Test Two Rework

One. Given two distributions: P(x|w_1) = 1 for 0 ≤ x ≤ 1, 0 elsewhere
P(x|w_2) = 1 for 1/2 ≤ x ≤ 3/2, 0 elsewhere

Assume equal priors.

One-A. Build a nearest neighbor classifier using two randomly draw points from each class.

Class 1 points: ½ & ¾ Class 2 points: 1 & 4/3
Use mean squared error to find distance between the random test points.
Distance = (point_a – point_b)^2.

Distance between points

	½	¾	1	4/3
½	**0**	**1/16**	**9/16**	**25/36**
¾	**1/16**	**0**	**1/16**	**49/144**
1	**9/16**	**1/16**	**0**	**1/9**
4/3	**25/36**	**49/144**	**1/9**	**0**

From the results assume k=2 to break the points into two clusters. The results indicate that 4/3 is a nearest-neighbor of 1 and that ½ is a nearest-neighbor of ¾. A bit of error comes into play in that ¾ shares the same distance with ½ and 1. However, as k=2 ½ and ¾ will be paired up while 1 and 4/3 make the other pairing.

Using these points, a classifier can be built and tested against the points 0, ½, 1, and 3/2. The new distances can be computed with the same mean squared error formula.

Distance Between Points, Classifier

	0	½	1	3/2
½	**¼**	**0**	**¼**	**1**
¾	**9/16**	**1/16**	**1/16**	**9/16**
1	**1**	**¼**	**0**	**¼**
4/3	**16/9**	**25/36**	**1/9**	**1/36**

With this approach the smallest aggregated distance between each classifier will be used to determine the class for each point. This results in Class One taking on 0 and ½, while Class Two would take 1 and 3/2. Despite the clear divide between the nearest-neighbor of each point, there is a large amount of error because ½ and 1 could come from either class. With two points that could have gone either way and the priors being equal, the error of this classifier is 25%.

One-B. Explain what happens as you allow the number of points drawn to increase. Show that your result in (A) converges to the correct result.

As the number of points drawn increases the error rate will settle at 25%. Values drawn between 0 and ½ and between 1 and 3/2 represent half the total distribution of values from the two classes. The other half of possibilities overlaps between ½ and 1 but both classes are equally likely over that range presenting 50% for either class to be correct. This zone contributes 25% error, 50% likelihood of data being in the range and 50% chance of accurate classification. This can also come from a Bayesian approach where the likelihood of a given class is 50% and the error associated with either class is also 50% this the overall error rate, the theoretical limit, is .5*.5 or 25%.

From part A, the probability of a point falling between 0 and ½ is 25%, from between ½ and 1 is 50%, and from 1 and 3/2 is 25%. The probability of error for Class one, P_e(x|w_1) = 50%, and for Class two, P_e(x|w_2) =50%. Regardless of the number of points drawn in a random fashion, as the number of points grows the system will tend toward the asymptotic overall error rate seen from the Bayesian approach.

Two. Given two HMM models that generate the sequence '%$%'

Model A.

Transition Matrix

	End State: 1	End State: 2	End State: 3
Start State: 1	**75%**	**25%**	**N/A**
Start State: 2	**N/A**	**25%**	**75%**

Emission Matrix

	Output $	Output %
State: 1	**25%**	**75%**
State: 2	**75%**	**25%**

Model B.

Transition Matrix

	End State: 1	End State: 2	End State: 3
Start State: 1	**25%**	**75%**	**N/A**
Start State: 2	**N/A**	**75%**	**25%**

Emission Matrix

	Output $	Output %
State: 1	**75%**	**25%**
State: 2	**25%**	**75%**

Two-A. Probability that Model A produced the sequence '%$%'?

$$P_1(\%)*a11*P_1(\$)*a12*P_2(\%)*a23 + P_1(\%)*a12*P_2(S)*a22*P_2(\%)*a23 = 2.64\%$$

Two-B. Which model most likely produced the sequence '%$%'?

From part A, it is known that Model A produces the desired output 2.64% of the time. The likelihood of Model B producing the sequence presents as a similar equation.

$$P_1(\%)*b11*P_1(\$)*b12*P_2(\%)*b23 + P_1(\%)*b12*P_2(S)*b22*P_2(\%)*b23 = 0.88\%$$

The resulting likelihood of Model B generating the output is far lower than that of Model A, which makes Model A more likely to produce the desired output sequence.

Two-C. Which state sequence most likely produced the sequence '%$%'. What was the probability of that state sequence?

First, determine which set of transitions are most likely to happen for a given sequence. As the only viable options are 1 → 1 → 2 → 3 and 1 → 2 → 2 → 3 so the likely hood of each sequence must be computed for each model's transition matrix.

TRANSITION	1 → 1→ 2 → 3	1 → 2 → 2 → 3
Model A	14.06%	4.68%
Model B	4.68%	14.06%

Each sequence appears to be equally likely given the transition matrices for Model A and Model B. The emission matrices will show which path is more likely to produce the desired output and should thus be used to distinguish which sequence is optimal.

EMISSION	1 → 1→ 2 → 3	1 → 2 → 2 → 3
Model A	4.68%	14.06%
Model B	14.06%	4.68%

Wait. This isn't helpful at all! Everything is equal! The trick is in checking each possible path through the system which has been computed in the part B.

Model A, Path 1,1,2,3 – 1.98%
Model A, Path 1,2,2,3 – 0.66%
Model B, Path 1,1,2,3 – 0.22%
Model B, Path 1,2,2,3 – 0.66%

From this set of data it seems clear that path 1 → 1 → 2 → 3 has the highest chance of producing the desired output at 2.2% over 1.32%.

Two-D. Give at least two reasons why the probabilities in (A) and (C) differ.

Model A benefits from possessing probabilities that only required two 25% options for the desired output for the route 1 → 1 → 2 → 3, despite Model B performing the worst for the same route, as it requires four 25% options. The alternative route for both models requires three 25% options.

Model A also presents the best option regardless of path because the initial required emission is highest in State 1 of Model A at 75% where Model B will suffer with 25%. This is the one 75% option that Model B is unable at access due to the required sequence and allows Model A to prevail for part A. Even though the alternative path performs equally well for both models, the advantage of the starting position in Model A for the best path is greater than the aggregate of all the Model and path combinations.