<https://machinelearningmastery.com/polynomial-features-transforms-for-machine-learning/>

^explaining polynomial features for PCA

Questions to Patrick

* How can I use PCA to find features to drop (e.g. perform feature extraction)? From the lesson it seems like we just create z-scores for features but then don’t actually identify which features to drop…?
* How to interpret the pca tables and how is it organized in hierarchy of importance?
* I want to use multiple “black-box” models like neural networks and random forest, but then I lose interpretability….without interpretability, I have a hard time solving a business problem. Can you think of any way I could use black box models and still have a good problem statement?

Pca will get you the most important features – but these features are not discernable, they don’t mean anything – you are extracting the directional eigenvectors that explain the most explanatory power –

Code he sent is to take PCA and turn into dataframe of important feature components

People want to say “this causes this”, but predictions are more about

What movie compositions create the optimal level of ratios to get the highest ratios

Given this script, what would the rating be? I can help with my model

Audience would be producers, not the writers, because artists wouldn’t care

**Current Questions**

Notes from Kelly on RF

* N\_estimators – can be as high as we want, but will require a lot of processing power
* Max\_features – try using lower max features
* Max\_depth – try higher max depth

**Next Steps**

* **LinReg** - Use polynomial features (degree 3 is *not* good)
  + Best so far is 1.1938 times better (test size of 25%, degree of 2, components=200)
  + If polynomial features prove to help, try using polynomial features on SVR and Random Forest!!
* **SVR** –
  + try finding new optimal value of C parameter
  + try using smaller test size? 30%?
  + Best so far is 1.203 times better
* **Random Forest**
  + try using smaller test size? 30%?
* Learn Tableau

SVR – with 200 components – 1.1937 times better

**Tableau sheets to build**

1. Build variance/std of rating rank by movie year (already done)
2. Build bar chart with main category being country, and subcategory being genre, and having rating rank be on y-axis (go to 30:30 in tutorial for this)
3. Build geographical chart of countries and their respective rating rank (see 48:18 in video)
4. ~~Build heatmap of most common ratios (as done in EDA notebook~~)
5. Build scatterplot of predictions vs actuals in your SVR model

^will need to use a new dataset with the predictions

1. Build scatterplot of residuals/errors in your SVR model

^will need to use a new dataset with the predictions

Try to combine multiple sheets into a dashboard and then use the filter function to showcase certain conditions (see 55:00 in video)





