Práctica Final

Clasificador con validación de ventanas deslizantes

1. Qué se ha hecho

Se ha creado una estructura en loop que va ejecutando el entrenamiento con una ventana cada vez mayor.
Para esto, se ha cogido el dataset entero y se han cogido las fechas de inicio y fin. Como el último año no está completo (y en la fecha de descarga, no habían pasado 50 días), no nos sirve para contrastar la predicción con la realidad, así que se ha descartado, siendo nuestro último año el último menos uno.
A continuación, se han extraído las rentabilidades pasadas a 10 y 50 días, y se ha consultado si la futura ha sido de subida o bajada, formando así el dataframe que se va a usar.
Dentro de un for del primer año a último año menos uno, se han ido cogiendo los valores de entrenamiento, y los de testeo, siendo el siguiente año.
Se han entrenado los modelos de entrenamiento de '*k* nearest neighbor', 'random forest' y redes neuronales. El control de entrenamiento ha sido el 'cross validation'. Los resultados, la matriz de confusión, se han recogido en una estructura para consultarlos a posteriori.

2. Resultados

kNN:

2012	Referencia		2013	Referencia		2014	Referencia		2015	Referencia	
Pred.	↓	↑	Pred.	↓	↑	Pred.	↓	↑	Pred.	↓	↑
↓	1	22	↓	1	83	↓	28	126	↓	9	24
↑	49	139	↑	16	111	↑	7	49	↑	27	151

2016	Referencia		2017	Referencia	
Pred.	↓	↑	Pred.	↓	↑
↓	15	73	↓	27	134
↑	18	101	↑	5	43

Random forest:

2012	Referencia		2013	Referencia		2014	Referencia		2015	Referencia	
Pred.	↓	↑	Pred.	↓	↑	Pred.	↓	↑	Pred.	↓	↑
↓	2	21	↓	5	79	↓	46	108	↓	11	22
↑	63	125	↑	29	98	↑	21	35	↑	62	116

2016	Referencia		2017	Referencia	
Pred.	↓	↑	Pred.	↓	↑
↓	23	65	↓	34	127
↑	31	88	↑	9	39

Red neuronal:

2012	Referencia		2013	Referencia		2014	Referencia		2015	Referencia	
Pred.	↓	↑	Pred.	↓	↑	Pred.	↓	↑	Pred.	↓	↑
↓	1	22	↓	0	84	↓	14	140	↓	5	28
↑	28	160	↑	3	124	↑	2	54	↑	6	172

2016	Referencia		2017	Referencia	
Pred.	↓	↑	Pred.	↓	↑
↓	6	82	↓	2	159
↑	3	116	↑	0	48

Random

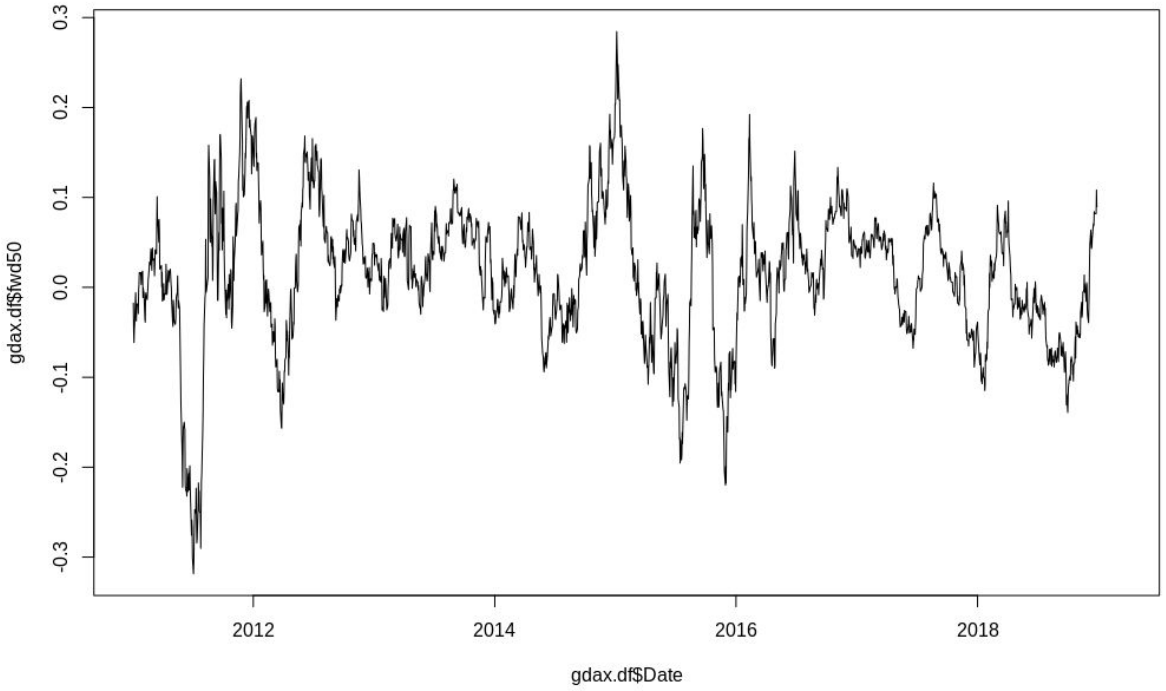
2012	Referencia		2013	Referencia		2014	Referencia		2015	Referencia	
Pred.	↓	↑	Pred.	↓	↑	Pred.	↓	↑	Pred.	↓	↑
↓	12	11	↓	10	23	↓	71	83	↓	10	23
↑	95	93	↑	103	75	↑	18	38	↑	103	75

2016	Referencia		2017	Referencia	
Pred.	↓	↑	Pred.	↓	↑
↓	49	39	↓	83	78
↑	63	56	↑	26	22

3. Interpretación

Aunque cada resultado sea diferente, por lo general se ha detectado un resultado pesimista (demasiados falsos negativos). Esto puede deberse a la falta de configuración de los algoritmos de aprendizaje por una parte, o que el set de datos tenía una tendencia ascendente que no ha sabido interpretar. Al ir subiendo, los datos anteriores no han servido para identificar la proxima subida.

Si se hace un plot de los resultados, se observa que no es el caso:



Así que deducimos que aunque la máquina haya calculado algo, la veracidad de estas predicciones no es del todo correcta. El predictor aleatorio no se ha alejado mucho de las demás predicciones.