

第2題

a小題

$\alpha_t(i)$ 是給定 model λ ，看到 observation $o_1 o_2 \dots o_t$ 且在時間 t 時的 state 為 i 的機率。 $\beta_t(i)$ 是給定 model λ ，時間 t 時在 state i 的機率，而 $\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)} = \frac{P(\hat{O}, q_t=i|\lambda)}{P(\hat{O}|\lambda)}$ 故 $\gamma_t(i)$ 是給定一 observation sequence $o_1 o_2 \dots o_N$ 時間 t 時 state 為 i 的機率。

b小題

定義一新變數 $\delta_t(i)$ 其值為時間 t 時到達 state i 的單一路徑之中最高的機率。則 $\delta_t(i)$ 數學定義為：

$$\delta_t(i) = \max_{q_1^* q_2^* \dots q_N^*} P[q_1, q_2, \dots, q_{t-1}, q_t = i, o_1, o_2, \dots, o_t | \lambda]$$

則我們可以進一步得到一遞迴定義：

$$\delta_{t+1}(j) = \max_i [\delta_t(i) a_{ij}] \cdot b_j(o_{t+1})$$

故 Viterbi 演算法利用此遞迴定義配合 Dynamic Programming 可快速找出最佳值並由 Backtracking 找到最佳路徑。

第4題

Decision tree 可以根據一連串的 feature 及利用 entropy 為 splitting criteria 來有效分類資料。

在 Triphone 的資料裡，許多的 event 在 training data 之中會成為 unseen event，因此造成 training 上的困難。因此根據聲學的知識，我們可以將各樣的 triphone 分類，使得在發聲上相近的 triphone 能夠被找出，讓 unseen triphone 能夠由類似的 triphone 得到適當的資料來估計。

第6題

Katz Smoothing 是由 Good-Turing Smoothing 而來，後者會將高 frequency 的 event 之機率搬動挪作低 frequency 的 event 的機率，但此一作法是設法總是有更高 frequency 的 event 存在。在實際應用時此一假設若不成立則會遺失最高 frequency event 的機率。同時高 frequency 的 event 往往是可信任的，因此 Katz Smoothing 針對 frequency 低於某一 threshold 的 event 做 smoothing，若高過此一 threshold 則不調整。將低於 threshold 的 event 做 discount，將 frequency 分給 unseen events，再依照 Good-Turing Smoothing。

第7題

a小題

voiced signals: 由振動聲帶所產生的聲音，稱為濁音。在 time domain 上的波形特徵是會有固定的形狀重覆出現。

voiced signals: 不經由振動聲帶所產生的聲音，稱為清音。在 time domain 上的波形特徵是沒有固定的形狀重覆出現。

b小題

pitch 指的是聲音的音高，頻率越高的聲調越高。在中文裡經由變化音高造成聲調，例是一聲是音高不變，二聲則是音高從低變高，三聲則是由高變低再拉高。

第9題

tree lexicon 的做法是將每個字的各聲學單元視為一個 node，則在辨識時對每 node 只需找出其最佳 path 即可得知到達該 node 的最大機率。這樣的好處是可以大量節省空間的使用，並且可以加快運算速度，但是因為會捨棄掉一部份資訊，因此不會找到最佳解。