Expectation-Maximization for Naïve Bayes Use **PENCIL** to do this exercise. **First Iteration.** You have data from a Naïve Bayes model with $N = 2$ (so there are two binary features, A_1 and A_2, and one binary class variable C. The structure of the Bayes net is $A_1 \leftarrow C \rightarrow A_2$. data doesn't include class variable C

The data is as follows: $M = 10$ values for $(A_1, A_2) = (t,t), (t,f), (t,f), (f,f), (t,t), (f,t), (f,f), (t,t), (t,t), (t,t)$. Fill in the values given in class for the initial guess for the parameters of this model:

start out by guessing all values $\theta_c = P(C = t) = \boxed{0.6}$ $\theta_{11} = P(A_1 = t|C = t) = \boxed{0.9}$ $\theta_{10} = P(A_1 = t|C = f) = \boxed{0.3}$

$\theta_{21} = P(A_2 = t|C = t) = \boxed{0.2}$ $\theta_{20} = P(A_2 = t|C = f) = \boxed{0.6}$

First, you will caculate $P(C|A_1, A_2)$ for all possible values of A_1,A_2 and C. Do this by filling in the following table

1 A_1	2 A_2	3 C	4 $P(A_1\|C)$	5 $P(A_2\|C)$	6 $P(C)$	7 $P(C, A_1, A_2) = 4 \times 5 \times 6$	8 $P(C\|A_1, A_2)$ = normalise 7 over C
t	t	t	0.9	0.2	0.6	0.108	0.6
t	t	f	0.3	0.6	0.4	0.072	0.4
t	f	t	0.9	0.8	0.6	0.432	0.9
t	f	f	0.3	0.4	0.4	0.048	0.1
f	t	t	0.1	0.2	0.6	0.012	0.07
f	t	f	0.7	0.6	0.4	0.168	0.93
f	f	t	0.1	0.8	0.6	0.048	0.3
f	f	f	0.7	0.4	0.4	0.112	0.7

Now, "complete" the table of data below with the values in the table you just filled in.

A_1	A_2	C	$P(C, A_1, A_2)$	$P(C\|A_1, A_2)$
t	t	t	0.108	0.6
t	t	f	0.072	0.4
t	f	t	0.432	0.9
t	f	f	0.048	0.1
t	f	t	0.432	0.9
t	f	f	0.048	0.1
f	f	t	0.048	0.3
f	f	f	0.112	0.7
t	t	t	0.108	0.6
t	t	f	0.072	0.4
f	t	t	0.012	0.07
f	t	f	0.168	0.93
f	f	t	0.048	0.3
f	f	f	0.112	0.7
t	t	t	0.108	0.6
t	t	f	0.072	0.4
t	t	t	0.108	0.6
t	t	f	0.072	0.4
t	t	t	0.108	0.6
t	t	f	0.072	0.4
put sum of all rows here:			2.36	10

When this value's change is less than a threashold, you can stop iterating

When the sum of all rows in the second-to-last column stops changing, you can stop iterating. If $d_j = \{a_{j1}, a_{j2}\}$ is the data for $j = 1 \dots M$, then this is :

$$\sum_{j=1}^{M} P(d_j) = \sum_{j=1}^{M} \sum_{c} P(c, a_{j1}, a_{j2})$$

Now, you will re-estimate the parameters using the last column in the table on the last page (the normalised $P(C|A_1, A_2)$)
First, compute the sum of all the rows in the table above where $C = t$, and divide by the sum of all rows to get

$$\theta_C = \frac{\text{sum of all weights where } C = t}{\text{sum of all weights}} = \frac{5.47}{4.53 + 5.47} = \boxed{0.547}$$

Now compute the sum of all the rows in the table above where $C = t$ **AND** $A_1 = t$, to get

$$\theta_{11} = \frac{\text{sum of all weights where } C = t \text{ and } A_1 = t}{\text{sum of all weights where } C = t} = \frac{4.8}{5.47} = \boxed{0.88}$$

Now compute the sum of all the rows in the table above where $C = f$ **AND** $A_1 = t$, to get

$$\theta_{10} = \frac{\text{sum of all weights where } C = f \text{ and } A_1 = t}{\text{sum of all weights where } C = f} = \frac{2.2}{4.53} = \boxed{0.49}$$

Now compute the sum of all the rows in the table above where $C = t$ **AND** $A_2 = t$, to get

$$\theta_{21} = \frac{\text{sum of all weights where } C = t \text{ and } A_2 = t}{\text{sum of all weights where } C = t} = \frac{3.07}{5.47} = \boxed{0.56}$$

Now compute the sum of all the rows in the table above where $C = f$ **AND** $A_2 = t$, to get

$$\theta_{20} = \frac{\text{sum of all weights where } C = f \text{ and } A_2 = t}{\text{sum of all weights where } C = f} = \frac{2.93}{4.53} = \boxed{0.65}$$

Finally, copy the values you just calculated back to the first boxes on the last (the model parameters), and start again. Continue doing this until the sum of all the rows stops changing.

AFTER CONVERGENCE: Now copy your values for all values of θ from the other page here:

$$\theta_c = P(C = t) = \boxed{0.6} \quad \theta_{11} = P(A_1 = t|C = t) = \boxed{0.9} \quad \theta_{10} = P(A_1 = t|C = f) = \boxed{0.3}$$
$$\theta_{21} = P(A_2 = t|C = t) = \boxed{0.2} \quad \theta_{20} = P(A_2 = t|C = f) = \boxed{0.6}$$

A_1	A_2	$P(C = t|A_1, A_2)$	prediction
t	t		
t	f		
f	t		
f	f		

Compute $P(C = t|A_1, A_2)$ for each test data, write it in the table to the left,
and then threshold at 0.5 to make a prediction for the test data