Comp550 Homework 2

Yue Violet Guo

(260606976)

1 Proofs

1.1 Q1

True. Proof: Base case:

When $t = 1$, $\delta_j(1) = P(O_1 = k, Q_1 = j|\theta)$

Since we use MLE parameters $\pi(i)$ and $b_j(O_1)$, $\delta_j(1)$ is the maximum for all observations.

Induction:

At $t = k$, suppose base case is true and $k \geq 1$,

$\delta_j(t) = max_i \delta_i(t-1) a_{ij} b_j(O_t)$

Since we assume base case holds, for $\delta_i(t-1)$, we have $Q_{1:t-1}$ that maximizes $\delta_i(t-1)$. Now, at $t = 1$, by definition of the algorithm, we choose MLE parameters, e.g. the a_{ij} is the maximum probability of state transitions, and $b_j(O_t)$ is the maximum probability of observation at the current time step. Therefore, at t, we have the maximum $\delta_j(t)$ possible. □

1.2 Q2

1. Gold standard: It is implied by $-\sum_i \sum_t log(p_t^i)$ that each label is independent since it does not condition on any other label.

2. LCRF

We have

$$-\sum_i \log P(Y^{(i)}|X^{(i)})$$
$$=\sum_i \log P(y_1^{(i)}, ..., y_t^{(i)}|x_1^{(i)}, ..., x_t^{(i)})$$

If we want to equate the two, we must have that in LCRF, we need the labels to be all independent given only the input at the current timestamp of the label. However, this is not true since LCRF uses features $f_k(y_t, y_{t-1}, x_t)$ □

2 French grammar

1. The advantages of using CFG are that we can parse multiple constituents, and use subcatgorization to represent many complex rules of French. Compared to finite state machines, this is more interpretable and scalable. For example, regex can model regular verb conjugations using common suffixes, but cannot handle irregular verbs, such as avoir.

2. The disadvantages are that CFG cannot model a complete set of French grammar. It also doesn't have a universal representation, since different linguist can come up with different subcategorizations of different terminals and non-terminals.

3. Some aspects of French grammar that my CFG cannot handle are pronoun and noun gender agreement. For example, the pronoun I/je itself doesn't carry a default, explicit gender like "elle" vs "il", but if the person speaking is female, she would say "Je suis étudiante", not "Je suis étudiant." My CFG does not handle this aspect of gender.

Another aspect that my CFG cannot handle is proper names that can be preceded by a determiner or an honorific. This CFG parses example 12 "Jonathan", but rejects "Monsieur Jonathan." It also rejects "le Canada."

The CFG cannot properly parse a noun phrase that has more than one adjective, where the ordering of the adjectives can be complicated.

The CFG does not subcategorize pronominal verbs under verbs, although we model direct object pronouns.

The CFG also does not handle any prepositions, such as "en", "au", "à", which vary with lexical meaning, gender, etc.

The CFG also cannot handle more complicated sentence structures, such as subordinate clauses because we do not have a rule that models conjunctions and conjunctive phrases, such as "lorsque ", "parce que", etc.

The CFG cannot handle negations with verbs.

The CFG cannot handle adverbs.

And of course, this CFG cannot handle other verb tenses (past, future) that the assignment outline excludes.

3 HMM

Table 1: Results

Model	cipher 1	cipher 2	cipher 3
basic	9.87	14.97	21.29
laplace	97.66	83.11	21.3
lm	6.58	5.36	7.87
lm, laplace	4.32	7.61	10.21

The basic HMM was within expectation since we have a small corpus, and tests contain unseen examples. We

achieve close to perfect accuracy with HMM and Laplacian smoothing. The speculation is that the Laplacian smoothing gives unseen examples a reasonably low probability so that the model doesn't assign a zero probability to any unseen example. The unexpected results are the experiments with an extra corpus. We did not have any significant increase. The reason is that the extra corpus is unlabelled. In code breaking, Vernam cipher can only be decrypted by exhaustive search in all possible key combinations, not natural frequencies of transition in the English language. Also, the transition counts of our cipher data may have a different distribution from the extra corpus, which may impact performance on test set of cipher data.