Feedback Summary

Project Proposal

* Financial terminology is a bit tough for non-finance readers to follow. For final writeup, include brief definitions and/or leave out details
* How is “strategy” defined for the Novy-Marx and Velikov data? Is it static or dynamic? If dynamic, how might that affect results?
* Tune the latent dimensions on a wider range, and/or explain why three or five dimensions is important to focus on
* Split into train, dev, and test sets, with dev set used for tuning hyperparameters
* From other students on Canvas:
  + Would ease of interpretation be affected using non-linear methods?

Project Update PDF

* Again, need to very clearly define what are the input features, what is being predicted, why a low-dimensional representation of these features is expected to serve as a good replacement for Fama-French. Need to carefully define these things so anyone in the class can follow what we are discussing
* Include more details about the autocorrelation correction (we can put in the report how we ultimately didn’t use it since autocorrelation ended up being insignificant on a monthly time scale)
* Don’t ever want to analyze the whole data set. Should have done the training/dev/test split up front so the test set is as “blind” as possible
* From other students on Canvas:
  + Document details of time series, including frequency, units, etc.
  + Give some more context on the 24 factors in the dataset and the interpretability of various methods that we used
  + Didn’t achieve very high dimensionality reduction, got from 24 down to 20 in some cases before starting to lose a lot of explanatory power
  + Whether we considered other performance metrics besides R^2

Presentation feedback from Canvas

* Address why used single assets vs. portfolios
* Would non-linear dimensionality reduction have worked better if we used a method other than linear regression to measure our results?
* Interpret the learned factors and see if they are sensible?
  + Think we explicitly decided against doing this – it’s actually really challenging and would take as much time as the rest of the project did
  + One of the key advantages to linear over non-linear methods is this interpretability question
* Say more about why non-linear methods didn’t work better than linear. Can we analyze/visualize that somehow?
* Why do you not want too many factors? There was a question on whether we could use R^2 on a held-out set to decide the number of factors, but that’s exactly what we did, so guess it wasn’t clear in the presentation and should be made clear.

Notes and points we need to hit in final report from the feedback:

* Interpretability question. Interpretation of the lower-dimensional feature representation is really challenging and requires a lot of analysis, and is especially hard when using non-linear methods. We decided early on that this was outside the scope of this project, but would make for interesting future work to build on our results.
* Address clearly how we chose the number of factors
* Why we used single assets
* Why we didn’t get an improvement using non-linear methods