

## Hidden Markov Models:

### Working out Viterbi:

We are expected to manually work out part of viterbi in the exams.

If they say you should do three steps of viterbi, then we should only do three steps.

You should show working out in maths.

### Assumptions of HMM's:

- Transition and Emission probabilities are invariant of time
- Output independence.
- Outputs are dependent on only  $n$  previous hidden states in an  $n$  order markov model.

When asked to evaluate a model, you should evaluate a model – how does it break HMM assumptions not “why is this probability equal to  $x$  not  $y$ ”. It is okay to suggest improvements.

Discrete modelling of continuous cases are not satisfactory.

The markov assumption is the one of a limited horizon.

The other assumptions of HMM's are output independence. Another is time invariance (transition and emission probabilities are constant with time).

With HMM's it's ambiguous whether 0 probabilities are impossible (true zero) or unseen (near-zero).

Be careful about what your transition and emission probabilities are. In general you should smooth emissions (since in general there are usually far fewer hidden states than emissions). Case in point: POS tagging and amino acids. If you have to smooth (which you should not in general), then you would be better off smoothing emission probabilities.

If viterbi gets wrong outcome then you've trained it on a niche dataset. The way to solve this is to get a better dataset.

Heaps' Law says you will never see all the words and so will have some zero-probability transitions. Solved by smoothing.

Check to see whether the matrix has been wholly or partially smoothed. verbs and nouns etc are “open class” – the size of this set is near-infinite. Other words are a closed-set (closed-class – as are connectives).

We don't really go to high order HMM's because there is too much data sparseness. IE if we have 5 states and a 10 order HMM then we have 9.7m states. Most will be zero.