

# CSE 325/425 (Spring 2021) Homework 6

Due on 11:55pm, May 7, 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>HW6.pdf

Other format will not be accepted.

## Questions:

1. Explain what the following equation is calculating in machine translation in plain English.

$$P(F|E) = \sum_A P(F, A|E), \quad (1)$$

$E$  is the target translation and  $F$  is a sentence in a foreign language to be translated.  $A$  is an alignment between  $E$  and  $F$ . (*Hint: explain each probability and then explain what the summation is doing.*)

2. Use the alignments in Figure 1 and the translation probabilities in Table 1 to run the E-step of the EM algorithm of training the simplified translation model

$$P(F, A|E) = \prod_{j=1}^J P(f_j|e_j). \quad (2)$$

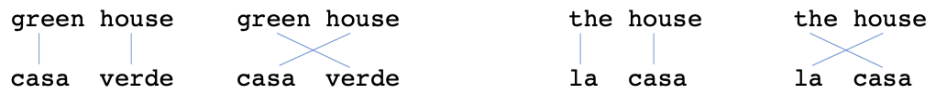


Figure 1: Two pairs of  $E$  and  $F$ . Each pair has two alignments indicated by the links between the words.

$P(\text{casa} \mid \text{green}) = 0.5$	$P(\text{verde} \mid \text{green}) = 0.5$	$P(\text{la} \mid \text{green}) = 0$
$P(\text{casa} \mid \text{house}) = 0.5$	$P(\text{verde} \mid \text{house}) = 0.25$	$P(\text{la} \mid \text{house}) = 0.25$
$P(\text{casa} \mid \text{the}) = 0.5$	$P(\text{verde} \mid \text{the}) = 0$	$P(\text{la} \mid \text{the}) = 0.5$

Table 1: Translation probabilities  $P(f|e)$

- Based on the above E-step, re-estimate the translation probability using the M-step.
- Find the alignment matrix for the following translation. The rows represent words in  $E$  and columns represent words in  $F$ .

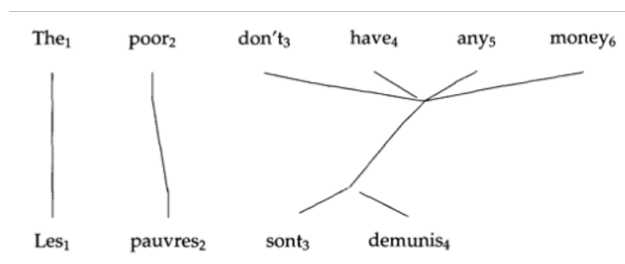


Figure 2: A translation of a French sentence into an English sentence

- Calculate the precision of candidates 1 and 2 using bigram and circle the matched bi-grams in each candidate. The matches are case-insensitive.

**Cand 1:** It is a guide to action which ensures that the military always obeys the commands of the party

**Cand 2:** It is to insure the troops forever hearing the activity guidebook that party direct

**Ref 1:** It is a guide to action that ensures that the military will forever heed Party commands

**Ref 2:** It is the guiding principle which guarantees the military forces always being under the command of the Party

**Ref 3:** It is the practical guide for the army always to heed the directions of the party

Figure 3: Example translations and ground truth for evaluation.