Stock analysis using Hidden Markov Model

Feature selection:

Search Method:

Attribute ranking.

Attribute Evaluator (unsupervised):

Principal Components Attribute Transformer

Correlation matrix

1	-0.88	-0.85	-0.91	-0.92	-0.89	-0.74	-0.9	0.68
-0.88	1	0.84	0.95	0.95	0.92	0.66	0.93	-0.66
-0.85	0.84	1	0.9	0.9	0.88	0.61	0.91	-0.61
-0.91	0.95	0.9	1	0.99	0.96	0.69	0.97	-0.67
-0.92	0.95	0.9	0.99	1	0.97	0.7	0.97	-0.69
-0.89	0.92	0.88	0.96	0.97	1	0.73	0.94	-0.64
-0.74	0.66	0.61	0.69	0.7	0.73	1	0.64	-0.57
-0.9	0.93	0.91	0.97	0.97	0.94	0.64	1	-0.67
0.68	-0.66	-0.61	-0.67	-0.69	-0.64	-0.57	-0.67	1

eigenvalue	proportion	cumulative	
7.57106	0.84123	0.84123	-0.359S6-0.356S5-0.352S9-0.351S7-0.346S3...
0.54122	0.06014	0.90136	-0.65S10+0.617S8-0.264S4-0.213S9-0.158S5...
0.44844	0.04983	0.95119	-0.701S10-0.686S8-0.117S7+0.102S9+0.091S2...

Eigenvectors

V1	V2	V3

0.3436 -0.0462 0.0911 S2

-0.3458 -0.1503 0.0509 S3

-0.3326 -0.2641 0.0407 S4

-0.3562 -0.1577 0.0085 S5

-0.3586 -0.1325 0.0075 S6

-0.3514 -0.1212 -0.1168 S7

-0.2759 0.6169 -0.6865 S8

-0.3523 -0.2129 0.1019 S9

0.2694 -0.65 -0.7015 S10

Ranked attributes:

0.1588 1 -0.359S6-0.356S5-0.352S9-0.351S7-0.346S3...

0.0986 2 -0.65S10+0.617S8-0.264S4-0.213S9-0.158S5...

0.0488 3 -0.701S10-0.686S8-0.117S7+0.102S9+0.091S2...

Selected attributes: 1,2,3 : 3

(1) Which variables matter for predicting S1?
S2, S3 and S4 matters most for predicting because they have more correlation with S1. I used Principal Components Attribute evaluator to get the most important features for this data.

(2) Does S1 go up or down over this period?
S1 goes up for some days and went down for some. I am attaching the excel sheet which 0/1 entries where 0 means stock went down and 1 depicts that stock went up.
On the last day, S1 went down.

(3) How much confidence do you have in your model? Why and when would it fail?
My model has good confidence because I took care of overfitting problem using dropout. Although it may fail in case of uncertainty of some events.

(4) What techniques did you use? Why?

Stock price prediction is regression problem but we can formulate the problem as binary classification problem because closing price is going to be lower or higher with respect to today. I used Hidden Markov Model and Deep neural network with 2 hidden layers and

sigmoid/tanh/relu as activation function. To overcome the problem of overfitting, I used dropout technique. To prove this problem as regression problem, I used linear regression