Principal Component Analysis

Prof Wells

STA 395: Machine Learning

February 1st, 2024

Outline

In today's class, we will. . .

- Discuss Principal Component Analysis (PCA) as an example of unsupervised learning
- Investigate matrix formulation for PCA
- Interpret PCA in context

Section 1

Principal Component Analysis

Dimensionality Reduction

Suppose you collect a sample of n observations on p predictors $X_1, \ldots, X_p$, where p is relatively large. Suppose further that some of the predictors are correlated with one another.

Dimensionality Reduction

Suppose you collect a sample of n observations on p predictors $X_1, \ldots, X_p$, where p is relatively large. Suppose further that some of the predictors are correlated with one another.

- Any predictive model for a response Y based on all of the correlated variables will underperform due to instability in parameter estimates.

Dimensionality Reduction

Suppose you collect a sample of n observations on p predictors $X_1, \ldots, X_p$, where p is relatively large. Suppose further that some of the predictors are correlated with one another.

- Any predictive model for a response Y based on all of the correlated variables will underperform due to instability in parameter estimates.

It may be difficult to fit complex models accurately, given limited number of observatiosn compared to predictors.

Dimensionality Reduction

Suppose you collect a sample of n observations on p predictors $X_1, \ldots, X_p$, where p is relatively large. Suppose further that some of the predictors are correlated with one another.

- Any predictive model for a response Y based on all of the correlated variables will underperform due to instability in parameter estimates.

It may be difficult to fit complex models accurately, given limited number of observatiosn compared to predictors.

- If p is larger than n, it may not be possible to fit certain models to the data (for example MLR models cannot be used)

Dimensionality Reduction

Suppose you collect a sample of n observations on p predictors $X_1, \ldots, X_p$, where p is relatively large. Suppose further that some of the predictors are correlated with one another.

- Any predictive model for a response Y based on all of the correlated variables will underperform due to instability in parameter estimates.

It may be difficult to fit complex models accurately, given limited number of observatiosn compared to predictors.

- If p is larger than n, it may not be possible to fit certain models to the data (for example MLR models cannot be used)

One solution is to perform variable selection and drop some less useful predictors.

Dimensionality Reduction

Suppose you collect a sample of n observations on p predictors $X_1, \ldots, X_p$, where p is relatively large. Suppose further that some of the predictors are correlated with one another.

- Any predictive model for a response Y based on all of the correlated variables will underperform due to instability in parameter estimates.

It may be difficult to fit complex models accurately, given limited number of observatiosn compared to predictors.

- If p is larger than n, it may not be possible to fit certain models to the data (for example MLR models cannot be used)

One solution is to perform variable selection and drop some less useful predictors.

- But dropping variables completely loses possible valuable information.

Dimensionality Reduction

Suppose you collect a sample of n observations on p predictors $X_1, \ldots, X_p$, where p is relatively large. Suppose further that some of the predictors are correlated with one another.

- Any predictive model for a response Y based on all of the correlated variables will underperform due to instability in parameter estimates.

It may be difficult to fit complex models accurately, given limited number of observatiosn compared to predictors.

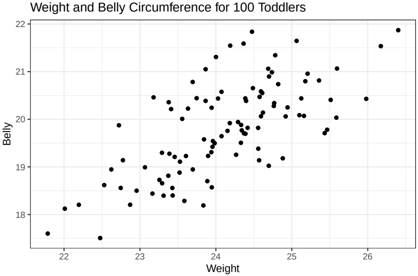
- If p is larger than n, it may not be possible to fit certain models to the data (for example MLR models cannot be used)

One solution is to perform variable selection and drop some less useful predictors.

- But dropping variables completely loses possible valuable information.
- Instead, we can combine variables into new ones that adequately describe the variance in the data, and drop those that have limited utility in explaining that variance.

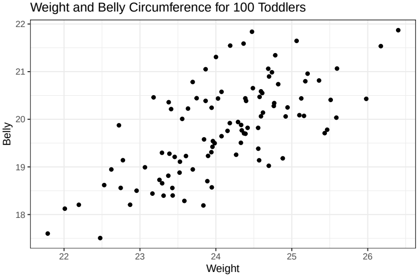
Data Cloud

Consider the weight and belly circumference for a random sample of 100 toddlers.



