Artificial Intelligence (CPSC 4383) Dr. Keith Bush
Homework 3: Bush Speed Figures Fall 2014

Important Dates:
Post Date: October 31 (12:00 PM).
Due Date: ~~November 17 (11:59 PM)~~. November 19 (11:59 PM)

Artificial Intelligence Concepts:
Temporal uncertainty, hidden Markov models, filtering, prediction

Task Summary:
Handicapping horse racing has long been dominated by the Beyer Speed Figure (abbreviated BeyerSF), a tool for evaluating the speed of a horse on a given day adjusting for all possible external factors (track conditions, length of race, starting gate, etc). The BeyerSF is remarkably good at predicting horse racing performance, but it does not account for two things. First, the BeyerSF is an observation of a horse's generated speed on a given day, which is a noisy measurement of a horse's true speed capability. Second, race horses (just like human athletes) are trained to attain a 'peak' true speed capability at specific times in the racing season such that their chances of winning a major purse are maximized. Thus, the true speed capability of a horse changes over time according to the training regimen, diet, and racing schedule of that particular horse.

To predict the outcome of a major horse race, such as the Arkansas Derby hosted at Oaklawn each Spring, we would need to be able to estimate both the true speed capability of a horse on derby day, which we will term the Bush Speed Figure (abbreviated BuSF), as well as the observed speed of that horse. Using some wizardry and mathematical acumen, the Bush Speed Figure system incorporates a transition model of true speed capability as a function of time (i.e., how the true speed capability of a horse changes from time t-1 to time t) as well as how the speed of the horse will be observed in the race (a noisy observation of BuSF at time t, termed obsBuSF). The BuSF and obsBuSF are integers on the range [0,5], where 0 and 5 indicate the respective physiological minimum and maximum speeds a thoroughbred horse can achieve. The transition and observation models of the BuSF handicapping system are provided below.

Transition Model: $p(\mathbf{BuSF_t} \mid BuSF_{t-1})$

0	**1**	**2**	**3**	**4**	**5**	$BuSF_{t-1}$
0.4	0.6	0.0	0.0	0.0	0.0	0
0.1	0.3	0.6	0.0	0.0	0.0	1
0.0	0.1	0.3	0.6	0.0	0.0	2
0.0	0.0	0.1	0.3	0.6	0.0	3
0.0	0.0	0.0	0.1	0.3	0.6	4
0.4	0.2	0.0	0.0	0.1	0.3	5

Observation Model: $p(\mathbf{obsBuSF_t} \mid BuSF_t)$

0	1	2	3	4	5	$BuSF_t$
0.9	0.1	0.0	0.0	0.0	0.0	0
0.1	0.8	0.1	0.0	0.0	0.0	1
0.0	0.1	0.8	0.1	0.0	0.0	2
0.0	0.0	0.1	0.8	0.1	0.0	3
0.0	0.0	0.0	0.1	0.8	0.1	4
0.0	0.0	0.0	0.0	0.1	0.9	5

Assume that the BuSF of a horse prior to the start of the data is unknown (use a uniform distribution over the BuSF values). Using obsBuSF values for 6 horses over 24 historical (but sequential) races leading up to the Arkansas Derby, implement the following calculations using an (approved) computer language:

1. filtering over the historical races to optimally compute the distribution of BuSF values for each horse at the time of race;
2. prediction to estimate the distribution of BuSF and obsBuSF values for each horse at the Arkansas Derby (i.e., race 25);
3. expected observed Bush Speed Figure, expObsBuSF, for each horse using the formula:

$$expObsBuSF = \sum_{i=1}^{6} p(obsBuSF_i) * i;$$

4. rank the horses in order of predicted finish according to their expObsBuSF; and,
5. report your rank of the horses for the Arkansas Derby in a short text file (or other allowable document type) as shown below.

Predicted Finish	Horse Name	expObsBuSF	Derby obsBuSF
1			
2			
3			
4			
5			
6			

Rubrics

Description	Point value
Submission, report, compilation, and execution	50
Filtering implementation	20
Prediction implementation	20
Expectation and ranking	10

Model parameters and example data

The instructor has provided a directory "hmwk3_release.zip" containing numerous resources to aide in implementing this assignment. Within this compressed directory there are several subdirectories:

params: contains *.txt versions of the prior, transition, and observation distributions as well as the race parameters (6 horses, 6 BuSF numbers, and 24 races).
observationsTrain: contains a *.txt version of the obsBuSF figures for each horse for a 24 race season as well as simulated Derby results.
truthsTrain: contains *.txt versions of the true prior BuSF, the true BuSF for each race and the true BuSF for the Derby.
observationsTest: contains a *.txt version of the obsBuSF figures for each horse for the 24 race season.

