Name: Vira Oleksyuk (redo exam)

Problem	Points	Score
1(a)	30	
1(b)	20	
2(a)	10	
2(b)	20	
2(c)	10	
2(d)	10	
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) If I can't read or follow your solution, it is wrong and no partial credit will be awarded.

Problem No. 1: Consider two probability distributions defined by:

$$p(x|\omega_1) = \left\{ \begin{array}{cc} 1 & 0 \le x \le 1 \\ 0 & elsewhere \end{array} \right\} \text{ and } p(x|\omega_2) = \left\{ \begin{array}{cc} 1 & 1/2 \le x \le 3/2 \\ 0 & elsewhere \end{array} \right\}$$

and assume equal priors.

(a) Draw two points at random from each class. Design a nearest-neighbor classifier based on these points. Compute the probability of error.

Fig. 1 shows distributions $p(x|w_1)$ and $p(x|w_2)$. I plotted 4 random points in red from those distributions. Values of the points are given in Table 1.

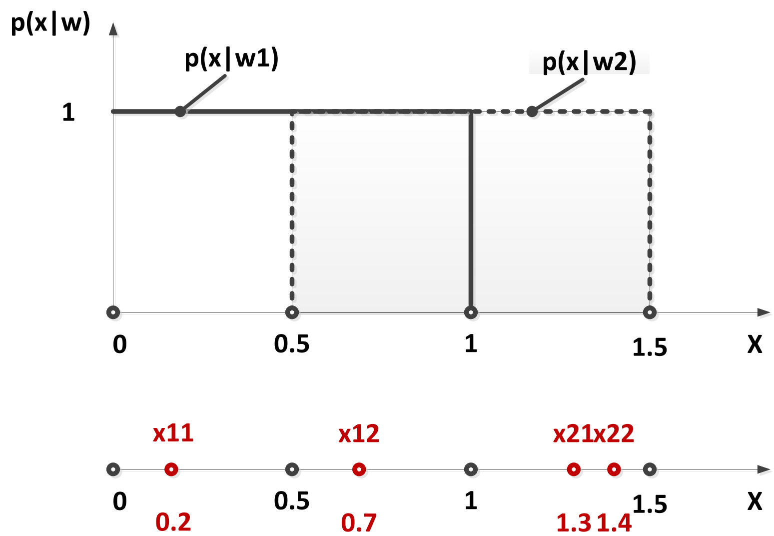

Figure 1. Drawn points from two distributions.

Table 1 shows randomly drawn data points from two distributions.

Table 1. Coordinates of the randomly drawn data points from the given distributions.

Data point	Coordinate	Assigned Class
X11	0.2	1
X12	0.7	1
X21	1.3	2
X22	1.4	2

The priors are equal:

$$P(w_1) = P(w_2) = 0.5$$

For the nearest neighbor classifier, we assume one neighbor (k=1).

If we draw an arbitrary x on 0X coordinate, it will be classified according to the class of its closest neighbor. For the case on Fig.1, the decision (F) will be at: $F = \frac{1.3-0.7}{2} = 1$, so all random x with value

less than 1 will be classified as *class 1*, and all the *x*, which values are greater than 1, will be classified as *class 2*.
Points of Class 1 will not be misclassified, yet points drawn from distribution $p(x|w_2)$ will be misclassified at x = (0.5,1) region. The error for the constructed classifier will be following equation and will contain only misclassified Class 2 points:

$$P(error) = \int_0^0 p(x|w_1)\,P(w1) + \int_{.5}^1 p(x|w_2)\,P(w2) =$$
$$= \int_{.5}^1 p(x|w_2)\,P(w2) = 0.5 - 0.5 \cdot 0.5 = 0.25$$

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.

When we will start to draw more and more points from those two distributions and will construct a nearest neighbor classifier with them, we will approach 0.25 error due to overlap of those distributions and their priors. Decision F will shift to $F = \frac{1-0.5}{2} = 0.75$, because we will draw points from those distributions equally likely and the middle of their overlap located at x=0.75. So the error will be:

$$P(error) = \int_{0.75}^1 p(x|w_1)\,P(w1) + \int_{0.5}^{0.75} p(x|w_2)\,P(w2) =$$
$$= (0.5 - 0.375) + (0.375 - 0.25) = 0.25$$

The result in a) was the same as for b) part. This is due to the particular four random points I draw for the first classifier.
If I would generated random data from given distributions in MATLAB, the value of error will converge to 0.25 when number of drawn points $n \to \infty$. I also simulated error from the part a). Results are placed in Table 2. When n=1000000, error for both is very close to 0.25.

