COMP 424 A4 Solution

Question 1. (POS tagging).

(a) Let C = "conjunction", N = "noun", V = "verb", A = "adjective".

then the smoothed inital probability would be:

$$\pi = \begin{pmatrix} \pi_C \\ \pi_N \\ \pi_V \\ \pi_A \end{pmatrix} = \begin{pmatrix} 2+1/4+4 \\ 0+1/4+4 \\ 2+1/4+4 \\ 0+1/4+4 \end{pmatrix} = \begin{pmatrix} 3/8 \\ 1/8 \\ 3/8 \\ 1/8 \end{pmatrix}$$

The smoothed transition matrix:

$$A = \begin{pmatrix} a_{CC} & a_{CN} & a_{CV} & a_{CJ} \\ a_{NC} & a_{NN} & a_{NV} & a_{NJ} \\ a_{VC} & a_{VN} & a_{VV} & a_{VJ} \\ a_{JC} & a_{JN} & a_{JV} & a_{JJ} \end{pmatrix} = \begin{pmatrix} \frac{0+1}{2+4} & \frac{2+1}{2+4} & \frac{0+1}{2+4} & \frac{0+1}{2+4} \\ \frac{0+1}{5+4} & \frac{1+1}{5+4} & \frac{3+1}{5+4} & \frac{1+1}{5+4} \\ \frac{0+1}{6+4} & \frac{3+1}{6+4} & \frac{3+1}{6+4} & \frac{0+1}{6+4} \\ \frac{0+1}{0+4} & \frac{0+1}{0+4} & \frac{0+1}{0+4} & \frac{0+1}{0+4} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{6} & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{9} & \frac{2}{9} & \frac{4}{9} & \frac{2}{9} \\ \frac{1}{10} & \frac{2}{5} & \frac{2}{5} & \frac{1}{10} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

The smoothed emission matrix:

$$B = \begin{pmatrix} b^C_{that} & b^C_{is} & b^C_{not} & b^C_{it} & b^C_{good} \\ b^N_{that} & b^N_{is} & b^N_{not} & b^N_{it} & b^N_{good} \\ b^V_{that} & b^V_{is} & b^V_{not} & b^V_{it} & b^V_{good} \\ b^J_{that} & b^J_{is} & b^J_{not} & b^J_{it} & b^J_{good} \end{pmatrix} = \begin{pmatrix} \frac{2+1}{2+5} & \frac{0+1}{2+5} & \frac{0+1}{2+5} & \frac{0+1}{2+5} & \frac{0+1}{2+5} \\ \frac{3+1}{6+5} & \frac{0+1}{6+5} & \frac{1+1}{6+5} & \frac{2+1}{6+5} & \frac{0+1}{6+5} \\ \frac{1+1}{8+5} & \frac{6+1}{8+5} & \frac{1+1}{8+5} & \frac{0+1}{8+5} & \frac{0+1}{8+5} \\ \frac{0+1}{1+5} & \frac{0+1}{1+5} & \frac{0+1}{1+5} & \frac{0+1}{1+5} & \frac{1+1}{1+5} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{3}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{4}{11} & \frac{1}{11} & \frac{2}{11} & \frac{3}{11} & \frac{1}{11} \\ \frac{2}{13} & \frac{7}{13} & \frac{2}{13} & \frac{1}{13} & \frac{1}{13} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} \end{pmatrix}$$

(b) Run the forward algorithm to fill those trellis:

	not	that	good
N	(1/8)*(2/11) = 0.0227	0.0227*(2/9)*(4/11) + 0.0578*(2/5)*(4/11) + 0.0536*(0.5)*(4/11) + 0.0208*(0.25)*(4/11) =0.0219	0.0219*(2/9)*(1/11) + 0.00728*(2/5)*(1/11) + 0.00962*(0.5)*(1/11) + 0.00416*(0.25)*(1/11) = 0.00124
V	(3/8)*(2/13) = 0.0578	0.0227*(4/9)*(2/13) + 0.0578*(2/5)*(2/13) + 0.0536*(1/6)*(2/13) + 0.0208*(0.25)*(2/13) =0.00728	0.0219*(4/9)*(1/13) + 0.00728*(2/5)*(1/13) + 0.00962*(1/6)*(1/13) + 0.00416*(0.25)*(1/13) = 0.00118
C	(3/8)*(1/7) = 0.0536	0.0227*(1/9)*(3/7) + 0.0578*(1/10)*(3/7) + 0.0536*(1/6)*(3/7) + 0.0208*(0.25)*(3/7) = 0.00962	0.0219*(1/9)*(1/7) + 0.00728*(1/10)*(1/7) + 0.00962*(1/6)*(1/7) + 0.00416*(0.25)*(1/7) = 0.000829
J	(1/8)*(1/6) = 0.0208	0.0227*(2/9)*(1/6) + 0.0578*(1/10)*(1/6) + 0.0536*(1/6)*(1/6) + 0.0208*(0.25)*(1/6) = 0.00416	0.0219*(2/9)*(1/3) + 0.00728*(1/10)*(1/3) + 0.00962*(1/6)*(1/3) + 0.00416*(0.25)*(1/3) = 0.00275