Data Cloud

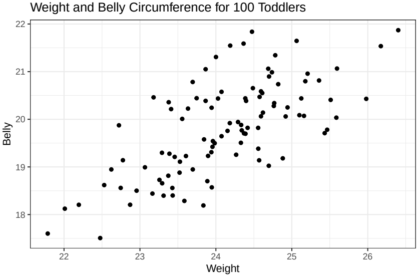
Consider the weight and belly circumference for a random sample of 100 toddlers.



What are the approximate standard deviations of `Weight` and `Belly`?

Data Cloud

Consider the weight and belly circumference for a random sample of 100 toddlers.

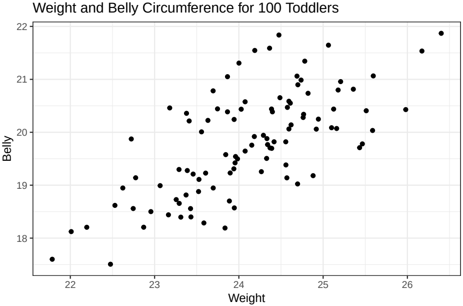


What are the approximate standard deviations of `Weight` and `Belly`?

```
##   sd_Weight  sd_Belly
## 1 0.8981994 0.9843542
```

Data Cloud

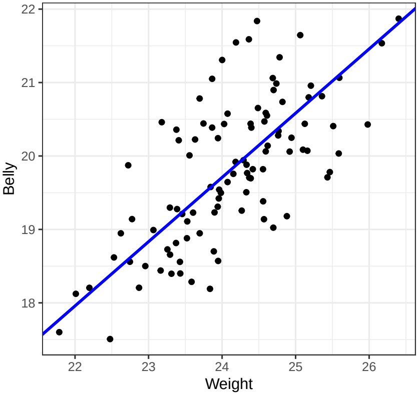
Consider the weight and belly circumference for a random sample of 100 toddlers.



But do either of these variables represent the direction of *maximal* variation in the data?

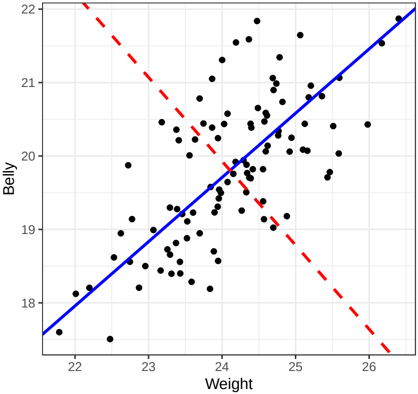
Maximal Variation

Can we find a line along which the observations vary the most?



Variation Decomposition

How much variation occurs perpendicular to this line?



First Principal Component

The first principal component of centered variables $X_1, \ldots, X_p$ is a normalized linear combination with largest variance, taking the form:

$$Z_1 = \phi_{11} X_1 + \cdots + \phi_{p1} X_p \qquad \text{with} \sum \phi_{i1}^2 = 1$$

First Principal Component

The first principal component of centered variables $X_1, \ldots, X_p$ is a normalized linear combination with largest variance, taking the form:

$$Z_1 = \phi_{11}X_1 + \cdots + \phi_{p1}X_p \qquad \text{with } \sum \phi_{i1}^2 = 1$$

- The vector $\phi_1 = \begin{pmatrix} \phi_{11} & \cdots & \phi_{p1} \end{pmatrix}^T$ is called the **loading** of the 1st PC

First Principal Component

The first principal component of centered variables $X_1, \ldots, X_p$ is a normalized linear combination with largest variance, taking the form:

$$Z_1 = \phi_{11}X_1 + \cdots + \phi_{p1}X_p \qquad \text{with } \sum \phi_{i1}^2 = 1$$

- The vector $\phi_1 = \begin{pmatrix} \phi_{11} & \cdots & \phi_{p1} \end{pmatrix}^T$ is called the **loading** of the 1st PC
    - The loading $\phi_1 \in \mathbb{R}^p$ points in the direction in feature space along which the data varies the most.