Table 2. Misclassification error for classifier from 4 random data points and from $n \to \infty$ data points.

Number of samples	**Error for 4 points classifier**	**Error for $n \to \infty$ points classifier**
100	0.2650	0.2600
10000	0.2539	0.2486
1000000	0.2498	0.2495

Problem No. 2: Consider the following models for a system that outputs sequences of the characters "$" and "%". For these models, you must start in state 1 and end in state 2.

(a) Compute the probability that model A produced the sequence "%$%".

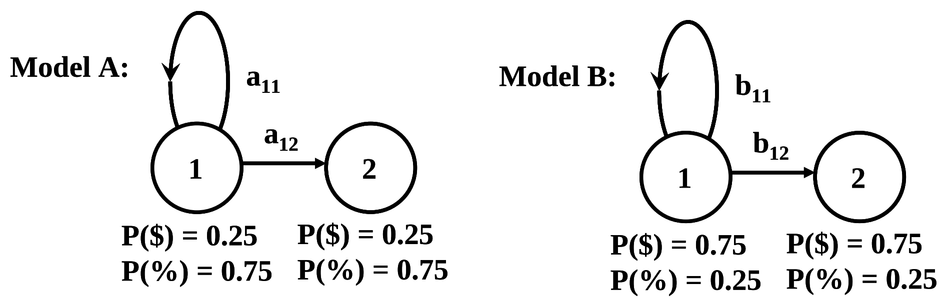

Fig.2 shows possible transitions for two models and their coefficients

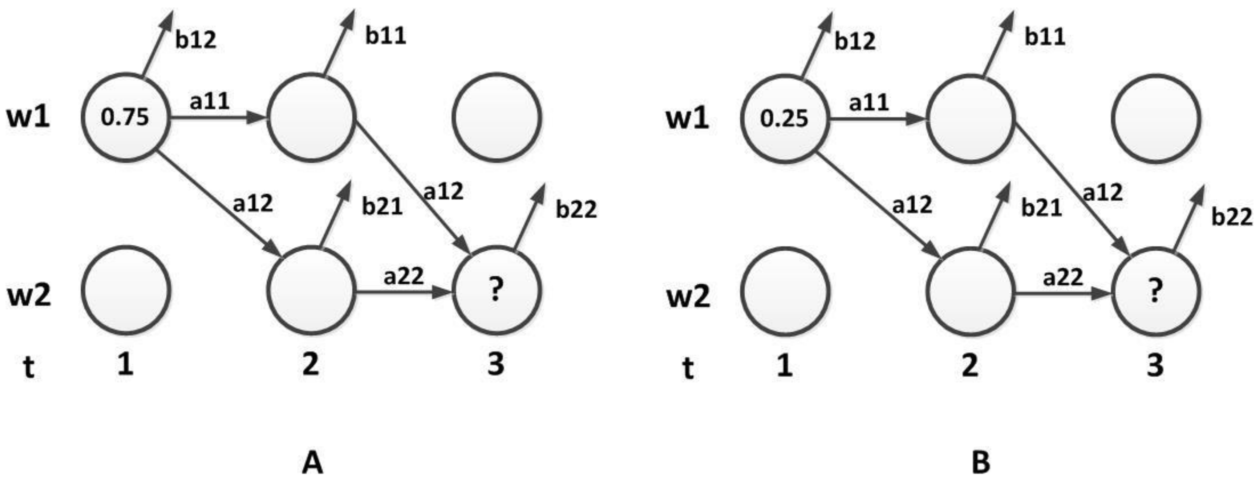

Figure 2. Models A and B of HMM

We want to find a probability of a "%$%" sequence for each model. Transition probabilities a are unknown, yet emission probabilities (b) are known. State 1 is the starting state, state 2 is the finishing state. Emission from state w1 of "$" is $b_{11.}$ Emission from state w1 of "%" is $b_{12.}$ Emission from state w1 of "$" is $b_{21.}$ Emission from state w2 of "%" is $b_{22.}$ Transition probability from state w1 to w1 is $a_{11.}$Transition probability from state w1 to w2 is $a_{12.}$ Transition probability from state w2 to w2 is $a_{22.}$

