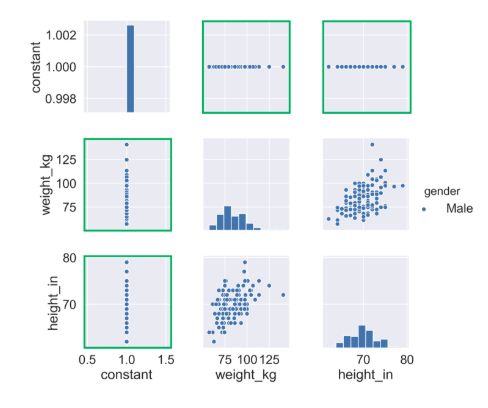
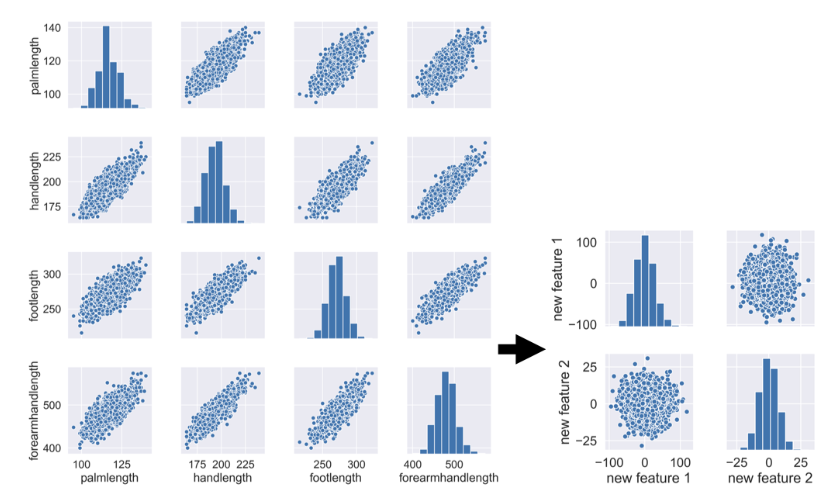


We can see that weight in pound is perfectly correlated with the weight in Kg. So it holds the same information. So we can remove one of the features.



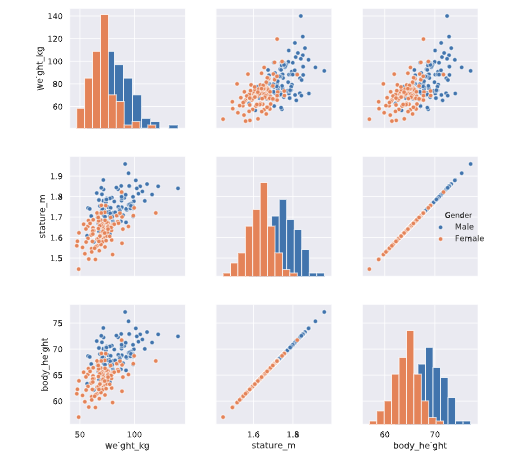
If it is a numeric feature with no variance in the column, such as constant, shown here, we can drop those as well.

Here the observation shows that we don’t see any other gender, so the gender feature can also be dropped.