First Principal Component

The first principal component of centered variables $X_1, \dots, X_p$ is a normalized linear combination with largest variance, taking the form:

$$Z_1 = \phi_{11} X_1 + \cdots + \phi_{p1} X_p \qquad \text{with } \sum \phi_{i1}^2 = 1$$

- The vector $\phi_1 = \begin{pmatrix} \phi_{11} & \cdots & \phi_{p1} \end{pmatrix}^T$ is called the **loading** of the 1st PC
  - The loading $\phi_1 \in \mathbb{R}^p$ points in the direction in feature space along which the data varies the most.
- The values

$$z_{i1} = \phi_1^T x_i = \phi_{11} x_{i1} + \phi_{21} x_{i2} + \dots \phi_{p1} x_{ip}$$

  for $1 \le i \le n$ are called the **scores** of the 1st PC

First Principal Component

The first principal component of centered variables $X_1, \ldots, X_p$ is a normalized linear combination with largest variance, taking the form:

$$Z_1 = \phi_{11} X_1 + \cdots + \phi_{p1} X_p \qquad \text{with } \sum \phi_{i1}^2 = 1$$

- The vector $\phi_1 = \begin{pmatrix} \phi_{11} & \cdots & \phi_{p1} \end{pmatrix}^T$ is called the **loading** of the 1st PC
  - The loading $\phi_1 \in \mathbb{R}^p$ points in the direction in feature space along which the data varies the most.
- The values

  $$z_{i1} = \phi_1^T x_i = \phi_{11} x_{i1} + \phi_{21} x_{i2} + \ldots \phi_{p1} x_{ip}$$

  for $1 \leq i \leq n$ are called the **scores** of the 1st PC
  - The score z_{i1} is the coordinate of the ith observation x_i in the 1st PC

Optimization Problem

The 1st PC has loading ϕ_1 whose scores $z_{i1} = \phi_1 x_i$ have largest possible variance

Optimization Problem

The 1st PC has loading ϕ_1 whose scores $z_{i1} = \phi_1 x_i$ have largest possible variance

- As Z_1 is centered, the first PC loading vector ϕ_1 solves the following optimization problem:

$$\begin{aligned}\phi_1 =& \text{argmax}_{\|\phi_1\|^2=1} \text{Var}(Z_1) \\ =& \text{argmax}_{\|\phi_1\|^2=1} \left\{ \frac{1}{n} \sum_{i=1}^{n} z_{i1}^2 \right\} = \text{argmax}_{\|\phi_1\|^2=1} \left\{ \frac{1}{n} \sum_{i=1}^{n} (\phi_1^T x_i)^2 \right\}\end{aligned}$$

Optimization Problem

The 1st PC has loading ϕ_1 whose scores $z_{i1} = \phi_1 x_i$ have largest possible variance

- As Z_1 is centered, the first PC loading vector ϕ_1 solves the following optimization problem:

$$\begin{aligned}\phi_1 =& \operatorname{argmax}_{\|\phi_1\|^2=1} \mathrm{Var}(Z_1) \\ =& \operatorname{argmax}_{\|\phi_1\|^2=1} \left\{ \frac{1}{n} \sum_{i=1}^{n} z_{i1}^2 \right\} = \operatorname{argmax}_{\|\phi_1\|^2=1} \left\{ \frac{1}{n} \sum_{i=1}^{n} (\phi_1^T x_i)^2 \right\}\end{aligned}$$

- But note by matrix multiplication,

$$\sum_{i=1}^{n} (\phi_1^T x_i)^2 = \phi_1^T X^T X \phi_1$$

Optimization Problem

The 1st PC has loading ϕ_1 whose scores $z_{i1} = \phi_1 x_i$ have largest possible variance

