Best accuracies for classifiers on the full test set

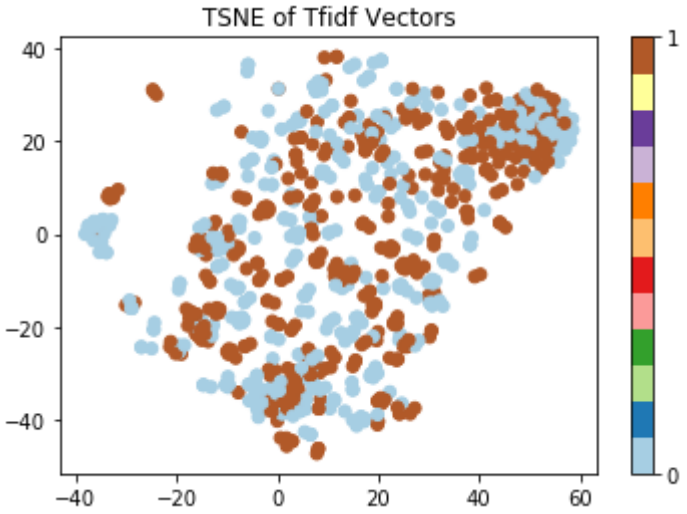
Accuracy	Number Correct	Test Size	Classifier	Model	Inferred
50%	22849	45333	SVC	Doc2Vec dm/c	False
50%	4548	9066	SVC	Doc2Vec dm/m	True
50%	22672	45333	Logistic Regression	Doc2Vec dm/c	False

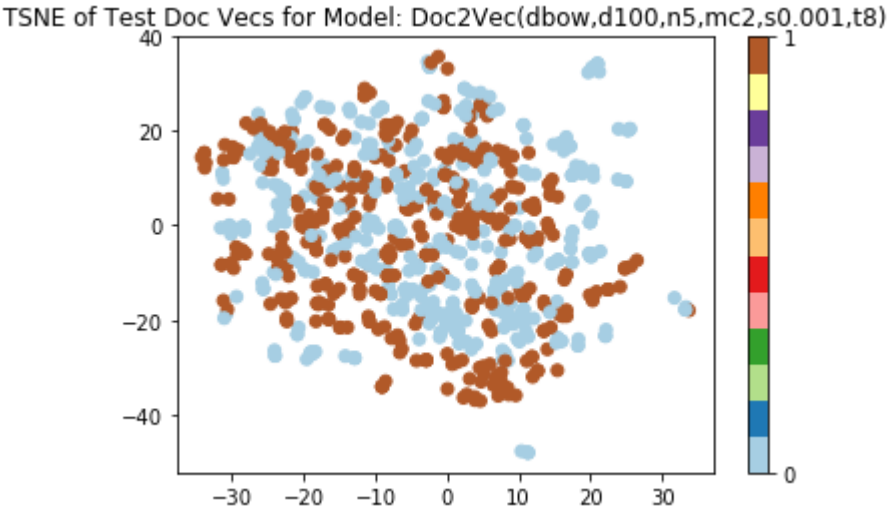
Results of grid search on SVC (There's an optimum?)

