FINAL PROJECT: STOCK MARKET

A network tour of data science

Jeroen Le Maire

Setup

- Dataset contains about 500 stocks

- Per stock: 27 features on 15 dates (quarterly)

- Which features are the most important ?

- Prediction through classification

Data preparation and cleaning

- Fetch the data in the right format

- Add the return to every data row (value + marker)

- Remove the data with unknown date

- Remove the incomplete data

- For all features calculate the Z-score and remove the outliers
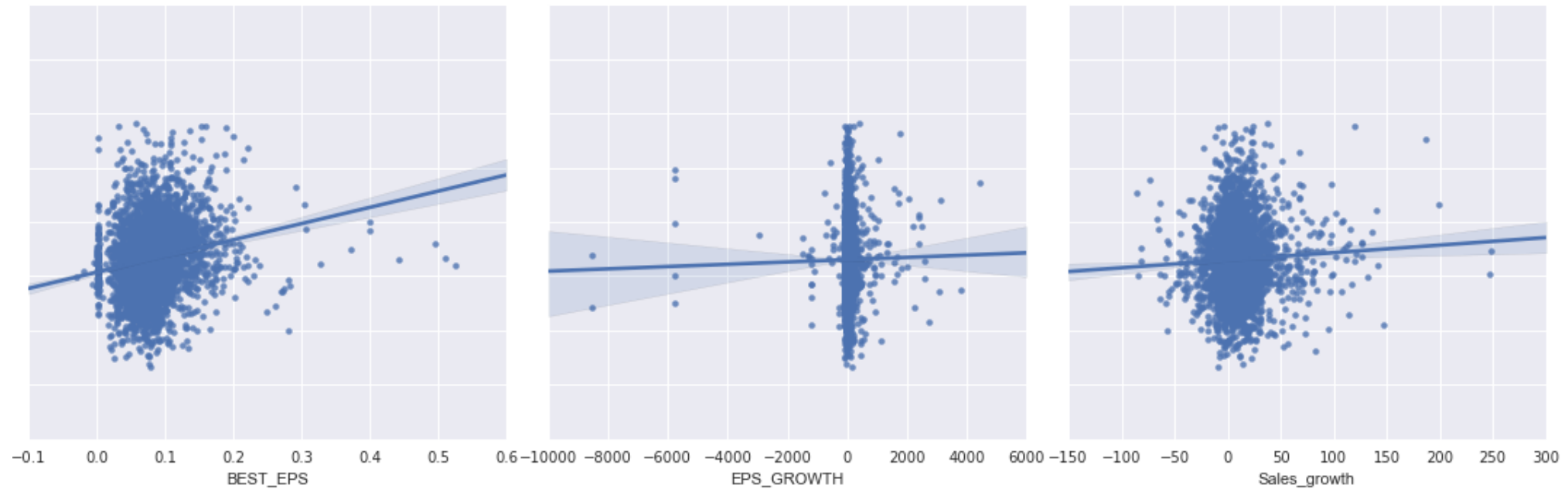 - Outliers cannot be used to make conclusions

Basic analysis

- Results
  - 73% of the data rows have a positive return
  - Average return is 15%

Basic analysis (PE vs return)
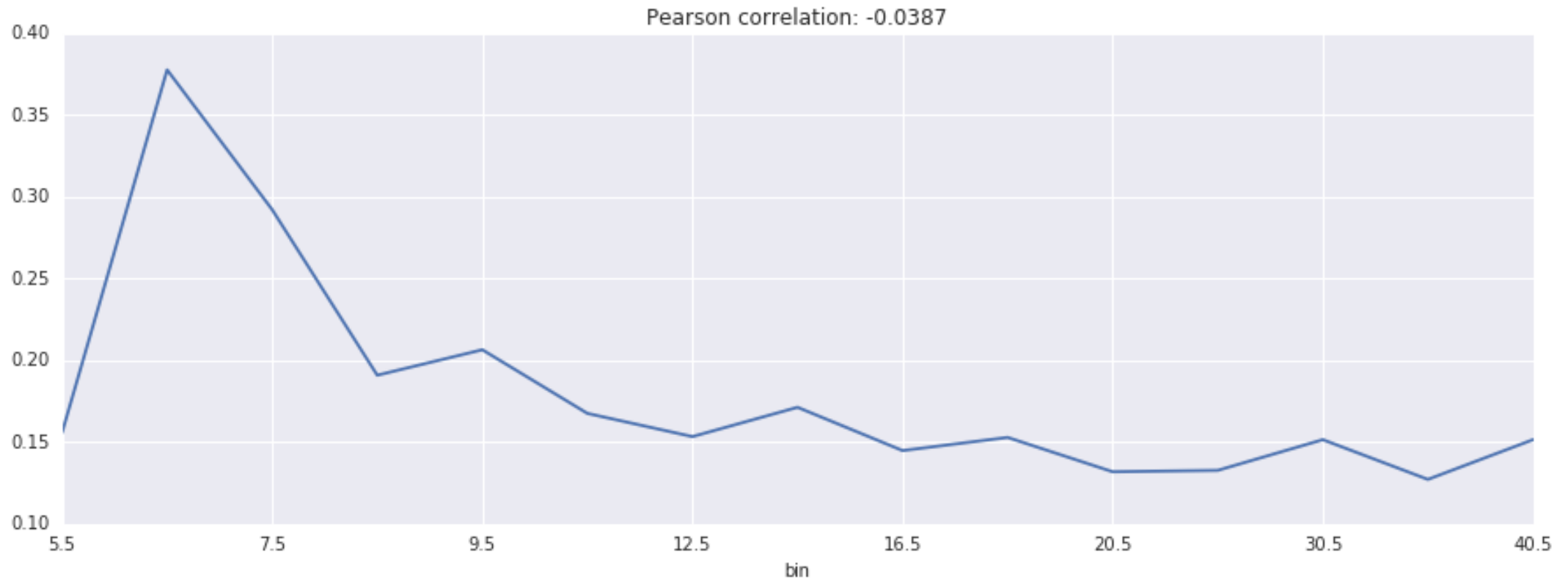
Basic analysis (pairplot)