- As Z_1 is centered, the first PC loading vector ϕ_1 solves the following optimization problem:

$$\begin{aligned}\phi_1 =& \text{argmax}_{\|\phi_1\|^2=1} \text{Var}(Z_1) \\ =& \text{argmax}_{\|\phi_1\|^2=1} \left\{ \frac{1}{n} \sum_{i=1}^{n} z_{i1}^2 \right\} = \text{argmax}_{\|\phi_1\|^2=1} \left\{ \frac{1}{n} \sum_{i=1}^{n} (\phi_1^T x_i)^2 \right\}\end{aligned}$$

- But note by matrix multiplication,

$$\sum_{i=1}^{n} (\phi_1^T x_i)^2 = \phi_1^T X^T X \phi_1$$

- And so equivalently, the 1st PC has normalized loading ϕ_1 which maximizes $\phi_1^T X^T X \phi_1$

Optimization Problem

The 1st PC has loading ϕ_1 whose scores $z_{i1} = \phi_1 x_i$ have largest possible variance

- As Z_1 is centered, the first PC loading vector ϕ_1 solves the following optimization problem:

$$\begin{aligned}\phi_1 =& \mathrm{argmax}_{\|\phi_1\|^2=1} \mathrm{Var}(Z_1)\\ =& \mathrm{argmax}_{\|\phi_1\|^2=1} \left\{ \frac{1}{n} \sum_{i=1}^{n} z_{i1}^2 \right\} = \mathrm{argmax}_{\|\phi_1\|^2=1} \left\{ \frac{1}{n} \sum_{i=1}^{n} (\phi_1^T x_i)^2 \right\}\end{aligned}$$

- But note by matrix multiplication,

$$\sum_{i=1}^{n} (\phi_1^T x_i)^2 = \phi_1^T X^T X \phi_1$$

- And so equivalently, the 1st PC has normalized loading ϕ_1 which maximizes $\phi_1^T X^T X \phi_1$
- A standard result in linear algebra:
  - The maximal value of $\phi_1^T X^T X \phi_1$ is the largest eigenvalue of the covariance matrix $X^T X$ and occurs when ϕ is the associated normalized eigenvector.

Additional Principal Components

The second principal component Z_2 is the linear combination of $X_1, \ldots, X_p$ that has maximal variance among all lin. combos. that are uncorrelated with Z_1, and takes the form

$$Z_2 = \phi_{12} X_1 + \cdots + \phi_{p2} X_p \text{ with } \|\phi_2\|^2 = 1 \text{ and } \mathrm{Corr}(Z_1, Z_2) = 0$$

Additional Principal Components

The second principal component Z_2 is the linear combination of $X_1, \ldots, X_p$ that has maximal variance among all lin. combos. that are uncorrelated with Z_1, and takes the form

$$Z_2 = \phi_{12} X_1 + \cdots + \phi_{p2} X_p \text{ with } \|\phi_2\|^2 = 1 \text{ and } \mathrm{Corr}(Z_1, Z_2) = 0$$

- Z_2 can also be obtained by projecting all observations onto the hyperplane perpendicular to ϕ_1 and finding the 1st principal component of the resulting data set.

Additional Principal Components

The second principal component Z_2 is the linear combination of $X_1, \dots, X_p$ that has maximal variance among all lin. combos. that are uncorrelated with Z_1, and takes the form

$$Z_2 = \phi_{12} X_1 + \cdots + \phi_{p2} X_p \text{ with } \|\phi_2\|^2 = 1 \text{ and } \mathrm{Corr}(Z_1, Z_2) = 0$$

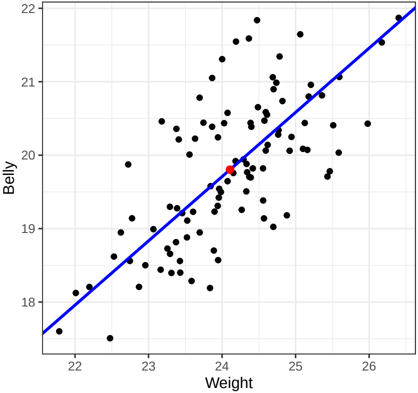
- Z_2 can also be obtained by projecting all observations onto the hyperplane perpendicular to ϕ_1 and finding the 1st principal component of the resulting data set.

