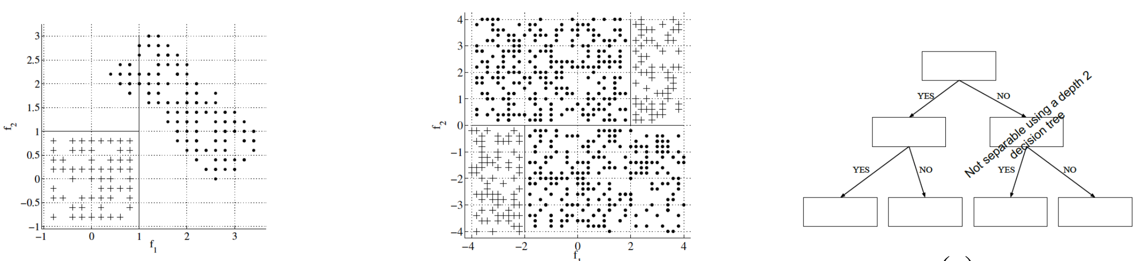
(i) Linear kernel: $K(x, z) = x^\top z = x \cdot z = x_1 z_1 + x_2 z_2$

(ii) Polynomial kernel of degree 2: $K(x, z) = (1 + x^\top z)^2 = (1 + x \cdot z)^2$

(iii) RBF (Gaussian) kernel: $K(x, z) = \exp\left(-\frac{1}{2\sigma^2}\|x - z\|^2\right) = \exp\left(-\frac{1}{2\sigma^2}(x - z)^\top (x - z)\right)$

1. Datasets are A, B, C and $x = [x_1, x_2]^T$. Which kernel(s) is suitable for which dataset? Do you need bias for the linear kernel? Name a classifier that is good for A and B but not for C.

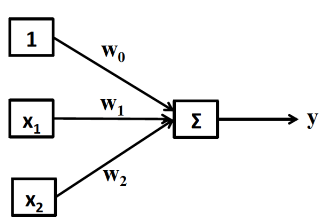
7 Decision Trees



(a) (b) (c)

1. Construct the decision tree in (c) for each of (a) and (b). If construction is not possible, write not separable as shown in (c).
2. The conditional independence assumption of Naive Bayes model is that features are conditionally independent when given the class. Do the set of points in (a) and (b) satisfy the assumption?

8 Perceptron weight update



$$y = \begin{cases} 1 & \text{if } (w_0 + w_1 \cdot x_1 + w_2 \cdot x_2) > 0 \\ 0 & \text{otherwise} \end{cases}$$

1. For what values of w_0 will the perceptron correctly classify $y = x_1 XOR x_2$ and $y = x_1 AND x_2$?

2. Suppose initial weights for a multi-class perceptron is $w_A = [1, 0, 0], w_B = [0, 1, 0]$ and $w_C = [0, 0, 1]$ and A is chosen over B over C in case of tie. What will be the weights after a feature vector (training data) $[1, -2, 3]$ with label A is added? In the next step, what will be the weights after a feature vector (training data) $[1, 1, -2]$ with label B is added? In the next step, what will be the weights after a feature vector (training data) $[1, -1, -4]$ with label B is added?

9 Naive Bayes

Given the table below where W stands for word and given W_1 = perfect and W_2 = note,

W	note	to	self	become	perfect
$P(W \mid Y = \text{spam})$	$1/6$	$1/8$	$1/4$	$1/4$	$1/8$
$P(W \mid Y = \text{ham})$	$1/8$	$1/3$	$1/4$	$1/12$	$1/12$

find the values of P(Y=spam) for which the classifier will produce label "spam".

10 Maximum likelihood and Laplacian smoothing

1. $P(X = k) = (1 - \theta)^{k-1}\theta$ for $k \in \{2, 4, 7, 9\}$. Write the log-likelihood expression as a function of θ and find the maximum likelihood estimate for θ.

2. Consider Bayes' net with two variables A, B and structure A $->$ B. Find the ML and k = 2 Laplace estimates for each of the table entries based on the following data: $(+a, -b), (+a, +b), (+a, -b), (-a, -b), (-a, -b)$.

A	$P^{\text{ML}}(A)$	$P^{\text{Laplace, k=2}}(A)$
$+a$		
$-a$		

A	B	$P^{\text{ML}}(\mathbf{B} \mid \mathbf{A})$	$P^{\text{Laplace, k=2}}(\mathbf{B} \mid \mathbf{A})$
$+a$	$+b$		
$+a$	$-b$		
$-a$	$+b$		
$-a$	$-b$		

11 Miscellaneous (FYI)

1. For MDPs, a small discount (close to 0) and/or a large negative living reward encourages shortsighted, greedy behaviour.

2. If H1 and H2 are both admissible heuristics, min(H1,H2), max(H1, H2) and (H1+H2)/2 are also admissible. Any weighted average of H1 and H2 is also admissible. However, they are not necessarily consistent.
   P.S. What will happen is H1 is admissible and H2 is not?
   P.P.S. Suppose a heuristic h's values are multiplied by 'n'. Will the solution of A* be still optimal and what will be the cost of that solution?

3. If the optimal Q-function, Q*, can't be represented as a weighted combination of features, then the feature-based representation can't find Q*.

4. In the limit of infinite data, Laplace smoothing doesn't change the parameter estimates.

5. How to determine and avoid overfitting and underfitting during classifier training?