For what Y_1 is Variance(Y_1) maximized?

- The coefficient of the first principal component correspond to the Eigen vector with the maximum Eigen value.

$$\max_{e_1} \; e_1^T \Sigma e \qquad \text{s-t} \quad e_1^T e = 1$$

$$\Rightarrow \max_{e_1} \; \underbrace{e_1^T \Sigma e + \tilde{\lambda}\,[e_1^T e - 1]}_{F(e_1)}$$

Langrangian mulltier based rewrite of original constraind objective.

$$\nabla_{e_1} F = 0$$

$$\Rightarrow \Sigma e_1 + \tilde{\lambda} e_1 = 0$$

$$\Rightarrow \Sigma e_1 = -\tilde{\lambda} e_1 \Rightarrow \boxed{\Sigma e_1 = \lambda e_1}$$

$\Rightarrow$ e_1 is an Eigen vector.

$$\max_{e_1} e_1^T \Sigma e_1 \equiv \max_{e_1, \lambda} e_1^T \lambda e = \max_{\lambda} \lambda$$

s.t e_1 is a Eigen vector.

Choose the e_1 corresponding to the largest Eigen value.

More generally

- The i-th principal component corresponds the i-th largest eigen vector.

The variance for the *i*th principal component is equal to the *i*th eigenvalue.

$$var(Y_i) = \mathrm{var}(e_{i1}X_1 + e_{i2}X_2 + \ldots e_{ip}X_p) = \lambda_i$$

$$\mathrm{cov}(Y_i, Y_j) = 0$$

The proportion of variance explained

- The total variance of X $\sum_{j=1}^{p} \sigma_{jj}^2 = \sum_{j=1}^{p} Var(x_j)$ $\quad \Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \cdots \\ \sigma_{12} & \sigma_{22}^2 & \\ & & \ddots \\ & & & \sigma_{pp}^2 \end{bmatrix}$
- We can show that sum of p Eigen values equals the total variance

show

$$\sum_{j=1}^{p} \sigma_{jj}^2 = \sum_{j=1}^{p} \lambda_j$$

$$\Sigma = \sum_{j=1}^{p} \lambda_j e_j e_j^T$$

$$e_j^T e_j = 1 = \sum_{r=1}^{p} e_{jr}^2$$

$$\lambda_j e_j e_j^T$$

$$\text{Dig}\left\{ \lambda_j \begin{bmatrix} e_{j1} \\ \vdots \\ e_{jp} \end{bmatrix} \begin{bmatrix} e_{j1} & \cdots & e_{jp} \end{bmatrix} \right\} = \lambda_j \begin{bmatrix} e_{j1}^2 & & \\ & \ddots & \\ & & e_{jp}^2 \end{bmatrix}$$

$$\text{Trace}(\lambda_j e_j e_j^T) = \lambda_j$$

Reducing number of dimensions

- Variance explained by first k Eigen values $\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}$

11.3 - Example: Places Rated

Example 11-2: Places Rated

We will use the Places Rated Almanac data (Boyer and Savageau) which rates 329 communities according to nine criteria:

1. Climate and Terrain
2. Housing
3. Health Care & Environment
4. Crime
5. Transportation
6. Education
7. The Arts
8. Recreation
9. Economics

11.3 - Example: Places Rated | STAT 505 (psu.edu)