In general, the kth principal component is a linear combination that has maximal variance among all combos that are uncorrelated with $Z_1, \dots, Z_{k-1}$

$$\begin{aligned} Z_k = &\phi_{1k} X_1 + \cdots + \phi_{pk} X_p \\ &\text{with } \|\phi_k\|^2 = 1 \text{ and } \mathrm{Corr}(Z_j, Z_k) = 0, \text{ for all } 1 \le j \le k-1 \end{aligned}$$

PCA Visual

The first principal component

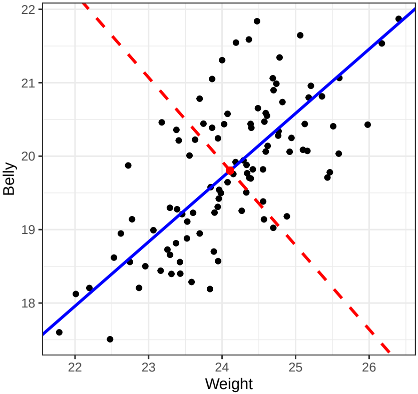


$$Z_1 = 0.67 \cdot (\text{Weight} - 24.1) + 0.75 \cdot (\text{Belly} - 19.8)$$

$$\phi_1 = \begin{pmatrix} 0.67 & 0.75 \end{pmatrix}^T$$

PCA Visual

The 2nd principal component is perpendicular to the 1st:

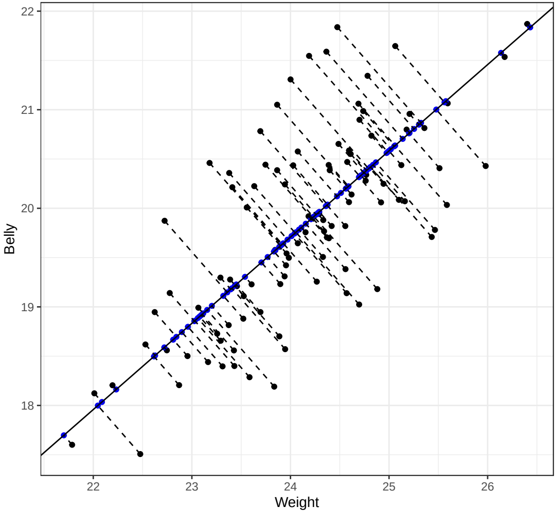


$$Z_2 = 0.75 \cdot (\text{Weight} - 24.1) - 0.67 \cdot (\text{Belly} - 19.8)$$

$$\phi_2 = \begin{pmatrix} 0.75 & -.67 \end{pmatrix}^T$$

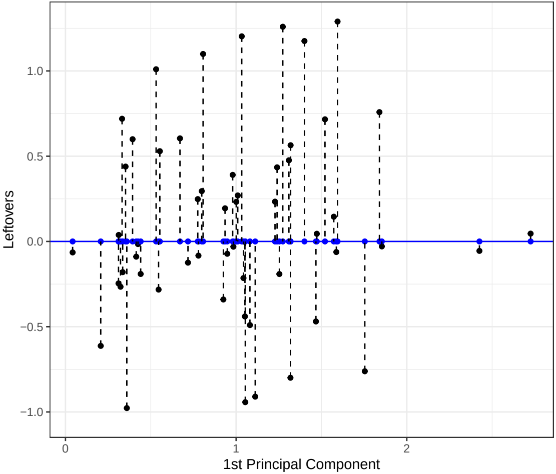
PCA Visual

What is leftover?



PCA Visual

Rotating axes so they lie along principal components:



Two Geometric Perspective

Perspective 1: Principal components are directions in feature space along which data vary the most.

Two Geometric Perspective

Perspective 1: Principal components are directions in feature space along which data vary the most.

Perspective 2: The first M principal components are the best M-dimensional approximation to the p-dimensional data set.

Two Geometric Perspective

Perspective 1: Principal components are directions in feature space along which data vary the most.

Perspective 2: The first M principal components are the best M-dimensional approximation to the p-dimensional data set.

