Name: ____

Problem	Points	Score
1(a)	20	
1(b)	10	
1(c)	10	
2(a)	10	
2(b)	10	
2(c)	10	
2(d)	10	
3	20	
Total	100	

Notes:

(1) The exam is closed books and notes except for one double-sided sheet of notes.

(2) Please indicate clearly your answer to the problem.

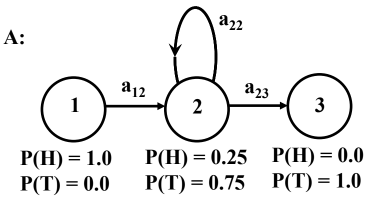
(3) Please try to make your solution legible and easy to follow. The better I can understand your thought process, the more generous I can be about partial credit. I will not give partial credit for ungrammatical sentences or fragmented answers. Please collect your thoughts and compose coherent answers.

(4) If you aren't sure how to work the details of a problem, at the very least write an outline of your solution indicating the step by step process that you think is needed to solve the problem.

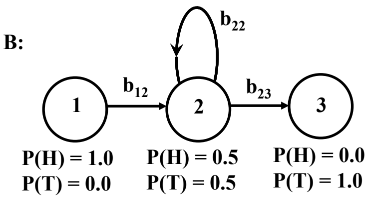
(40 pts) Problem No. 1:

Problem No. 1: Consider the following models for a system that outputs sequences of the characters "H" and "T". For these models, you must start in state 1 and end in state 3.

Model A:



Model B:



(a) Compute the probability that model A produced the sequence "HHH". Be as specific as possible in explaining your solution.

(b) Which model most likely produced the sequence "HTH." Explain.

(c) Which state sequence most likely produced the sequence "HTH." What was the probability of that state sequence?

(40 pts) Problem No. 2: You are given a training set that consists of random samples from two uniform distributions. The first distribution, associated with Class 1, is a two-dimensional uniform distribution with a mean of (0,0) and extends from $-1 \le x \le 1$ and $-1 \le y \le 1$. The second distribution has a mean of (0.5,0.5) and extends from $0 \le x \le 1$ and $0 \le y \le 1$.

(a) Explain how you would design a Support Vector Machine (SVM) to model this data that would minimize the error rate on the training data. Could you find an SVM that would produce 0% error? What is the minimum achievable error rate?

(b) Compute the probability of error that will be achieved on the training set (closed-set testing) by a classifier designed using a nearest-neighbor algorithm (e.g., kNN) based on a majority voting scheme (as we discussed in class). Assume k is infinite – all the training data is used in each decision.

(c) Suppose Class 1 was represented by a Gaussian distribution constrained to have an identity covariance matrix. Suppose Class 2 was represented the same way. The means of these distributions were trained on the specific training data above. How would the probability of error change from part (a)?

(d) How would this error rate change if you used a Gaussian Mixture Model (GMM) with two mixture components per class? Sketch the resulting model (show where the means would fall if training was done properly).

(20 pts) Problem No. 3: It is Day 15 of the outbreak of a worldwide pandemic. As the nation's chief health officer, you are asked to perform a grim prediction of how many patients in the U.S. will be infected from COVID-19. You are provided the data shown in the table below by your expertly trained staff (the third row of the header are the populations of the respective countries). Place your estimates, which you will deliver to the media this afternoon, in the final column. You must provide an answer of course since the public is hungry for answers!

Justify your estimates using as many concepts that we have learned in this course as possible. Your numbers should be based on some mathematical model you derive from the data. You should explain why that model is appropriate, how you would estimate the parameters of that model, etc. Discuss your confidence in these estimates, citing best-case, worst-case and typical scenarios. As you have seen in the past few weeks, the societal and economic implications of your decisions are far-reaching, so you better be sure of your model.

	Country			
	China	Italy	Spain	U.S.
Day	1,386M	60M	46M	327M
01	1	0	0	0
02	2	0	0	0
03	4	0	0	0
04	8	2	0	0
05	16	8	0	0
06	32	32	0	0
07	64	128	0	0
08	128	512	32	0
09	256	2,048	64	0
10	512	8,192	128	1
11	1,024	32,768	512	2
12	2,048	131,072	1,024	4
13	4,096	524,288	512	8
14	8,192	2,097,152	256	128
15	16,384	1,048,576	64	
16	32,768	524,288	128	
17	65,536	262,144	1,024	
18	131,072	131,072	8,192	
19	262,144	65,536	131,072	
20	524,288	32,768	1,048,576	
21	1,048,576	16,384	524,288	
22	524,288	8,192	262,144	
23	262,144	4,096	131,072	
24	131,072	2,048	65,536	
25	65,536	1,024	32,768	
26	32,768	512	16,384	
27	16,384	256	8,192	
28	8,192	128	4,096	
29	4,096	64	2,048	
30	2,048	32	1,024	
31	1,024	16	512	
32	512	8	256	
33	256	4	128	
34	128	2	64	
35	64	1	32	
36	32	2	16	
37	16	3	8	
38	4	2	4	
39	6	1	2	
40	3	3	1	

	Country			
	China	Italy	Spain	U.S.
Day	1,386M	60M	46M	327M
41	5	2	2	
42	2	1	3	
43	7	3	2	
44	3	2	1	
45	1	4	2	