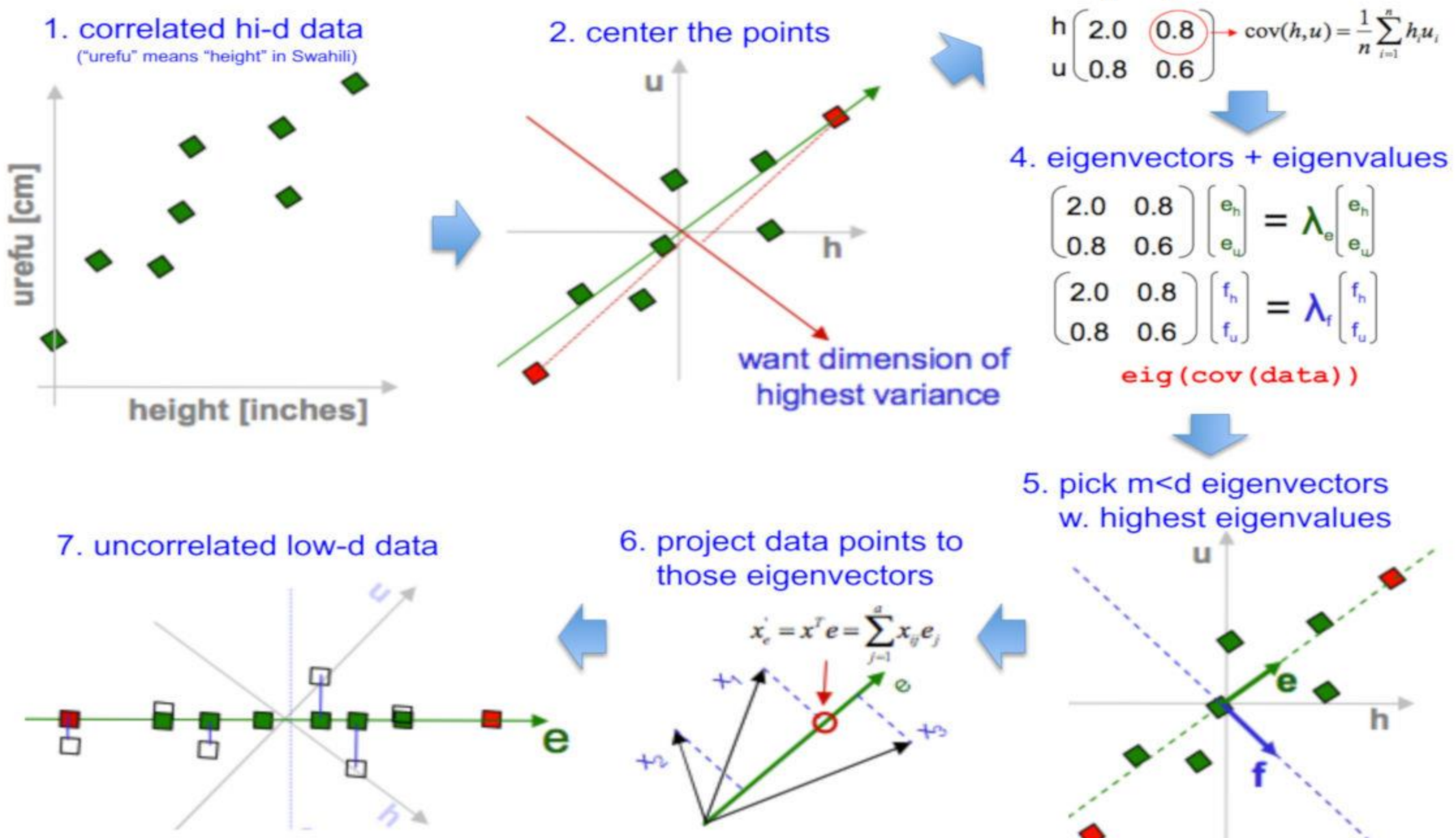
Notes

- The data for many of the variables are strongly skewed to the right.
- The log transformation was used to normalize the data.

More demos

- https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/05.09-Principal-Component-Analysis.ipynb

PCA in a nutshell



T-SNE: T-distributed stochastic neighbourhood embedding

Reading material:

https://www.dailydoseofds.com/formulating-and-implementing-the-t-sne-algorithm-from-scratch/

T-SNE

- Another data projection method, specifically designed for visualizing high dimensional data in two dimensions.
- Preserves local similarities and clusters better than PCA

- Creates non-linear projection