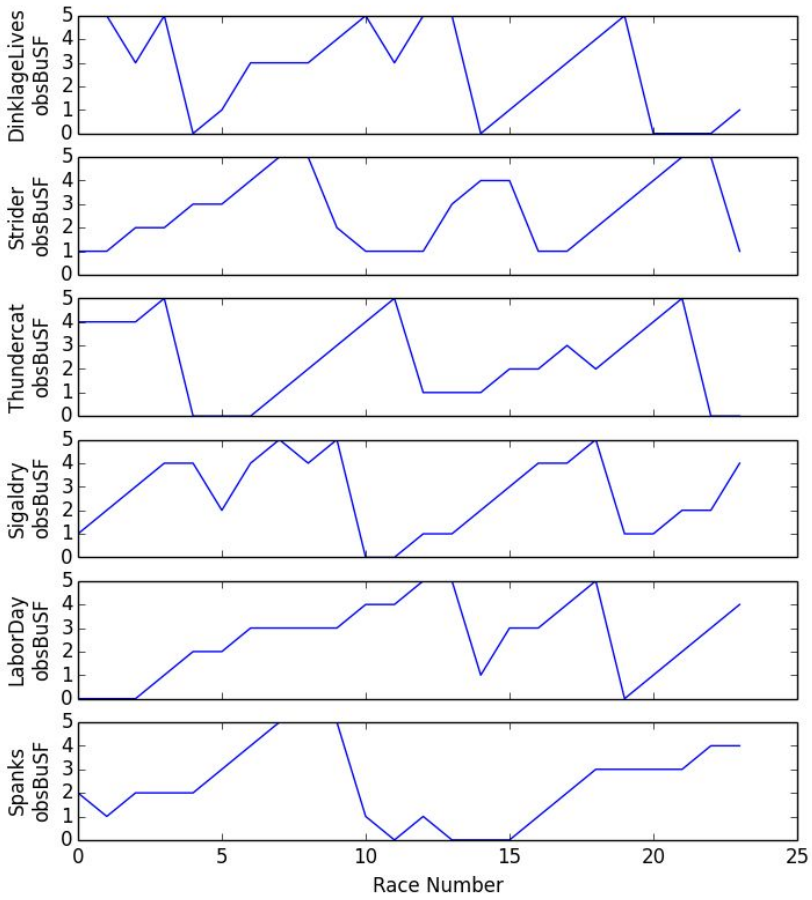
Note: the training observations and truths allow you to see all of the data (both hidden variables and observed variables), which will allow you to debug your code. The test observations are the "real" races on which you will predict the outcome of the Arkansas Derby.

Submission Format
You must submit your code and report as a compressed archive of a directory named <userid>.hmwk.3 that includes all of your data and source files as well as a README file that describes the purpose of each file and how to compile and execute your code.

How to go about completing this assignment:
1. read the specification and make a work plan/timeline;
2. ask questions!;
3. break down filtering into the prediction and refinement steps;
4. implement the prediction step;
5. implement the refinement step;
6. compute the distributions of the race horses;
7. implement the expected value calculation;
8. calculate the expected values of the horses;
9. sort the results;
10. write the report and submit.

Appendix: Example Observations and Solution of the Training Dataset

Simulated Racing Season Results
The figure below depicts the observed Bush Speed Figures across a twenty-four race season for the six horses competing in the Arkansas Derby. It should be clear from the plot below that the each horse exhibits multiple performance cycles per racing season.



Simulated Arkansas Derby Results

The data below depict a predicted outcome of the Arkansas Derby based on the transition and observation models, provided above, as well as the "training" racing season observations depicted in the above figure (and provided to you in observationsTrain). The first two columns show the predicted finish order of the horses and their names. The third column depicts the expected observed speed of the horse and the fourth column depicts the true observed speed of the horse.

```
1       'LaborDay'    4.3772921872635591    4.0
2       'Sigaldry'    4.0775448006826771    4.0
3         'Spanks'     3.998051881253804    3.0
4 'DinklageLives'    1.4646717530050333    2.0
5        'Strider'    1.3359999999999999    1.0
6     'Thundercat'  0.76574218749999989    0.0
```