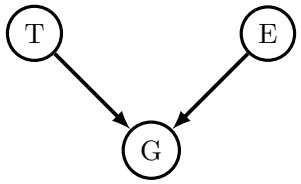
Quiz7

1 Learning in BN [50 pts]

The following Bayesian Network consists of three variables: T, E and G, which represent the Talent, Effort and Grade of a student respectively.



The possible values for each variable are as follows:

$$T \in \{\text{High}, \text{Medium}\}, E \in \{\text{High}, \text{Medium}, \text{Low}\}, G \in \{\text{A}, \text{B}, \text{C}, \text{F}\}.$$

(10pt) 1. How many parameters are required at least?

For variable T, we only need 1 parameter (since the sum of probabilities equals 1).
For variable E, we need 2 parameters.
For variable G, we need to specify the probability of G for each combination of T and E, so the number of parameters is 2 * 3 * 3.
Total parameters: 1 + 2 + 2 * 3* 3 = 21.

(15pt) 2. Given the training data, where each tuple is provided in the order of (T, E, G). Please estimate the conditional distributions for each variable: $P(T)$, $P(E)$, and $P(G)$. The training data is shown in the following table:

No.	T	E	G	No.	T	E	G
1	High	High	A	6	Medium	Medium	B
2	High	Medium	A	7	Medium	Low	C
3	High	Low	C	8	Medium	High	A
4	High	Medium	B	9	Medium	Low	F
5	High	Low	C	10	Medium	High	B

Hint: You can use tables to present your answer more clearly and write on both sides of the sheet.

Note: the question is to estimate the **conditional** distribution of each variable, so you need to estimate $P(G|T,E)$ actually.
But due to the potential ambiguity of $P(G)$ in the question, those who only write $P(G)$ will receive full points for this question, while those who write $P(G|T,E)$ will get bonus points. The total points will not exceed 100.

T	P_T
High	0.5
Medium	0.5

E	P_E
High	0.3
Medium	0.3
Low	0.4

G	T	E	$P_G(G\|T,E)$
A	High	High	1
A	High	Medium	0.5
A	Medium	High	0.5
B	High	Medium	0.5
B	Medium	High	0.5
B	Medium	Medium	1
C	High	Low	1
C	Medium	Low	0.5
F	Medium	Low	0.5

(10pt) 3. Use maximum likelihood estimation with Laplace smoothing to estimate the probability of a student get grade **A**, given that the student has **medium** talent and applies **medium** effort to study. Besides, give a brief explanation of why we need Laplace smoothing in probability estimation of Bayesian Network.
Assume that the amount of smoothing $\lambda = 1$.

$P(G = A|T = \text{Medium}, E = \text{Medium}) = \frac{0+1}{1+4} = \frac{1}{5}$
To avoid the zero probability problem in probability estimation. (Any reasonable answer is OK)
Zero probability problem: some events are not observed in the dataset, leading to the probability to be 0. This can be a disaster when we predict unseen data. Their probability will be predicted to be 0, which means this event will never happen.
To solve this problem, Laplace smoothing assumes that all events will occur one more time.

(15pt) 4. In unsupervised learning, we can use Expectation Maximization to estimate the maximum likelihood of parameters. In order to ease the calculation, we only focus on two variables: Talent and Grade. The training data is given as: $\mathcal{D}_{\text{train}} = \{(?, A), (?, B)\}$. Use the following conditional probability tables to write down one iteration of EM.

T	P_T
High	0.5
Medium	0.5

G	T	$P_G(G\|T)$
A	High	0.4
A	Medium	0.1
B	High	0.3
B	Medium	0.2
C	High	0.2
C	Medium	0.3
F	High	0.1
F	Medium	0.4

E-step (Since the training dataset does not include Grade C or F, we do not need to consider them.):

G	T	$P_G(G,T)$	$P(T\|G)$
A	High	0.5*0.4	$\frac{0.2}{0.2+0.05} = 0.8$
A	Medium	0.5*0.1	$\frac{0.05}{0.2+0.05} = 0.2$
B	High	0.5*0.3	$\frac{0.15}{0.15+0.1} = 0.6$
B	Medium	0.5*0.2	$\frac{0.1}{0.15+0.1} = 0.4$

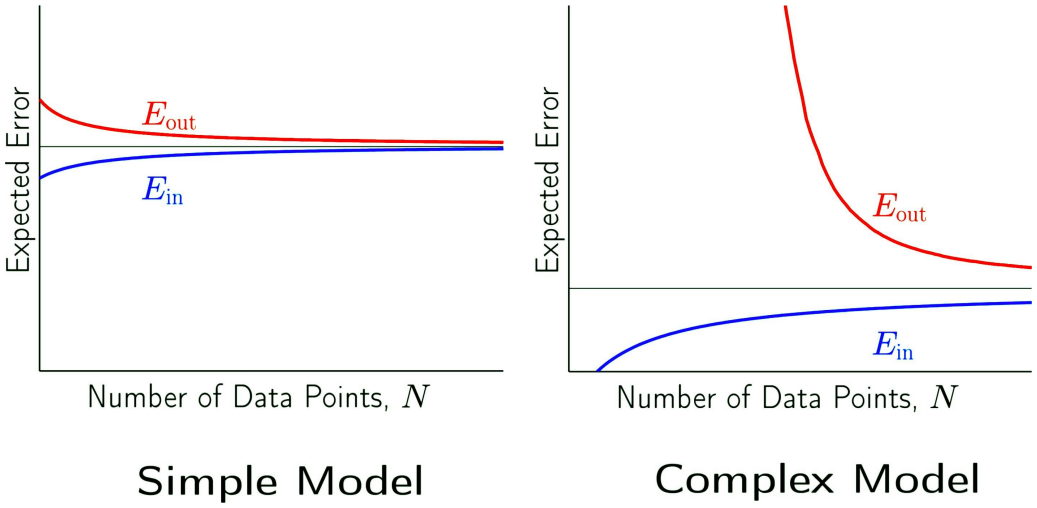
M-step:

T	P_T
High	$\frac{0.8+0.6}{0.8+0.6+0.2+0.4} = 0.7$
Medium	$\frac{0.2+0.4}{0.8+0.6+0.2+0.4} = 0.3$

G	T	$P_G(G\|T)$
A	High	$\frac{0.8}{0.8+0.6} = \frac{4}{7} = 0.57$
A	Medium	$\frac{0.2}{0.2+0.4} = \frac{1}{3} = 0.33$
B	High	$\frac{0.6}{0.8+0.6} = \frac{3}{7} = 0.43$
B	Medium	$\frac{0.3}{0.2+0.4} = \frac{1}{3} = 0.67$

Note:
purpose of E-step: given parameters (CPTs in this question), calculate the posterior distribution of hidden variables (T in this question).
purpose of M-step: maximize the likelihood function to update parameters (CPTs in this question).

2 Learning Curves [50 pts]



During class we provides an analysis of the above figure across six aspects. Please write down at least five of them.

refer to lecture.