$\Rightarrow P(\text{not that good}) = 0.00124 + 0.00118 + 0.000829$
$+ 0.00275 = 0.006$

(c). Following (b) we know:

$$P(z_4 = \text{noun} \mid x_{1:3} = \text{"not that good"})$$

$$= \frac{\sum_i P(z_3 = i, x_{1:3} = \text{"not that good"})\, P(z_4 = \text{noun} \mid z_3 = i)}{P(x_{1:3} = \text{"not that good"})}$$

$$= \frac{0.00124 \times \frac{2}{9} + 0.00118 \times \frac{2}{5} + 0.000829 \times \frac{1}{2} + 0.00275 \times \frac{1}{4}}{0.006}$$

$$= 0.308$$

(d)

	not	that	good
N	(1/8)*(2/11) = 0.0227	max{0.0227*(2/9)*(4/11) 0.0578*(2/5)*(4/11) 0.0536*(0.5)*(4/11) 0.0208*(0.25)*(4/11)} = 0.00975	max{ 0.00975*(2/9)*(1/11) 0.00357*(2/5)*(1/11) 0.00383*(0.5)*(1/11) 0.00149*(0.25)*(1/11)} =0.000197
V	(3/8)*(2/13) = 0.0578	max{ 0.0227*(4/9)*(2/13) 0.0578*(2/5)*(2/13) 0.0536*(1/6)*(2/13) 0.0208*(0.25)*(2/13)} =0.00357	max{ 0.00975*(4/9)*(1/13) 0.00357*(2/5)*(1/13) 0.00383*(1/6)*(1/13) 0.00149*(0.25)*(1/13)} = 0.000333
C	(3/8)*(1/7) = 0.0536	max{ 0.0227*(1/9)*(3/7) 0.0578*(1/10)*(3/7) 0.0536*(1/6)*(3/7) 0.0208*(0.25)*(3/7)} = 0.00383	max{ 0.00975*(1/9)*(1/7) 0.00357*(1/10)*(1/7) 0.00383*(1/6)*(1/7) 0.00149*(0.25)*(1/7)} =0.000155
J	(1/8)*(1/6) = 0.0208	max{ 0.0227*(2/9)*(1/6) 0.0578*(1/10)*(1/6) 0.0536*(1/6)*(1/6) 0.0208*(0.25)*(1/6)} = 0.00149	max{ 0.00975*(2/9)*(1/3) 0.00357*(1/10)*(1/3) 0.00383*(1/6)*(1/3) 0.00149*(0.25)*(1/3)} =0.000722

⇒ The POS tags are:

Not (conjunction) that (noun) good (adjective).

Q2.

(a). $P(A) = (0.05\times0.75\times0.85) + (0.1\times0.75\times0.15) + (0.1\times0.25\times0.85)$

$+ (0.75\times0.25\times0.15) = 0.0925$

$EU(\text{with insurance}) = -100\times0.0925 - 20\times(1-0.0925)$

$= -27.4$

$EU(\text{without insurance}) = -500\times0.0925 = -46.25$

$\Rightarrow \text{price} = -27.4 + 46.25 = 18.85$

(b) $Pr(A=1 \mid S=1, D=0) = 0.1$

$Pr(A=0 \mid S=1, D=0) = 0.9$

$EU(\text{with insurance}) = -100\times0.1 - 20\times0.9 = -28$

$EU(\text{without insurance}) = -500\times0.1 = -50$

$\Rightarrow \text{price} = -28 + 50 = 22$

(c). $Pr(A=1 \mid D=1) = Pr(A=1 \mid D=1, S=0)\cdot Pr(S=0)$

$+ Pr(A=1 \mid D=1, S=1)\cdot Pr(S=1)$

$= 0.1\times0.75 + 0.75\times0.25 = 0.2625$

$Pr(A=0|D=1) = 0.7375$

$EU(\text{with insurance}) = -100 \times 0.2625 - 20 \times 0.7375 = -41$

$EU(\text{without} \sim) = -500 \times 0.2625 = -131.25$

$\Rightarrow$ price $= -41 + 131.25 = 90.25$

(d) $EU(\text{with new insurance}) = -100 \times 0.0925 \times 0.9$

$- 500 \times 0.0925 \times 0.1 - 20 \times (1 - 0.0925)$

$= -31.1$

$\Delta EU = -27.4 + 31.1 = 3.7$

$\Rightarrow$ new price should be $18.85 - 3.7 = 15.15$

Q3.

(a). $N = 6^6$ (3^6 also makes sense).