The total probability of the sequence for both models is $\alpha_2(3)$. Probabilities of each step are multiplied. Prior is π=[1,0], means the w1 is a start state. Calculation is done as following.

Model A	**Model B**
$P(A)_{w1w1w2} = \alpha_2(3)$	$P(B)_{w1w1w2} = \alpha_2(3)$
$= (\pi \cdot b_{12}) \cdot a_{11}b_{11} \cdot a_{12}b_{22}$	$= (\pi \cdot b_{12}) \cdot a_{11}b_{11} \cdot a_{12}b_{22}$
$= (\pi \cdot b_{12}) \cdot \alpha_1(1) \cdot \alpha_1(2)$	$= (\pi \cdot b_{12}) \cdot \alpha_1(1) \cdot \alpha_1(2)$
$= (1 \cdot 0.75) \cdot 0.25a_{11} \cdot 0.75a_{12}$	$= (1 \cdot 0.25) \cdot 0.75a_{11} \cdot 0.25a_{12}$
$= 0.1406\ a_{11}a_{12}$	$= 0.0469\ a_{11}a_{12}$

$P(A)_{w1w2w2} = \alpha_2(3)$	$P(B)_{w1w2w2} = \alpha_2(3)$
$= (\pi \cdot b_{12}) \cdot a_{12}b_{21} \cdot a_{22}b_{22}$	$= (\pi \cdot b_{12}) \cdot a_{12}b_{21} \cdot a_{22}b_{22}$
$= (\pi \cdot b_{12}) \cdot \alpha_1(1) \cdot \alpha_2(2)$	$= (\pi \cdot b_{12}) \cdot \alpha_1(1) \cdot \alpha_2(2)$
$= (1 \cdot 0.75) \cdot 0.25a_{11} \cdot 0.75a_{12}$	$= (1 \cdot 0.25) \cdot 0.75a_{11} \cdot 0.25a_{12}$
$= 0.1406\, a_{12}a_{22}$	$= 0.0469\, a_{12}a_{22}$
$(P(A)_{w1w1w2} + P(A)_{w1w2w2})$	$(P(B)_{w1w1w2} + P(B)_{w1w2w2})$
$= \mathbf{0.1406(a_{11}a_{12} + a_{12}a_{22})}$	$= \mathbf{0.0469(a_{11}a_{12} + a_{12}a_{22})}$

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

If we assume that transition probabilities are equal for Model A and Model B,

$$(a_{11}a_{12} + a_{12}a_{22})_{A\ Model} = (a_{11}a_{12} + a_{12}a_{22})_{B\ Model}\ .$$

From here we can conclude, that Model A will give the desired output of "%$%" with higher probability of $0.1406(a_{11}a_{12} + a_{12}a_{22})$ as calculated earlier. Otherwise, we will have to recalculate our probabilities considering new values of $a_{11}, a_{12},$ and a_{22} for each model.

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

There will be two possible state sequences (Fig. 2):

(w1 w1 w2) or (w1 w2 w2)

Only second step transition will be different for two combinations.

As we can see from the part a) calculations, the sequence of states does not change the probability of getting "%$%" from Model A or Model B (assuming transition probabilities equal).

If we consider the situation, when assigned transitions will be $a_{11}a_{12} > a_{12}a_{22}$, it will make the path w1w1w2 optimal. Or if assigned transitions will be $a_{11}a_{12} < a_{12}a_{22}$, it will make the path w1w2w2 optimal instead.

(d) Give at least two reasons why the probabilities in (a) and (c) differ.

1) Transition probabilities are equal between states
2) States are limited to start at the state 1 and to stop at the state 2.
3) We considering each path probabilities in (c) yet were interested in overall probability of the models in (a) part.