Basic analysis – correlation with return

DIVIDENDY	-0.104225575413	EPS_GROWTH	0.0117654999064
CUR_MKT_CAP	-0.0720714047831	DebtEQ	0.021673495901
Pricesales	-0.0482633624651	CURratio	0.0268171530037
PXnow	-0.0400802070805	Quick	0.027724935736
oper_margin	-0.0389658987233	PX1YR	0.0292110084654
Prof_margin	-0.00435273379013	Sales_growth	0.0447181953908
PE	0.00295107421611	Rating	0.0460916895817
Pricebook	0.00325980783593	assetTurnover	0.0701406100772
fiveyrAvPriceEarnings	0.0100711755587	BEST_EPS	0.18738466211

Linear Classification

- Splitting and permutation

- Normalize the data (seperately)

- linear SVM and logistic regression

- Accuracy of 51 to 57 %

- Can outperform the market

Linear classifier - Result
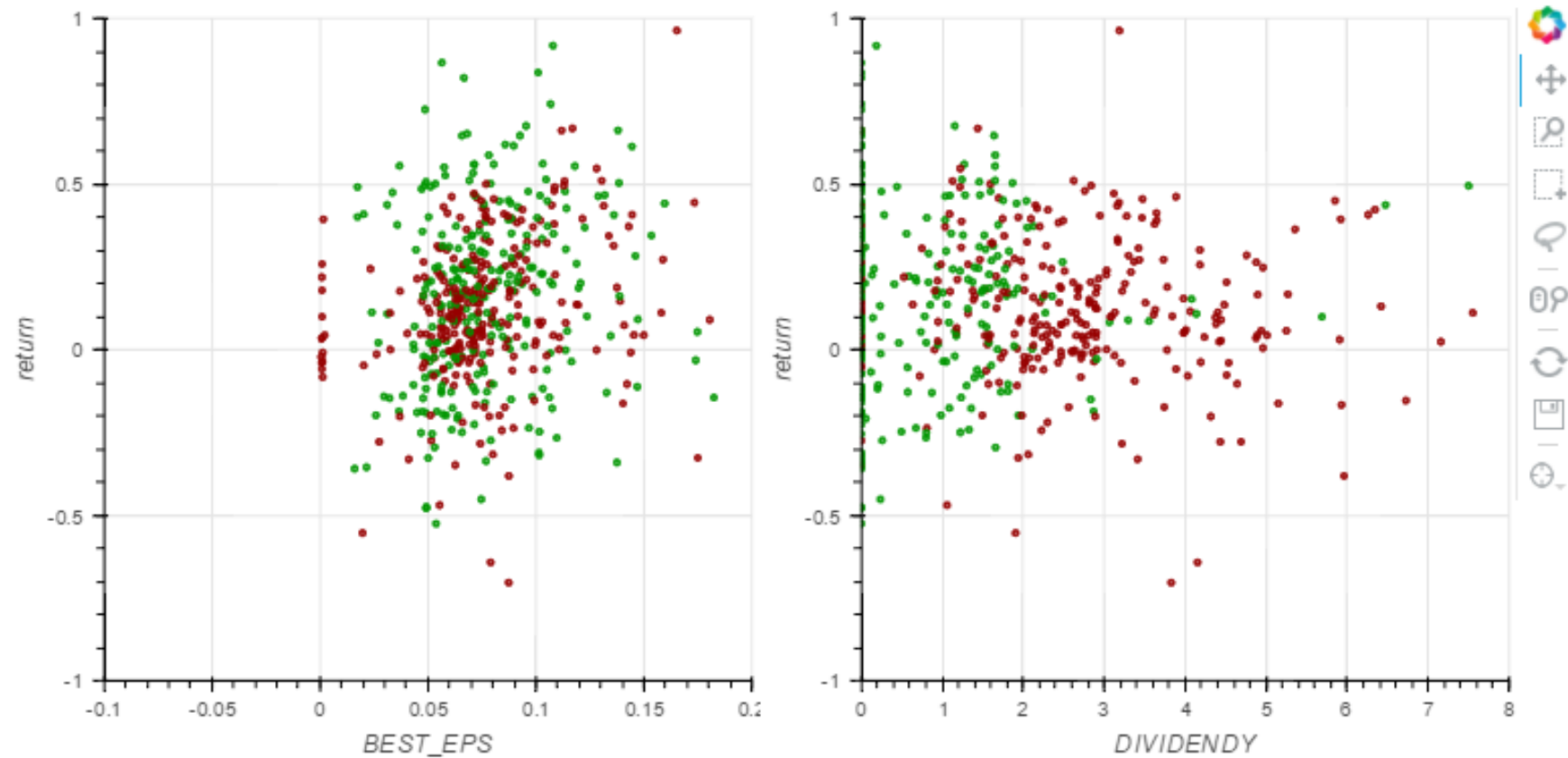
Neural network

- Didn't converge properly

- Tried to adapt:
  - learning rate
  - size of network
  - dropout
  - regularization
  - loss functions
  - amount of runs

Encountered problems

- Python object refference

- Some incomplete data that was not filtered out

- Neural network didn't converge properly

Possible improvements

- Classifiers
  - Feature extraction
  - Adaptations of the algorithms

- Neural network
  - Other loss function

- Code
  - More general code

Conclusion

Hard to classify economical data

It seems that classifiers can beat the market, but some more research is necessary to be sure.

?