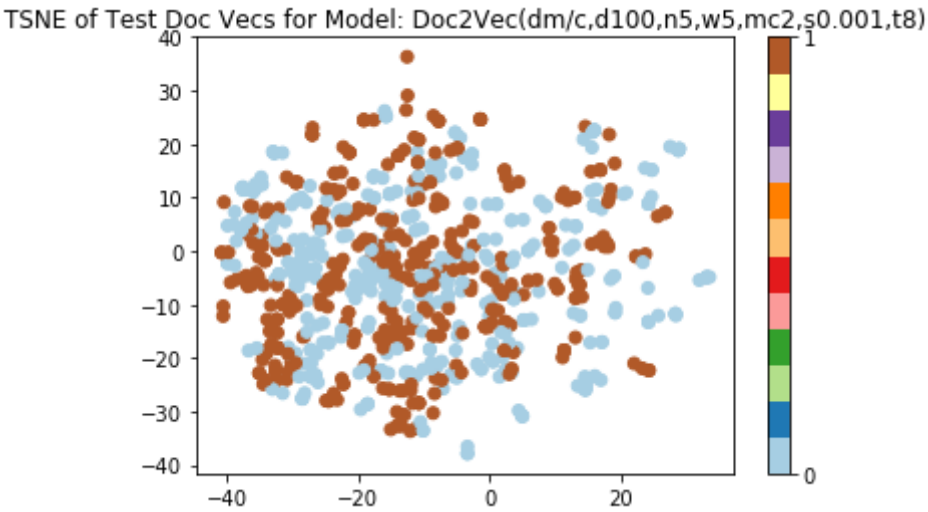
Accuracy	Classifier	C	Kernel	Model	Test Size
58.2%	SVC	10000	rbf	Doc2Vec dm/c	804
61.4%	SVC	100000	rbf	Doc2Vec dm/c	804
60.9%	SVC	1000000	rbf	Doc2Vec dm/c	804

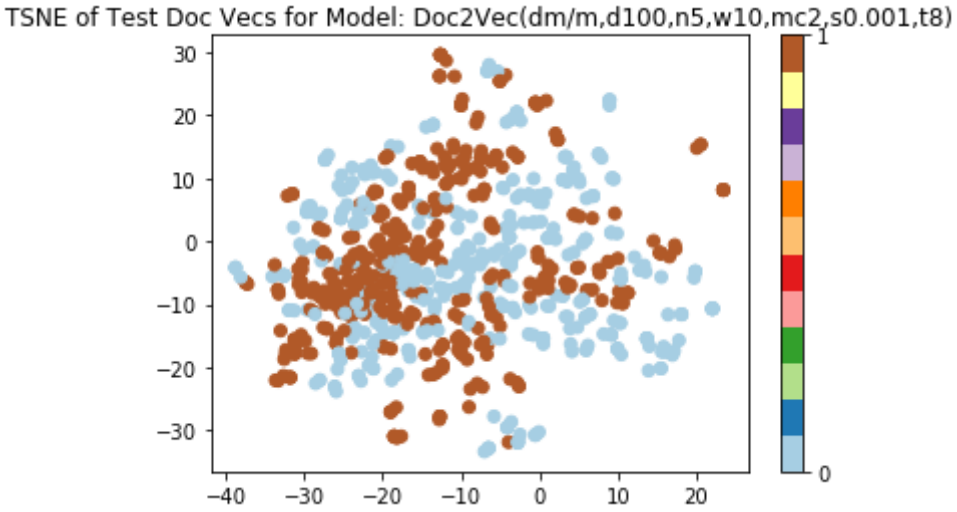
Prediction accuracy for day-of alphas

Accuracy	Number Correct	Test Size	Classifier	Model
82%	328	402	Logistic Regression	TfidfVectorizer
98%	394	402	Decision Tree	TfidfVectorizer
69%	279	402	SVC	TfidfVectorizer
51%	410	804	Logistic Regression	Doc2Vec - dbow
51%	407	804	Decision Tree	Doc2Vec - dbow
46%	370	804	SVC	Doc2Vec - dbow
51%	410	804	Logistic Regression	Doc2Vec - dm/c
47%	378	804	Decision Tree	Doc2Vec - dm/c
49%	397	804	SVC	Doc2Vec - dm/c
58%	469	804	Logistic Regression	Doc2Vec - dm/m
55%	439	804	Decision Tree	Doc2Vec - dm/m
55%	441	804	SVC	Doc2Vec - dm/m


TSNE of Tfidf Vectors
40
20
0
−20
−40
−40
−20
0
20
40
60
1
0


TSNE of Test Doc Vecs for Model: Doc2Vec(dbow,d100,n5,mc2,s0.001,t8)
40
20
0
−20
−40
−30
−20
−10
0
10
20
30
1
0