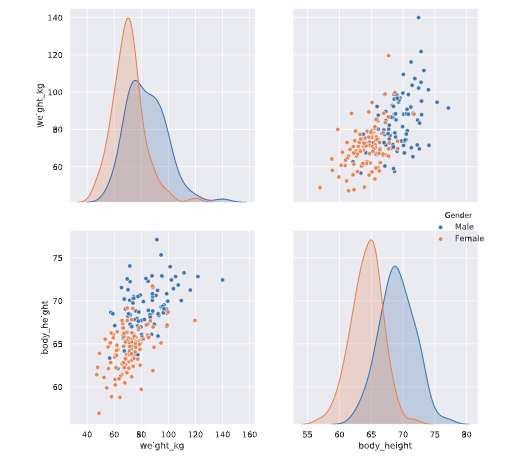


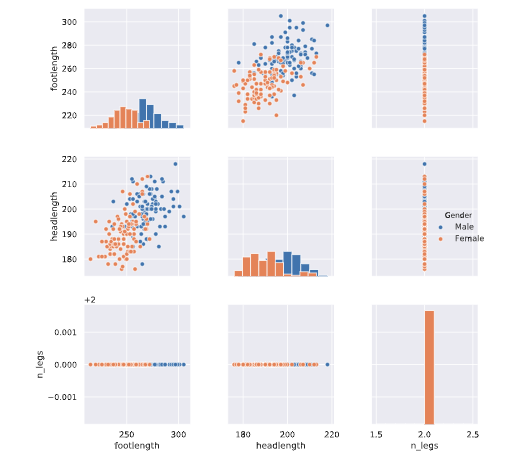
Data set like these, where there is a strong correlation between the features, we can use PCA to reduce to a lower dimension, still keeping the information.

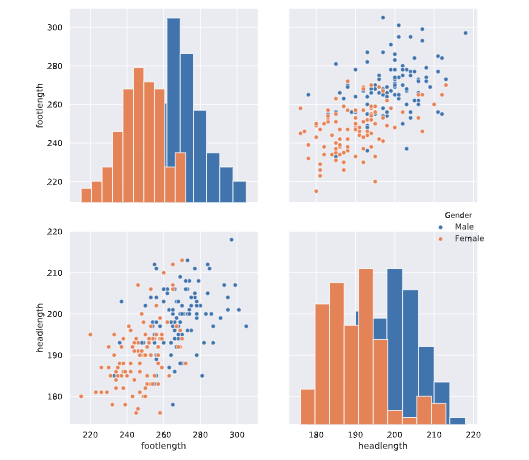
Excerise 1:



Stature\_m and body\_height are perfectly correlated.







feature extraction algorithms try to preserve as much information as possible and are quite good at this. Extracted features can be quite hard to interpret.

T distributed Stochastic Neighbor Embedding (t-SNE):

tSNE will maximize the distance in a two-dimensional space, when there is a lot of distance between the points in the higher dimensional space. Because of this observation that close to one and other may become clusters.

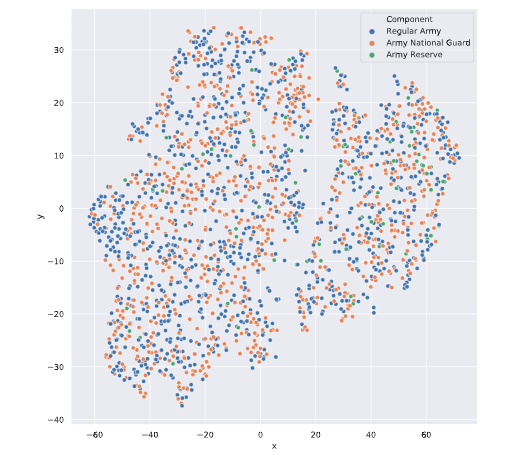
Learning\_rates : 10 -1000

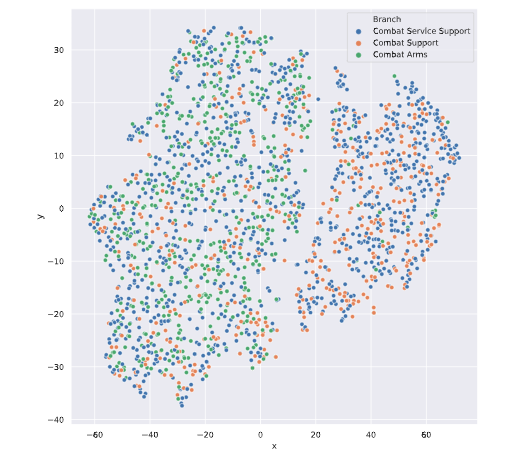
t-SNE helps to visually explore the dataset and help identify the most important drivers of variance in target variable.