(b). For the initall policy π_0, the transition matrix would be:
(Salad = S, yogurt = Y, hamburger = H, Burrito = B, Pizza = Z, poutine = P)

$$T^{\pi_0} = \begin{array}{c} \\ S \\ Y \\ H \\ B \\ Z \\ P \end{array} \begin{array}{c} \begin{array}{cccccc} S & Y & H & B & Z & P \end{array} \\ \begin{pmatrix} 0.2 & 0.8 & 0 & 0 & 0 & 0 \\ 0 & 0.2 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0.8 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$R^{\pi_0} = \begin{array}{c} S \\ Y \\ H \\ B \\ Z \\ P \end{array} \begin{pmatrix} 40 \\ 30 \\ 20 \\ 10 \\ 0 \\ 0 \end{pmatrix}.$$

$$(I - \gamma T^{\pi_0}) V^{\pi_0} = R^{\pi_0}$$

therefore, by Bellman equation:

$$V^{\pi_0} = (I - \gamma T^{\pi_0})^{-1} R^{\pi_0} = \begin{pmatrix} 107.96 \\ 67.40 \\ 35.10 \\ 12.20 \\ 0 \\ 0 \end{pmatrix} \begin{array}{c} S \\ Y \\ H \\ B \\ Z \\ P \end{array}$$

(c).

$$\pi' = \begin{array}{c} S \\ Y \\ H \\ B \\ Z \\ P \end{array} \begin{pmatrix} S \\ S \\ Y \\ H \\ B \\ Z \end{pmatrix}$$

(d) After one iteration, we'll reach the optimal policy.

$$R^{\pi_4} = \begin{pmatrix} 50 \\ 50 \\ 40 \\ 30 \\ 20 \\ 10 \end{pmatrix} = R^{\pi^*}$$

$$\Rightarrow V^* = (I - \gamma T^{\pi^*})^{-1} R^{\pi^*}$$

$$= \begin{pmatrix} 500 \\ 500 \\ 487.80 \\ 464.90 \\ 432.60 \\ 392.04 \end{pmatrix} \begin{array}{c} S \\ Y \\ H \\ B \\ Z \\ P. \end{array}$$

(e) V^* is unique in this case, since $V^*(s)$ is the unique solution for the system of non-linear Bellman equations: $V^*(s) = \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s,a,s')V^*(s')\right)$

f).

$$\pi^* = \begin{array}{c} S \\ Y \\ H \\ B \\ Z \\ P \end{array} \begin{pmatrix} S \\ S \\ Y \\ H \\ B \\ Z \end{pmatrix}$$

(g) Here both "yes" and "no" would make sense: since there's only one unique deterministic optimal policy, while there're still many non-deterministic policies that yield the same V^*.

(h). By the matrix-form Bellman equation:

$$V^{(t)} = (I - \gamma T^{\pi_t})^{-1} R^{\pi_t}$$

If we divide all health scores by 10, we'll reduce both $V^{(t)}$ and R^{π_t} into $\frac{1}{10}$ of their values. Therefore, V^* would also be $\frac{1}{10}$ of the original values, while each old π^* is still a valid optimal policy in this case.

Q4.

(a).

	t	R_t	A_t	Q_t
	0	/	/	$\{1, 2, 2, 1, 0, 3\}$
r	1	2	1	$\{2, 2, 2, 1, 0, 3\}$
r	2	$\sqrt{3}$	2	$\{2, \sqrt{3}, 2, 1, 0, 3\}$
r	3	1	3	$\{2, \sqrt{3}, 1, 1, 0, 3\}$
r	4	0	4	$\{2, \sqrt{3}, 1, 0, 0, 3\}$
r	5	-1	5	$\{2, \sqrt{3}, 1, 0, -1, 3\}$
	6	$-\sqrt{3}$	6	$\{2, \sqrt{3}, 1, 0, -1, -\sqrt{3}\}$
	7	-2	1	$\{0, \sqrt{3}, 1, 0, -1, -\sqrt{3}\}$
	8	$-\sqrt{3}$	2	$\{0, 0, 1, 0, -1, -\sqrt{3}\}$
	9	-1	3	$\{0, 0, 0, 0, -1, -\sqrt{3}\}$
	10	0	4	$\{0, 0, 0, 0, -1, -\sqrt{3}\}$
r	11	1	5	$\{0, 0, 0, 0, 0, -\sqrt{3}\}$
r	12	$\sqrt{3}$	6	$\{0, 0, 0, 0, 0, 0\}$.

(b)

At $t = 1, 2, 3, 4, 5, 11, 12$, the agent must have selected a random action.

(c) For each state $i \in A$, the observed sequence of rewards would be: $2\cos\left[\frac{\pi}{6}(i-1)\right]$, $-2\cos\left[\frac{\pi}{6}(i-1)\right]$, $2\cos\left[\frac{\pi}{6}(i-1)\right]$, ... Hence, the estimated expected reward would be:

$$2\cos\left[\frac{\pi}{6}(i-1)\right],\ 0,\ \frac{2\cos\left[\frac{\pi}{6}(i-1)\right]}{3},\ 0,\ \frac{2\cos\left[\frac{\pi}{6}(i-1)\right]}{5},\ \cdots$$

so $\forall i \in A$, $Q^*(i)$ will converge to 0.