TSNE of Test Doc Vecs for Model: Doc2Vec(dm/c,d100,n5,w5,mc2,s0.001,t8)
40
30
20
10
0
−10
−20
−30
−40
−40
−20
0
20
1
0



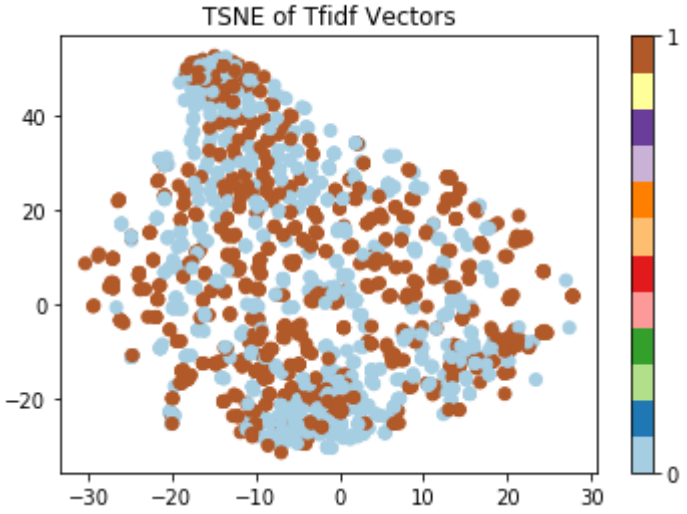
Data Split Sizes (Odd...)

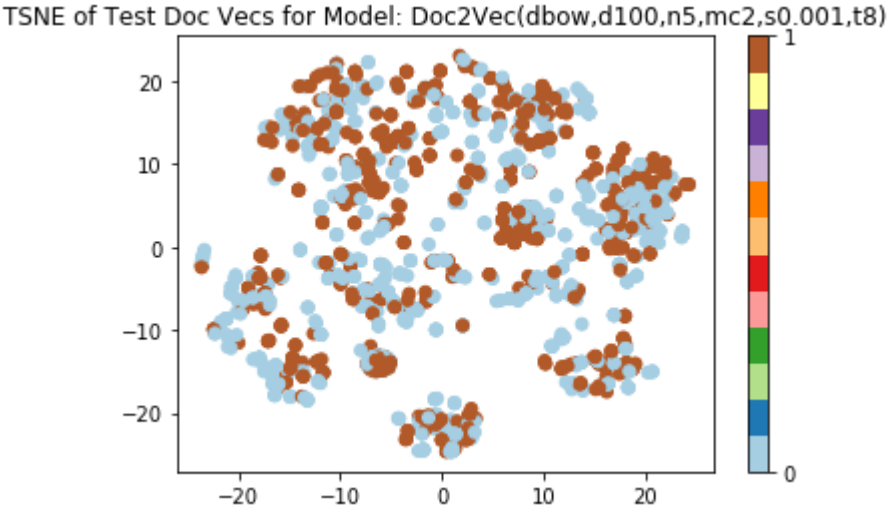
Data Path	Number of Docs
train-pos 1 day	50122
train-pos 30 day	37838
train-neg 1 day	52040
train-neg 30 day	397188
test-pos 1 day	22085
test-pos 30 day	9547
test-neg 1 day	23238
test-neg 30day	99402

