CSE 150. Assignment 4 Solutions (*30 pts total*)

4.1 Node Clustering (*4 pts total, 1 pt per column*)

$$P(Y|X\!=\!0) \;=\; P(Y_1,Y_2,Y_3|X\!=\!0) \;=\; P(Y_1|X\!=\!0)\,P(Y_2|X\!=\!0)\,P(Y_3|X\!=\!0)$$
$$P(Y|X\!=\!1) \;=\; P(Y_1,Y_2,Y_3|X\!=\!1) \;=\; P(Y_1|X\!=\!1)\,P(Y_2|X\!=\!1)\,P(Y_3|X\!=\!1)$$

Y_1	Y_2	Y_3	Y	$P(Y\|X\!=\!0)$	$P(Y\|X\!=\!1)$	$P(Z_1\!=\!1\|Y)$	$P(Z_2\!=\!1\|Y)$
0	0	0	1	0.048	0.03	0.1	0.8
1	0	0	2	0.192	0.02	0.2	0.7
0	1	0	3	0.112	0.03	0.3	0.6
0	0	1	4	0.012	0.27	0.4	0.5
1	1	0	5	0.448	0.02	0.5	0.4
1	0	1	6	0.048	0.18	0.6	0.3
0	1	1	7	0.028	0.27	0.7	0.2
1	1	1	8	0.112	0.18	0.8	0.1

Note that $\sum_{y'} P(Y\!=\!y'|X\!=\!0) = \sum_{y'} P(Y\!=\!y'|X\!=\!1) = 1.$

4.2 Maximum likelihood estimation (*8 pts total, 2 pts per item*)

(a) For the DAG on the left:

$$P_{\text{ml}}(X\!=\!x) \;=\; \frac{\text{count}(X\!=\!x)}{\sum_{x'}\text{count}(X\!=\!x')} \;=\; \frac{C(x)}{\sum_{x'} C(x')} \;=\; \frac{C(x)}{T}$$
$$P_{\text{ml}}(Y\!=\!y|X\!=\!x) \;=\; \frac{\text{count}(X\!=\!x,Y\!=\!y)}{\text{count}(X\!=\!x)} \;=\; \frac{C(x,y)}{C(x)}$$
$$P_{\text{ml}}(Z\!=\!z|Y\!=\!y) \;=\; \frac{\text{count}(Y\!=\!y,Z\!=\!z)}{\text{count}(Y\!=\!y)} \;=\; \frac{C(y,z)}{C(y)}$$

(b) Likewise for the DAG on the right:

$$P_{\text{ml}}(Y\!=\!y) \;=\; \frac{\text{count}(Y\!=\!y)}{\sum_{y'}\text{count}(Y\!=\!y')} \;=\; \frac{C(y)}{\sum_{y'} C(y')} \;=\; \frac{C(y)}{T}$$
$$P_{\text{ml}}(X\!=\!x|Y\!=\!y) \;=\; \frac{\text{count}(X\!=\!x,Y\!=\!y)}{\text{count}(Y\!=\!y)} \;=\; \frac{C(x,y)}{C(y)}$$
$$P_{\text{ml}}(Z\!=\!z|Y\!=\!y) \;=\; \frac{\text{count}(Y\!=\!y,Z\!=\!z)}{\text{count}(Y\!=\!y)} \;=\; \frac{C(y,z)}{C(y)}$$

(c) The joint distribution from part (a) is given by:

$$\begin{aligned} P_{\mathrm{ml}}(X\!=\!x, Y\!=\!y, Z\!=\!z) &= P_{\mathrm{ml}}(X\!=\!x)\, P_{\mathrm{ml}}(Y\!=\!y|X\!=\!x)\, P_{\mathrm{ml}}(Z\!=\!z|Y\!=\!y) \\ &= \frac{C(x)}{T}\, \frac{C(x,y)}{C(x)}\, \frac{C(y,z)}{C(y)} \\ &= \frac{1}{T}\, \frac{C(x,y)\, C(y,z)}{C(y)} \end{aligned}$$

Likewise the joint distribution from part (b) is given by:

$$\begin{aligned} P_{\mathrm{ml}}(X\!=\!x, Y\!=\!y, Z\!=\!z) &= P_{\mathrm{ml}}(Y\!=\!y)\, P_{\mathrm{ml}}(X\!=\!x|Y\!=\!y)\, P_{\mathrm{ml}}(Z\!=\!z|Y\!=\!y) \\ &= \frac{C(y)}{T}\, \frac{C(x,y)}{C(y)}\, \frac{C(y,z)}{C(y)} \\ &= \frac{1}{T}\, \frac{C(x,y)\, C(y,z)}{C(y)} \end{aligned}$$

These joint probabilities are equal for all values of x, y, and z.

(d) The graphs imply the same relations of conditional independence. In particular, in both graphs X and Z are conditionally independent given Y.