- Observe that the loading vector ϕ_1 generates the line in p-dim space that is *closest* to the n observations in the data set.

Two Geometric Perspective

Perspective 1: Principal components are directions in feature space along which data vary the most.

Perspective 2: The first M principal components are the best M-dimensional approximation to the p-dimensional data set.

- Observe that the loading vector ϕ_1 generates the line in p-dim space that is *closest* to the n observations in the data set.
- Together, the loading vectors ϕ_1, ϕ_2 generate the 2D plane in p-dim space that is closest to the n observations

Two Geometric Perspective

Perspective 1: Principal components are directions in feature space along which data vary the most.

Perspective 2: The first M principal components are the best M-dimensional approximation to the p-dimensional data set.

- Observe that the loading vector ϕ_1 generates the line in p-dim space that is *closest* to the n observations in the data set.
- Together, the loading vectors ϕ_1, ϕ_2 generate the 2D plane in p-dim space that is closest to the n observations
- Generally, the first M loading vectors $\phi_1, \ldots, \phi_M$ generate an M-dimensional hyperplane in p-dim space that is closest to the n observations.

Two Geometric Perspective

Perspective 1: Principal components are directions in feature space along which data vary the most.

Perspective 2: The first M principal components are the best M-dimensional approximation to the p-dimensional data set.

- Observe that the loading vector ϕ_1 generates the line in p-dim space that is *closest* to the n observations in the data set.
- Together, the loading vectors ϕ_1, ϕ_2 generate the 2D plane in p-dim space that is closest to the n observations
- Generally, the first M loading vectors $\phi_1, \ldots, \phi_M$ generate an M-dimensional hyperplane in p-dim space that is closest to the n observations.

$$x_{ij} \approx \sum_{m=1}^{M} z_{im}\phi_{jm} \qquad \text{where } z_{im} = \phi_m^T x_i = \phi_{1m}x_{im} + \cdots + \phi_{pm}x_{ip}$$

Properties of PCA

How much information is lost when we project the data set onto the hyperplane spanned by the first M principal component loading vectors?

Properties of PCA

How much information is lost when we project the data set onto the hyperplane spanned by the first M principal component loading vectors?

- The *Total Variance* (TV) of the data set is

$$\mathrm{TV} = \sum_{j=1}^{p} \mathrm{Var}(X_j) = \sum_{j=1}^{p} \frac{1}{n} \sum_{i=1}^{n} x_{ij}^2$$

Properties of PCA

How much information is lost when we project the data set onto the hyperplane spanned by the first M principal component loading vectors?

- The *Total Variance* (TV) of the data set is

$$\mathrm{TV} = \sum_{j=1}^{p} \mathrm{Var}(X_j) = \sum_{j=1}^{p} \frac{1}{n} \sum_{i=1}^{n} x_{ij}^2$$

- While the variance explained by the mth principal component V_m is

$$V_m = \frac{1}{n} \sum_{i=1}^{n} z_{im}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{p} \phi_{jm} x_{ij} \right)^2$$

Properties of PCA

How much information is lost when we project the data set onto the hyperplane spanned by the first M principal component loading vectors?

- The *Total Variance* (TV) of the data set is

$$\mathrm{TV} = \sum_{j=1}^{p} \mathrm{Var}(X_j) = \sum_{j=1}^{p} \frac{1}{n} \sum_{i=1}^{n} x_{ij}^2$$

- While the variance explained by the mth principal component V_m is

$$V_m = \frac{1}{n} \sum_{i=1}^{n} z_{im}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{p} \phi_{jm} x_{ij} \right)^2$$

- Thus, the *Proportion of Variance Explained* by the mth principal component PVE_m is

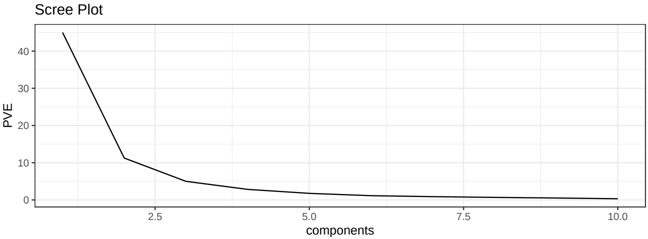
$$\mathrm{PVE}_m = \frac{V_m}{TV} = \frac{\sum_{i=1}^{n} \left(\sum_{j=1}^{p} \phi_{jm} x_{ij} \right)^2}{\sum_{j=1}^{p} \sum_{i=1}^{n} x_{ij}^2}$$

How many principal components?

