CSE 325/425 (Spring 2021) Homework 2

Due on 11:55pm, Mar 3, 2021

Grading: All questions have the same points (25 each). We will randomly grade some of the questions.

Submitting: Only electronic submissions on Coursesite are accepted. You can handwrite your answers on papers and then scan them to images. If you need to plot figures using a computer, the plotted files should be saved and included in the submitted pdf file. Submit a single pdf file named

<Your LIN>HW2.pdf

Other format will not be accepted.

Questions:

1. Given four POS tags *Noun, Determiner, Verb, Adjective*, construct a valid example of transition probability matrix A for an HMM that use these tags as hidden states. Pay attention to the shape of the matrix and any constraint that the elements in the matrix should satisfy.

   [[[The matrix is of shape 4×4, and each row sums to 1. An example is as follows: <==

$$\begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix} \tag{1}$$

   ]]]

2. Given observed sentence $O = [o_1, o_2, o_3]$ and a given POS tag sequence $Q = [q_1, q_2, q_3]$, use the forward algorithm to expand the probability $\Pr([o_1, o_2, o_3])$ in terms of the elements in the HMM parameters A, B, and π. You need to write down each step when the recursive equation in the algorithm is used.

   [[[The forward probability is $\alpha_t(i) = \Pr(o_1, \ldots, o_t, q_t = i)$. <==

$$
\begin{aligned}
\Pr([o_1, o_2, o_3]) &= \sum_{q_3}\sum_{q_2}\sum_{q_1} \Pr(o_1, o_2, o_3, q_1, q_2, q_3) && (2)\\
&= \sum_{q_3}\sum_{q_2}\sum_{q_1} \Pr(o_1, q_1)\Pr(o_2, q_2|q_1, o_1)\Pr(o_3, q_3|q_1, q_2, o_1, o_2) && (3)\\
&= \sum_{q_3}\sum_{q_2}\sum_{q_1} \Pr(o_1|q_1)\Pr(q_1)\Pr(o_2|q_2)\Pr(q_2|q_1)\Pr(o_3|q_3)\Pr(q_3|q_2) && (4)\\
&= \sum_{q_3}\left[\sum_{q_2}\left[\sum_{q_1}\left[\Pr(o_1|q_1)\Pr(q_1)\right]\Pr(o_2|q_2)\Pr(q_2|q_1)\right]\Pr(o_3|q_3)\Pr(q_3|q_2)\right] && (5)\\
&= \sum_{q_3}\left[\sum_{q_2}\left[\sum_{q_1}\alpha_1(q_1)\Pr(o_2|q_2)\Pr(q_2|q_1)\right]\Pr(o_3|q_3)\Pr(q_3|q_2)\right] && (6)\\
&= \sum_{q_3}\left[\sum_{q_2}\alpha_2(q_2)\Pr(o_3|q_3)\Pr(q_3|q_2)\right] && (7)\\
&= \sum_{q_3}\alpha_3(q_3). && (8)
\end{aligned}
$$

Eq. (2) uses the definition of marginal distribution; Eq. (3) uses the definition of conditional probabilities; Eq. (4) uses the HMM assumption; Eq. (5) uses the distribitivity of summation of products and pulls out the common factors; The remaining equations use the definition of the forward probabilities.]]]

3. Re-do the previous question with the same requirements, but use the backward algorithm.

   [[[<==

$$
\begin{aligned}
\Pr([o_1, o_2, o_3]) &= \sum_{q_1}\sum_{q_2}\sum_{q_3} \Pr(o_1|q_1)\Pr(q_1)\Pr(o_2|q_2)\Pr(q_2|q_1)\Pr(o_3|q_3)\Pr(q_3|q_2) && (9)\\
&= \sum_{q_1}\sum_{q_2} \Pr(o_1|q_1)\Pr(q_1)\Pr(o_2|q_2)\Pr(q_2|q_1)\left[\sum_{q_3}\Pr(o_3|q_3)\Pr(q_3|q_2)\right] && (10)\\
&= \sum_{q_1}\left[\sum_{q_2}\Pr(o_1|q_1)\Pr(q_1)\Pr(o_2|q_2)\Pr(q_2|q_1)\beta_2(q_2)\right] && (11)\\
&= \sum_{q_1}\Pr(o_1|q_1)\Pr(q_1)\left[\sum_{q_2}\Pr(o_2|q_2)\Pr(q_2|q_1)\beta_2(q_2)\right] && (12)\\
&= \sum_{q_1}\Pr(o_1|q_1)\Pr(q_1)\beta_1(q_1) && (13)
\end{aligned}
$$

   ]]]

4. Run the Viterbi algorithm to predict the best POS-tag sequence Q^* given the input $O = [o_1, o_2, o_3]$ with the four POS tags *Noun, Determiner, Verb, Adjective.* You need to draw a trellis, with the v values (based on Viterbi) annotating the correct nodes. Then write down the process of computing $v_1(Noun)$ and $v_3(Verb)$ using the Viterbi algorithm.

   [[[The trellis is shown in Figure 1. <==

$$
v_1(Noun) = \Pr(Noun)\Pr(o_1|Noun) \qquad (14)
$$

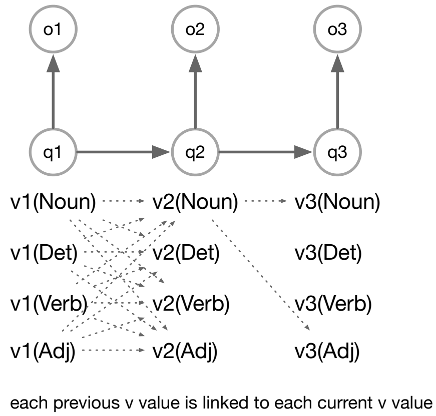

Figure 1: Trellis for computing v_t.

$$
\begin{aligned}
v_3(Verb) &= \max_{q_2} v_2(q_2)\Pr(q_2|q_1)\Pr(o_2|q_2) && (15)\\
&= \max_{q_2}\left[\max_{q_1} v_1(q_1)\Pr(q_1)\Pr(o_1|q_1)\right]\Pr(q_2|q_1)\Pr(o_2|q_2) && (16)\\
&= \max_{q_2}\left[\max_{q_1}\left[\Pr(q_1)\Pr(o_1|q_1)\right]\Pr(q_1)\Pr(o_1|q_1)\right]\Pr(q_2|q_1)\Pr(o_2|q_2) && (17)\\
& && (18)
\end{aligned}
$$

]]]

5. Run the forward and backward algorithms to compute the necessary probabilities to find $\xi_2(Noun, Verb)$ on the input training sequence $O = [o_1, o_2, o_3]$.

[[[<==

$$
\begin{aligned}
\alpha_2(Noun) &= \sum_{q_1} \alpha_1(q_1)\Pr(Noun|q_1)\Pr(o_2|Noun) && (19)\\
&= \sum_{q_1} \Pr(q_1)\Pr(o_1|q_1)\Pr(Noun|q_1)\Pr(o_2|Noun). && (20)
\end{aligned}
$$

$$\beta_3(Verb) = 1. \quad (21)$$

$$\xi_2(Noun, Verb) = \alpha_2(Noun)\Pr(Verb|Noun)\Pr(o_3|Verb)\beta_3(Verb) \quad (22)$$

]]]