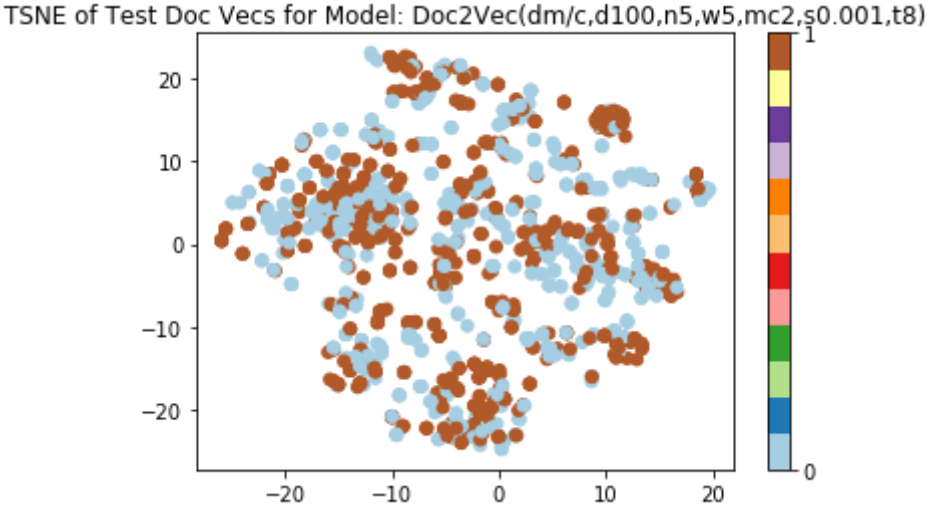
Prediction accuracy for 30-day returns

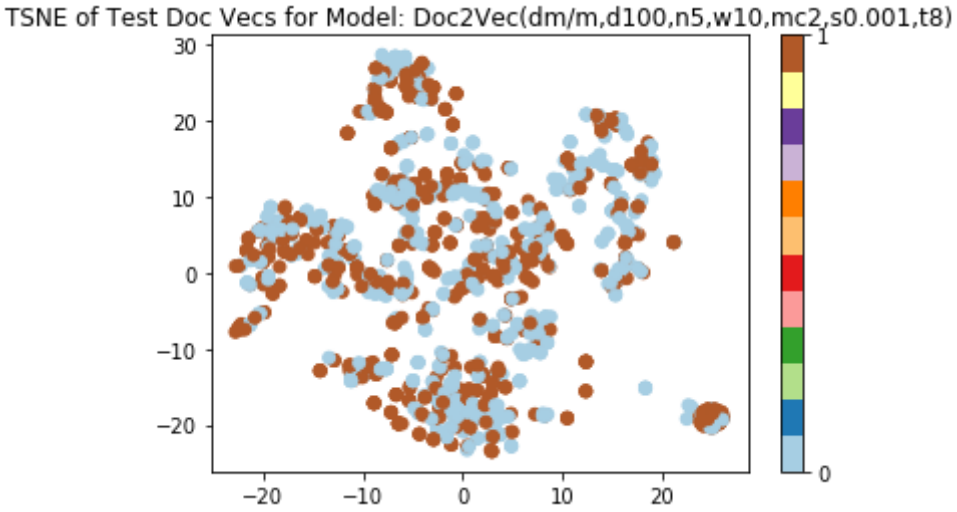
Accuracy	Number Correct	Test Size	Classifier	Model
77%	614	801	Logistic Regression	TfidfVectorizer
98%	784	801	Decision Tree	TfidfVectorizer
50%	401	801	SVC	TfidfVectorizer
51%	812	1602	Logistic Regression	Doc2Vec - dbow
45%	722	1602	Decision Tree	Doc2Vec - dbow
52%	833	1602	SVC	Doc2Vec - dbow
45%	728	1602	Logistic Regression	Doc2Vec - dm/c
51%	813	1602	Decision Tree	Doc2Vec - dm/c
48%	766	1602	SVC	Doc2Vec - dm/c
47%	749	1602	Logistic Regression	Doc2Vec - dm/m
52%	827	1602	Decision Tree	Doc2Vec - dm/m

Accuracy	Number Correct	Test Size	Classifier	Model
47%	759	1602	SVC	Doc2Vec - dm/m








TSNE of Test Doc Vecs for Model: Doc2Vec(dm/m,d100,n5,w10,mc2,s0.001,t8)
30
20
10
0
−10
−20
−20
−10
0
10
20
1
0

In [ ]: