

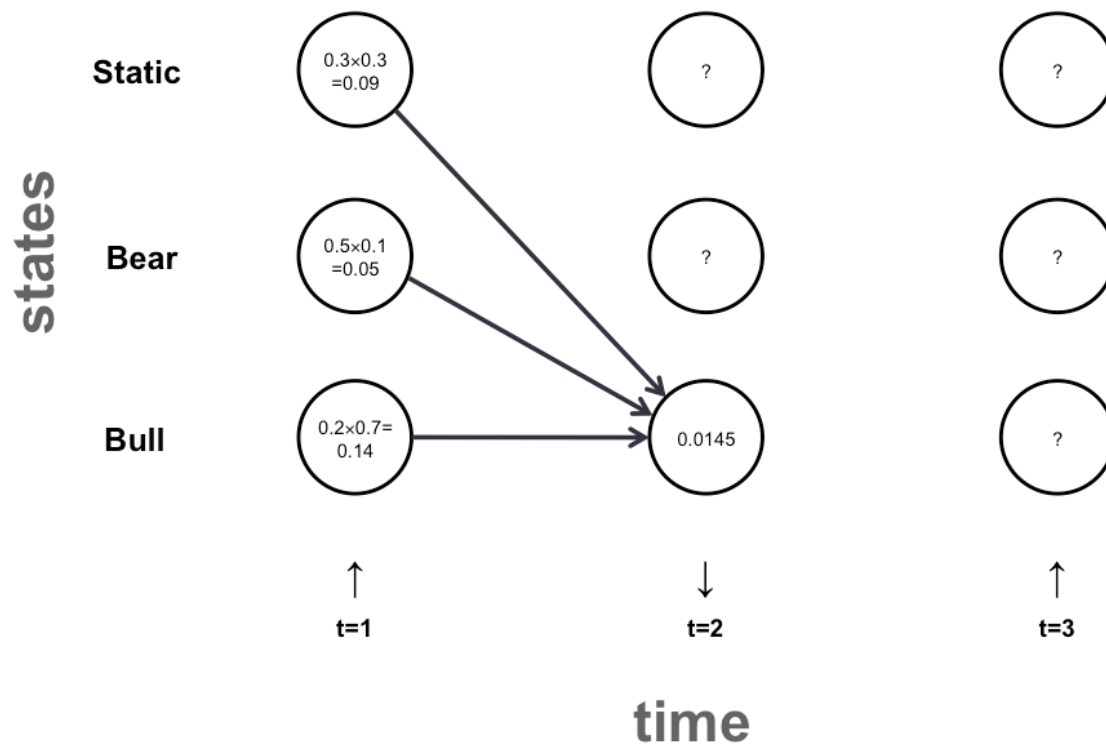
Computational Linguistics 1 - Fall 2011  
 CMSC/LING723, LBSC 744  
 Homework 3 - Due 13 October 2011

## Submission Guidelines

1. Print out any written portion of the assignment and submit in class on October 13, *as well as* emailing your writeup, in .pdf format, to `compling723.fall2011@gmail.com`.
2. Any code you write for this homework should also be submitted to `compling723.fall2011@gmail.com`.
3. Be sure include your full name and e-mail in your submission, and clearly indicate each problem number in your solution set.

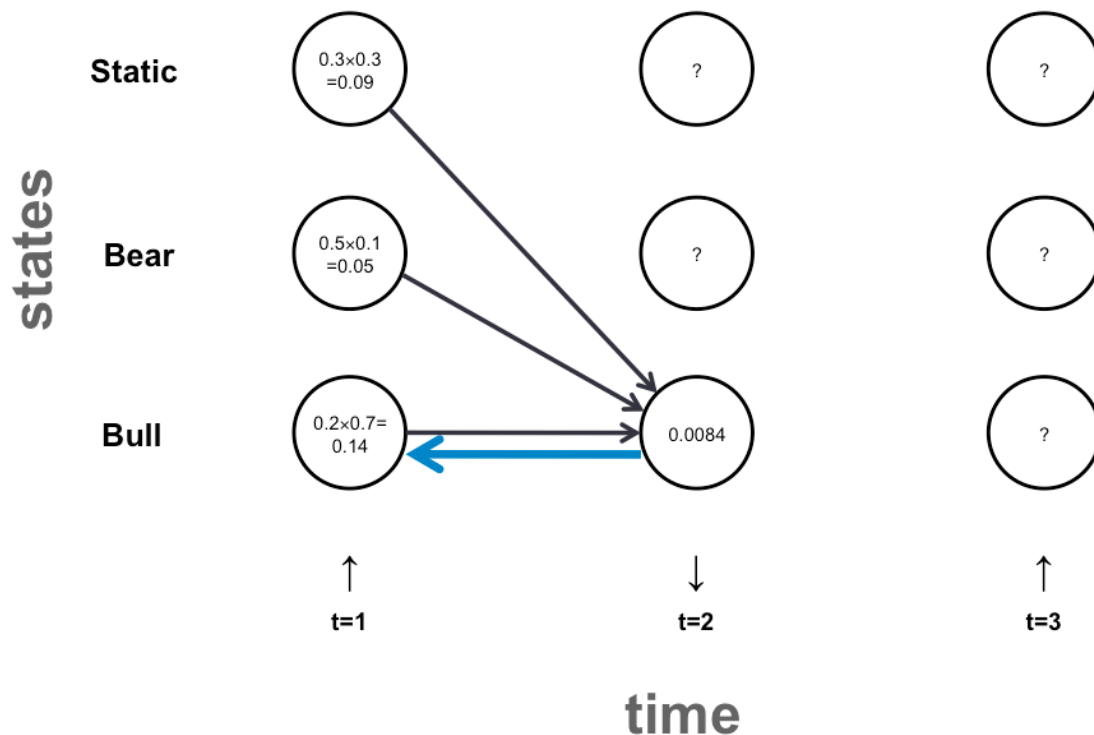
## Problem 1 (5 points)

Complete the Forward ‘trellis’ below, from Lecture 9. Use the Stock Market model defined on slide 13 of Lecture 9 (also on the handouts from class). Show your work, i.e. the values along each transition and at each state for each time step. What is the probability of the observation sequence “↑ ↓ ↑”?



## Problem 2 (5 points)

Complete the Viterbi ‘trellis’ below, from Lecture 9. Use the Stock Market model defined on slide 13 of Lecture 9 (also on the handouts from class). Show your work, i.e. the values along each transition, at each state for each time step, and the backpointer at each state for each time step. What is the most likely state sequence for the observation sequence “ $\uparrow \downarrow \uparrow$ ”?



## Problem 3 (40 points)

Write a Laplacian-smoothed bigram part-of-speech (POS) tagger, trained on our Wall Street Journal corpus, f2-21.train.pos (available from the class mailing list); a.k.a., induce a bi-tag HMM POS tagging model from our training data, following slide 41 of lecture 9.

For all words that occur fewer than 5 times in the training corpus, replace these words—in both the training and test corpus—with the symbol  $\langle \text{unk} \rangle$ .

Write a Viterbi decoder to make use of your HMM POS tagging model to find the highest probability tag sequence for each string in the test set, f2-21.test.pos. What is the POS-tag sequence for the first sentence in our test set?

How accurate is your POS tagger? (Evaluate the number of tags correctly predicted by your tagger, divided by the total number of words in the test corpus.)