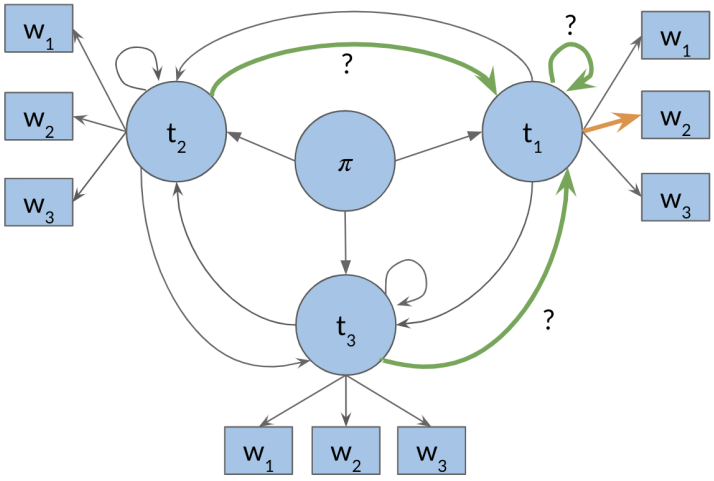
Viterbi: Forward Pass

This will be best illustrated with an example:



$C =$

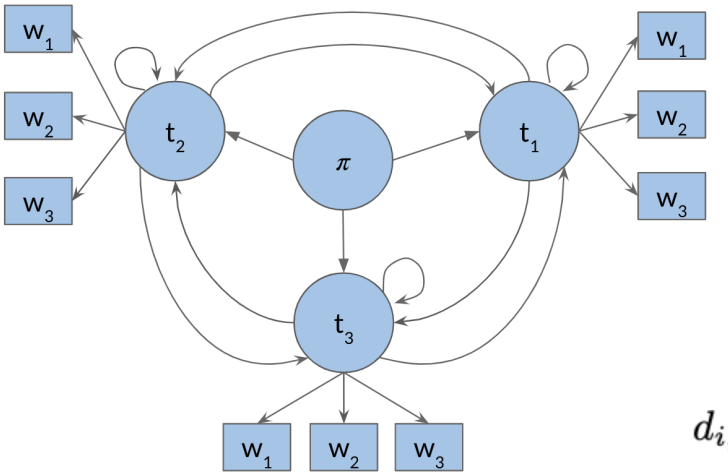
	w_1	w_2	...	w_K
t_1	$c_{1,1}$	$c_{1,2}$		$c_{1,K}$
...				
t_N	$c_{N,1}$	$c_{N,2}$		$c_{N,K}$

$$c_{1,2} = \max_k c_{k,1} * a_{k,1} * b_{1,cindex(w_2)}$$

So to populate a cell (i.e. 1,2) in the image above, you have to take the max of [kth cells in the previous column, times the corresponding transition probability of the kth POS to the first POS times the emission probability of the first POS and the current word you are looking at]. You do that for all the cells. Take a paper and a pencil, and make sure you understand how it is done.

The general rule is $c_{i,j} = \max_k c_{k,j-1} * a_{k,i} * b_{i,cindex(w_j)}$

Now to populate the **D** matrix, you will keep track of the argmax of where you came from as follows:



$D =$

	w_1	w_2	...	w_K
t_1	$d_{1,1}$	$d_{1,2}$		$d_{1,K}$
...				
t_N	$d_{N,1}$	$d_{N,2}$		$d_{N,K}$

$$c_{i,j} = \max_k c_{k,j-1} * a_{k,i} * b_{i,cindex(w_j)}$$
$$d_{i,j} = \operatorname{argmax}_k c_{k,j-1} * a_{k,i} * b_{i,cindex(w_j)}$$

Note that the only difference between c_{ij} and d_{ij}, is that in the former you compute the probability and in the latter you keep track of the index of the row where that probability came from. So you keep track of which k was used to get that max probability.