4.3 Statistical language modeling (*18 pts total, 8 pts source code plus 2 pts per item*)

(a) The following are the frequencies (i.e., maximum likelihood estimates) for the words starting with the letter ‘A’:

Word	Frequency	Word	Frequency
A	0.018407	ACCORDING	0.000348
AND	0.017863	AIR	0.000311
AT	0.004313	ADMINISTRATION	0.000292
AS	0.003992	AGENCY	0.000280
AN	0.002999	AROUND	0.000277
ARE	0.002990	AGREEMENT	0.000263
ABOUT	0.001926	AVERAGE	0.000259
AFTER	0.001347	ASKED	0.000258
ALSO	0.001310	ALREADY	0.000249
ALL	0.001182	AREA	0.000231
A.	0.001026	ANALYSTS	0.000226
ANY	0.000632	ANNOUNCED	0.000227
AMERICAN	0.000612	ADDED	0.000221
AGAINST	0.000596	ALTHOUGH	0.000214
ANOTHER	0.000428	AGREED	0.000212
AMONG	0.000374	APRIL	0.000207
AGO	0.000357	AWAY	0.000202

(b) The following are the frequencies (maximum likelihood estimates) for top-10 most likely words following “THE”:

Word	Frequency
<UNK>	0.615020
U.	0.013372
FIRST	0.011720
COMPANY	0.011659
NEW	0.009451
UNITED	0.008672
GOVERNMENT	0.006803
NINETEEN	0.006651
SAME	0.006287
TWO	0.006161

(c) The sentence is "THE STOCK MARKET FELL BY ONE HUNDRED POINTS LAST WEEK". Let $\mathcal{L}_u$ and $\mathcal{L}_b$ denote, respectively, the log-likelihood of the unigram and bigram models for this sentence. Then we have:

$$\mathcal{L}_u = -64.509440$$

$$\mathcal{L}_b = -40.918132$$

Since $\mathcal{L}_u < \mathcal{L}_b$, the bigram model yields the highest log-likelihood.

(d) The sentence is "THE SIXTEEN OFFICIALS SOLD FIRE INSURANCE". The log-likelihoods are:

$$\begin{aligned} \mathcal{L}_u &= -44.291934 \\ \mathcal{L}_b &= -\infty \end{aligned}$$

Since $\mathcal{L}_u > \mathcal{L}_b$, the unigram model yields highest log-likelihood. And we have the following probabilities for each bigram in the sentence:

$$P_b(\text{THE}|\text{<s>}) = 0.158653$$

$$P_b(\text{SIXTEEN}|\text{THE}) = 0.000229$$

$$P_b(\text{OFFICIALS}|\text{SIXTEEN}) = 0$$

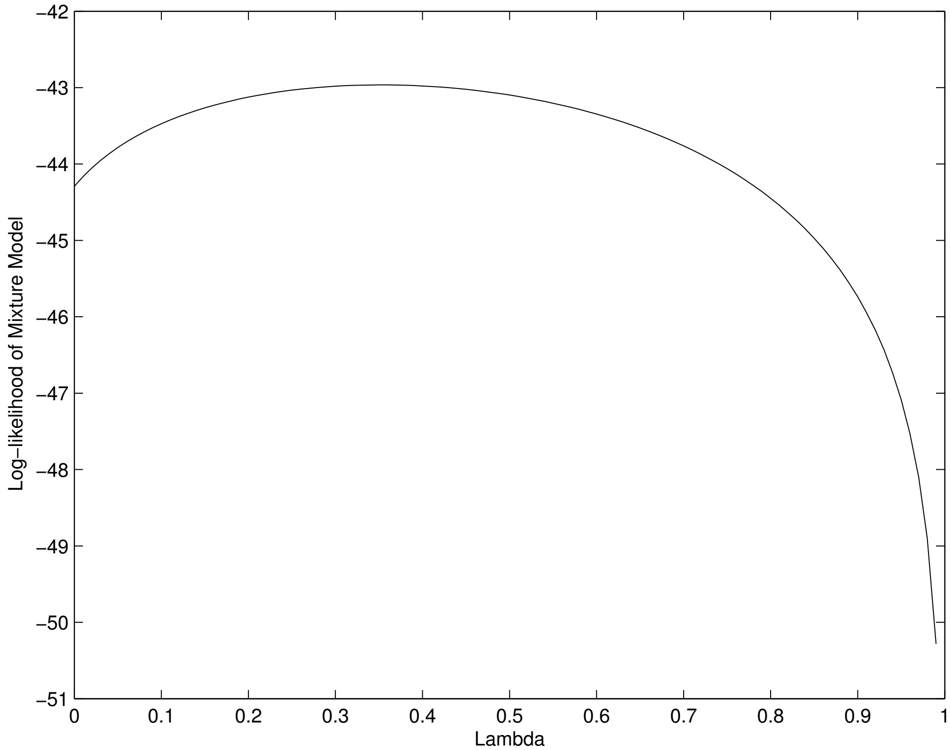
$$P_b(\text{SOLD}|\text{OFFICIALS}) = 0.000092$$

$$P_b(\text{FIRE}|\text{SOLD}) = 0$$

$$P_b(\text{INSURANCE}|\text{FIRE}) = 0.003052$$

The bigrams "SIXTEEN OFFICIALS" and "SOLD FIRE" are not observed in the training set. This causes the log-likelihood for the bigram model to be undefined.

(e) The figure shows the log-likelihood $\mathcal{L}_m$ for $\lambda \in [0, 1]$. The optimal value is $\lambda = 0.35$, which yields a log-likelihood of $\mathcal{L}_m = -42.96$.



(f) Source code