Name:

|  |  |  |
| --- | --- | --- |
| Problem | Points | Score |
| 1(a) | 20 |  |
| 1(b) | 10 |  |
| 2(a) | 20 |  |
| 2(b) | 10 |  |
| 3(a) | 20 |  |
| 3(b) | 20 |  |
| Total | 100 |  |

Notes:

1. The exam is closed books and notes except for four double-sided sheets of notes.
2. Please indicate clearly your answer to the problem.
3. If I can’t read or follow your solution, it is wrong and no partial credit will be awarded.

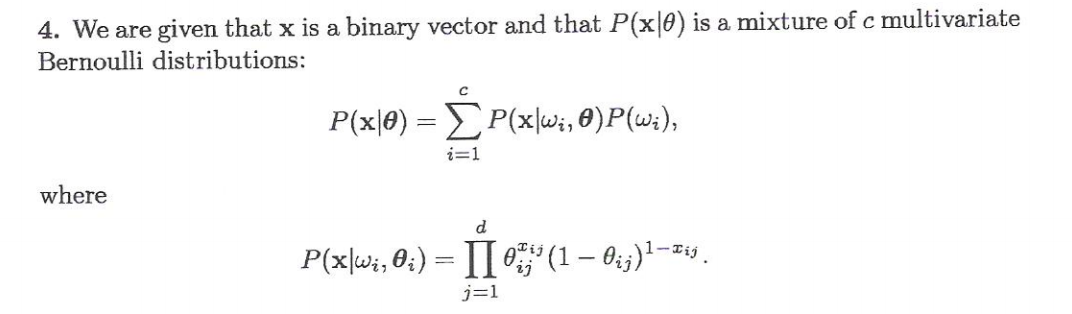
**Problem No. 1**: Consider the two distributions shown.

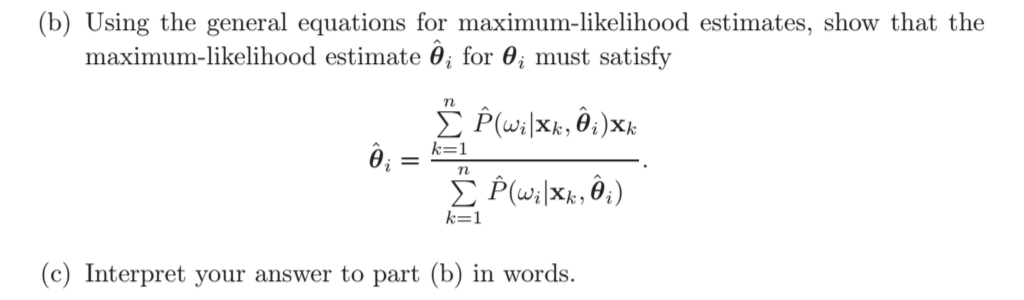


(a) Assuming equal priors, for what values of the mean, m1 and m2, will the minimum error rate for a maximum likelihood classifier be 50%?

(b) How would these values change if the prior for class 1 was 0.9? (An approximate answer is fine.) What if the prior for class 1 was 0.9 and the variance was changed to 10?

**Problem No. 2:**





**Problem No. 3:** For the three data sets shown on the next page (assume equal priors):

(a) Design an optimal classifier for each, following the principle of Occam’s Razor (keep it as simple as possible yet it should give optimal performance). Compare and contrast these classifiers from a recognition performance point of view, explaining why each one is most appropriate for the corresponding data sets. For example, the solution for the third data set might not be the most appropriate solution for the first data set. Why?

(b) Compare and contrast the computational complexity of these approaches. Discuss how they scale with the number of training data points. I want you to specifically discuss evaluation (recognition) and training in terms of computational complexity.