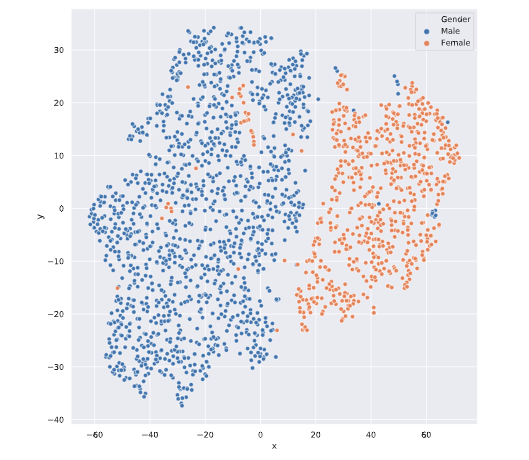
When you want to visually explore the patterns in a high dimensional dataset.

Press

Excerise 4:





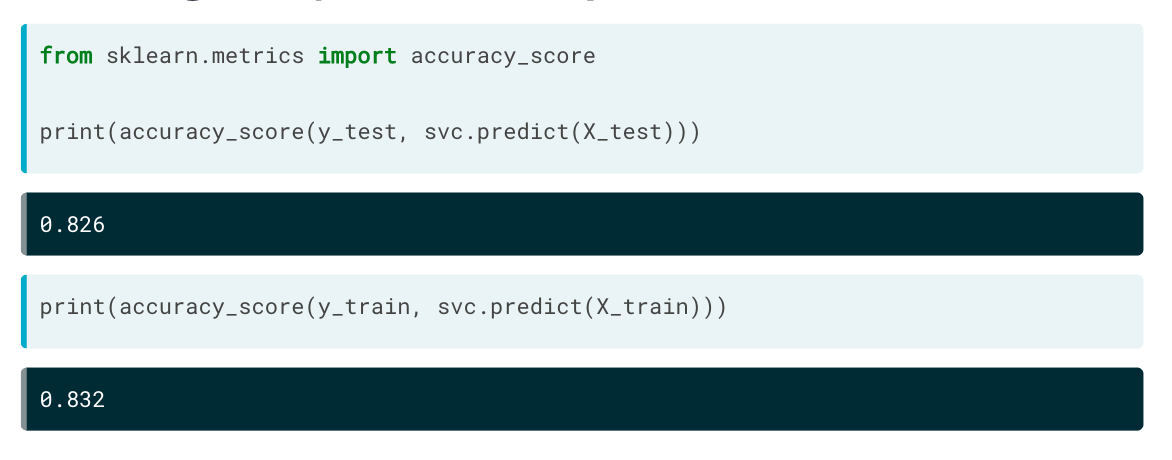


# The cure of dimensionality:

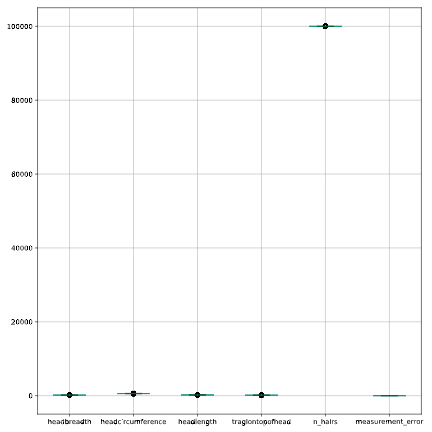
Why models overfit?

How to see if a model is overfit?

