Interpreting the Components/Synthetic Features

We have shown that

$$\tilde{\mathbf{X}} = \mathbf{X}V$$

This means that the j^{th} Component (Synthetic feature) $\tilde{\mathbf{X}}_j$

- is a linear combination of the n original features $\mathbf{X}_1, \ldots, \mathbf{X}_n$
- combined with weights V_j

$$\begin{array}{lll}
\tilde{\mathbf{X}} & = & \mathbf{X}V & \text{from the inverse transformation} \\
\tilde{\mathbf{X}}_j & = & (\mathbf{X}V)_j & \text{focus on synthetic feature } j \\
 & = & \begin{pmatrix} \mathbf{X}^{(1)} \cdot V_j \\ \mathbf{X}^{(2)} \cdot V_j \\ \vdots \\ \mathbf{X}^{(m)} \cdot V_j \end{pmatrix} & \text{definition of matrix multiplication}
\end{array}$$

We can try to interpret the meaning of $\tilde{\mathbf{X}}_j$ by looking at the weights V_j

- It is often the case that, for the first component $\tilde{\mathbf{X}}_1$:
    - all n elements of V_1 are approximately equal
    - leading to an interpretation of $\tilde{\mathbf{X}}_1$ as being an *average* across features
        - equally weighted market index when the features are the returns of different equities

It is also often the case that V_j

- contains a subset of indices $P = \{i_1, i_2, \ldots\}$ with high positive values
- contains a subset of indices $N = \{i'_1, i'_2, \ldots\}$ with high negative values
- leading to an interpretation of $\tilde{\mathbf{X}}_j$ as expressing a *dichotomy* between the features in P and those in N
    - For example: the returns of large-cap equities versus small-cap equities

Similarly, we can examine the relationship

$$\mathbf{X} = \tilde{\mathbf{X}} V^T$$

Let's examine the sensitivity of raw feature $\mathbf{X}_j$ to a change in synthetic feature $\tilde{\mathbf{X}}_{j'}$

$$\frac{\partial \mathbf{X}_j}{\partial \tilde{\mathbf{X}}_{j'}}$$

Let $\Delta(j')$ be the length n vector of all 0's except at index j'

$$\Delta(j')_k = \begin{cases} 0 & \text{if} \quad k \neq j' \\ 1 & \text{if} \quad k = j' \end{cases}$$

That is, $\Delta(j')$ represents a unit change to synthetic feature j' while having 0 change to all other features

So a *unit change* in synthetic feature j' results in a change of $V_{j'}^{(j)}$ in feature $\mathbf{X}_j$.

Recall

$$\tilde{\mathbf{X}} = U\Sigma$$

By examining the sensitivity of raw feature $\mathbf{X}_j$ to a change in *standardized* synthetic feature $U_{j'}$

$$\frac{\partial \mathbf{X}_j}{\partial U_{j'}}$$

we instead find the change in raw feature $\mathbf{X}_j$ for a *one standard deviation change* in $\tilde{\mathbf{X}}_{j'}$.

Given the index j' of one component/synthetic feature

- We can vary the index j of raw features
- To see how much a unit change in component j' changes each raw feature j

We can try to interpret component/synthetic feature j' in terms of how it affects raw features.

For example, it is often the case that (indices of) raw feature $\{1, 2, \ldots, n\}$

- contains a subset of indices $P = \{i_1, i_2, \ldots\}$ with positive response to a change in component/synthetic feature j'
- contains a subset of indices $N = \{i'_1, i'_2, \ldots\}$ with negative response to a change in component/synthetic feature j'

We can then interpret component/synthetic feature j' as a feature that creates a dichotomy of behavior among raw features P and N

We will see such dichotomies in our examples for PCA in Finance

- component/synthetic feature 2 affects the short end of the Yield Curve in an opposite manner from the long end of the Yield Curve
- component/synthetic feature 2 affects the returns of Large-Cap equities in an opposite manner from Small-Cap equities

To find a component/synthetic feature j' that expresses a dichotomy, one needs to find sets P and N that have some "natural" meaning

- Each raw feature (e.g., equity) may posses a set of "attributes"
    - Market Cap
    - Cyclical/Non-Cyclical
    - Industry
- By partitioning/sorting raw feature indices according to one such attribute, we might observe a dichotomy

Bottom line

- There is not automatic method to find a good interpretation
- Form a theory as to what attributes each raw feature possesses
- See whether a recognizable pattern of responses to unit change in component/synthetic feature j' emerges
    - When grouping raw features according to common values of an attribute
    - When sorting features according to the level of an attribute