MGMT 59000 Machine Learning

Final Project – Part 1

Selecting the greater annual Sharpe amongst one generated by Keras tuner[1] and manually tuned hyperparameters, we reached a Sharpe ratio of 5.502

	Keras tuner	Manually tuned model
Batch	**128**	**64**
Learning rate	**0.001**	**0.001**
Optimizer	**Adam**	**Adam**
Patience	**3**	**3**

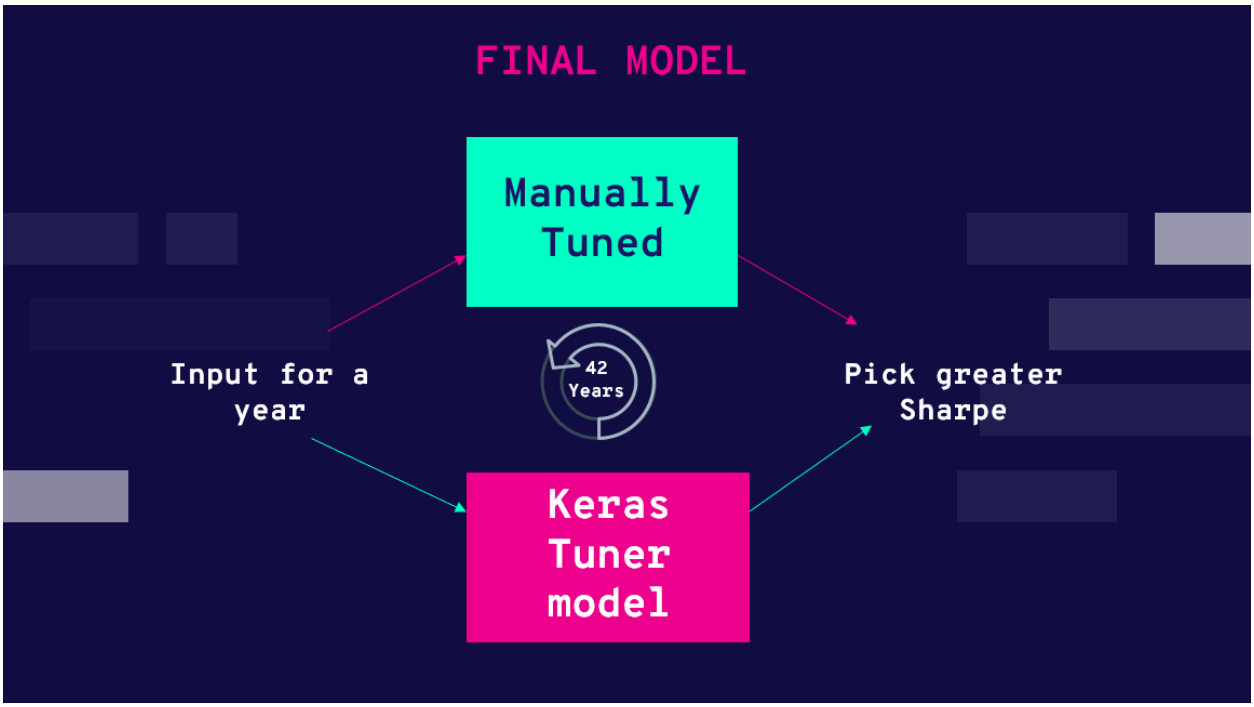
[1]The Keras auto tuner generated 42 distinct models based on each year's stock data.

Manually tuned Hyperparameters

Architecture	Batch Size	Learning Rate	Optimizer	Patience	Sharpe	Comments
(32,64,16)	64	0.001	Adam	3	5.33	Best
(64,32,16)	32	0.001	Adam	3	5.4371	Not Repeatable
(32,64,16)	70	0.001	Adam	3	5.20	
(32, 64, 16)	96	0.001	Adam	4	4.802	
(32, 64, 16)	64	0.001	RMSprop	3	4.67	Adam more suited
(32,64,16)	64	0.0001	Adam	3	4.92	
(32,64,16)	64	0.005	Adam	3	5.09	0.001 best suited
(64,128,64)	128	0.001	Adam	3	4.46	Higher units overfitting
(64,128,128,64,16)	128	0.001	Adam	3	5.22	Very time consuming for minimal increase in Sharpe
(64,128,128,64,16)	128	0.01	Adam	3	3.72	
(64,128,128,64,16)	64	0.001	Adam	3	2.46	
(64,32,16) (Sigmoid)	32	0.001	Adam	3	4.46	Not Repeatable

Many more variations of parameters were experimented upon after using only 6 or 12 years of data, however, did not track these changes if the Sharpe was very low or the change was minimal.

Final Model



Final Project – Part 2

Selecting the best R^2 by trying out different parameter combinations helped us achieve a R^2 of -0.5616. We have only used Adam optimizer for all iterations.

Dropout Layers?	Architecture	Learning rate	Patience	Batch Size	Epochs	R^2
Yes	(1024,512,256,128,128,128,128,64,32,1)	0.000001	1	64	100	-0.5616
No	(1024,512,256,128,128,128,128,64,32,1)	0.000001	1	64	100	-0.9656
No	(1024,512,256,128,128,128,128,64,32,1)	0.00001	3	64	100	-1.3456
No	(1024,256,128,64,32,1)	0.0001	1	64	100	-2.0887
No	(1024,256,128,64,32,1)	0.0001	3	256	100	-2.3290
No	(512,256,128,64,32,1)	0.01	1	256	100	-39.1675
No	(256,128,64,32,1)	0.00001	3	64	100	-2.4831
No	(256,128,64,32,1)	0.0001	2	64	100	-3.4850
No	(256,128,64,32,1)	0.001	1	64	100	-3.6770

Architecture of best model:

```
Model: "sequential_551"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 layer_normalization_551 (La  (None, 137)              274
 yerNormalization)

 dense_5510 (Dense)          (None, 1024)              141312

 dropout_3857 (Dropout)      (None, 1024)              0

 dense_5511 (Dense)          (None, 512)               524800

 dropout_3858 (Dropout)      (None, 512)               0

 dense_5512 (Dense)          (None, 256)               131328

 dropout_3859 (Dropout)      (None, 256)               0

 batch_normalization_1653 (B  (None, 256)              1024
 atchNormalization)

 dense_5513 (Dense)          (None, 128)               32896

 dropout_3860 (Dropout)      (None, 128)               0

 dense_5514 (Dense)          (None, 128)               16512

 dropout_3861 (Dropout)      (None, 128)               0

 dense_5515 (Dense)          (None, 128)               16512

 dropout_3862 (Dropout)      (None, 128)               0

 dense_5516 (Dense)          (None, 128)               16512

 dropout_3863 (Dropout)      (None, 128)               0

 batch_normalization_1654 (B  (None, 128)              512
 atchNormalization)

 dense_5517 (Dense)          (None, 64)                8256

 batch_normalization_1655 (B  (None, 64)               256
 atchNormalization)

 dense_5518 (Dense)          (None, 32)                2080

 dense_5519 (Dense)          (None, 1)                 33

=================================================================
Total params: 892,307
Trainable params: 891,411
Non-trainable params: 896
_________________________________________________________________
```