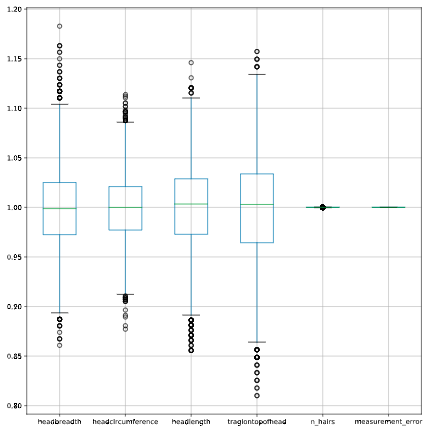
The test set and training set results could be observed: If the accuracy on the training set is much higher than the one in the test set. We can conclude the the model has not generalized well.



As we add more features to improve the model, we need to add more data (rows). Otherwise the model would be able to memorize the data and predict based on that.



Normalize



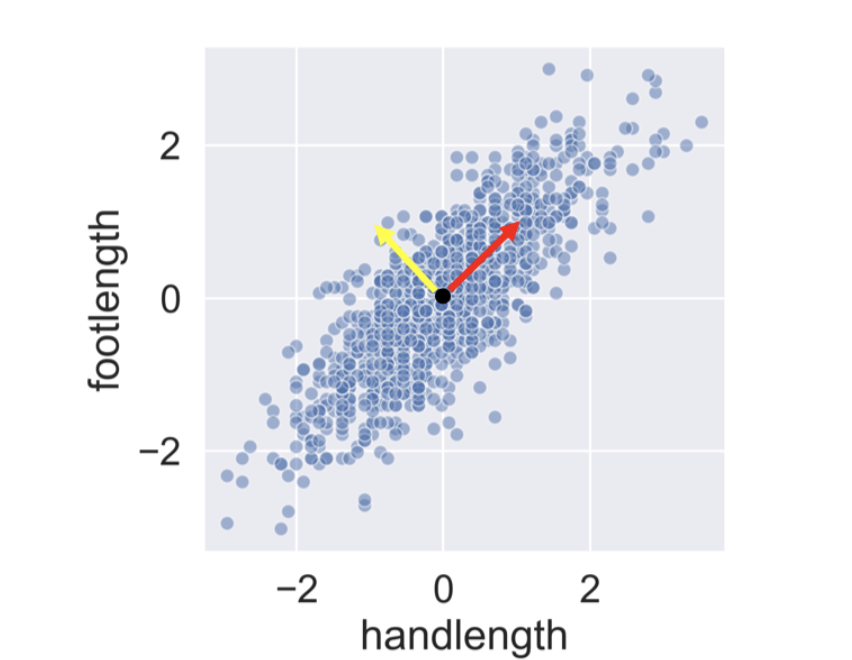
Pairwise Correlation Plot:

Identify the features that are related.



Can be used with any model that gives feature coef or feature importance.

Feature Extraction - PCA:



Here the red vector:

+ive : People have long hands and foot

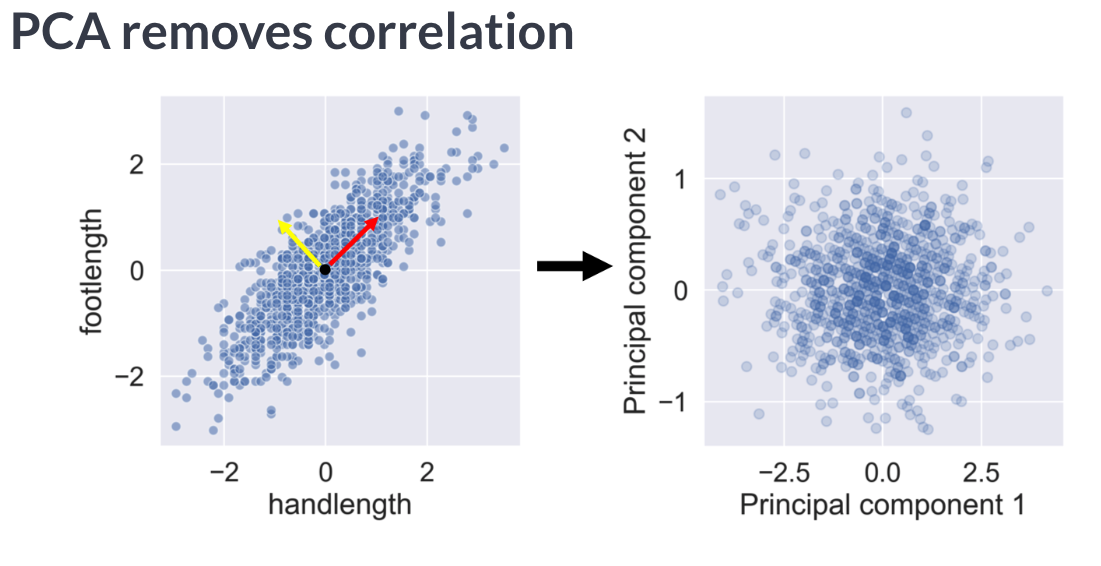
-ive: Short hands and foots.

Yellow vector describes that

+ive: People have a longer foot than the hands

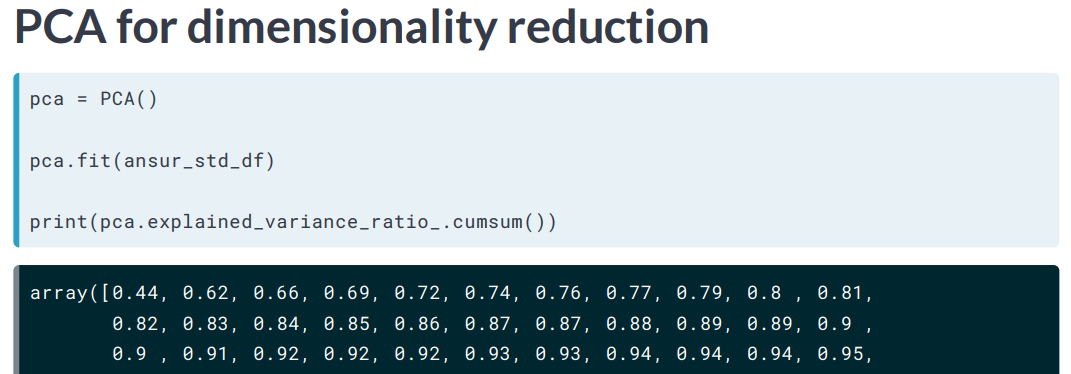
-ive: People have smaller foot then hands.

Always scale the data before you do PCA on it. Otherwise it would underperform.



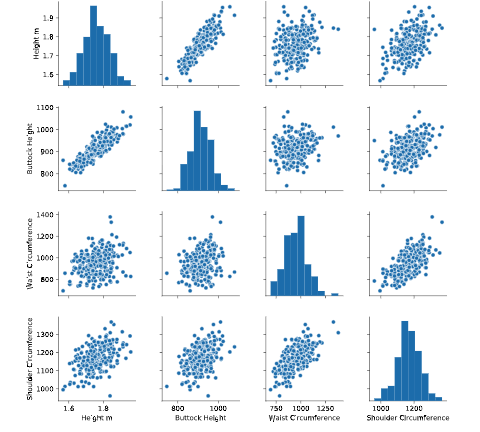
Therefore, it has no duplicate information. The components share no duplicate information. They are ranked from most to least important.

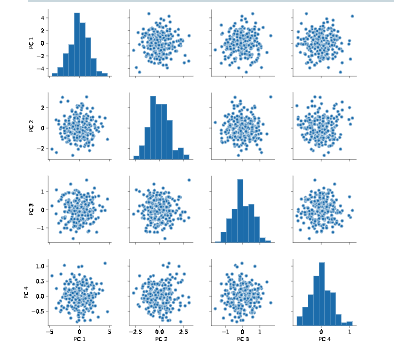
****

****

* Cumulative sum of the explained variance.

3. Calculating Principal Components





Good job! Notice how, in contrast to the input features, none of the principal components are correlated to one another.