We can create the *scree plot* of PVE_m versus *m* and look for the point of diminishing returns (called the *elbow*)

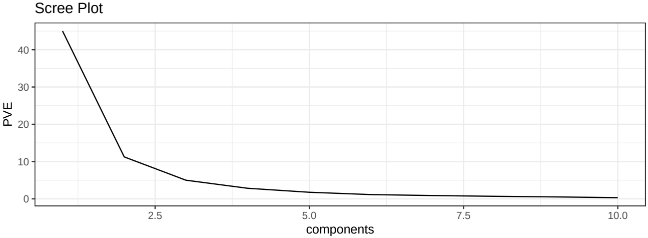
How many principal components?

We can create the *scree plot* of PVE_m versus m and look for the point of diminishing returns (called the *elbow*)



How many principal components?

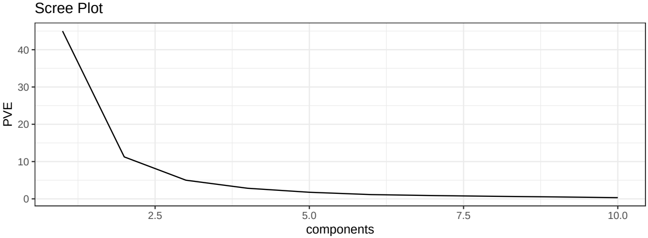
We can create the *scree plot* of PVE_m versus *m* and look for the point of diminishing returns (called the *elbow*)



- Here, 2 or 3 PCs seem sufficient.

How many principal components?

We can create the *scree plot* of PVE_m versus *m* and look for the point of diminishing returns (called the *elbow*)



- Here, 2 or 3 PCs seem sufficient.

Alternative: look data structure present in the first several principal components, and then add more components until the structures of interest stops changing

Section 2

PCA Example

Perfumes

12 perfumers were asked to rate 12 perfumes on 11 scent adjectives

```
##  [1] "spicy"    "heady"    "fruity"   "green"    "vanilla"  "floral"
##  [7] "woody"    "citrus"   "marine"   "greedy"   "oriental"
```

Perfumes

12 perfumers were asked to rate 12 perfumes on 11 scent adjectives

```
##  [1] "spicy"    "heady"    "fruity"   "green"    "vanilla"  "floral"
##  [7] "woody"    "citrus"   "marine"   "greedy"   "oriental"
```

Each was rated on a scale of 1-10, and ratings for each perfume were averaged across experts.

```
## # A tibble: 6 x 12
##   perfume      spicy heady fruity green vanilla floral woody citrus marine greedy
##   <chr>        <dbl> <dbl>  <dbl> <dbl>   <dbl>  <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1 "Angel"      3.22  8.26   1.9   0.133   7.75    2.09 1.05   0.142 0.125   8.28
## 2 "Aromatics~ 7.41  8.17   0.575 0.35    1.75    3.71 3.39   0.375 0.0583  0.258
## 3 "Chanel N5" 3.93  8.42   1.18  0.5     1.73    4.66 1.02   0.6   0.05    0.458
## 4 "Cin\xe9ma" 0.983 2.07   5.2   0.267   4.18    5.32 1.25   0.775 1.02    3.66
## 5 "Coco Made~ 0.925 0.717  4.58  1.2     2.02    7.31 1.13   1.17  1.14    2.72
## 6 "J'adore E~ 0.108 1.03   6.85  1.62    0.183   8.51 0.925  2.13  1.91    1.47
## # i 1 more variable: oriental <dbl>
```

Fitting the PCA

We use software (Python, R, etc.) to fit a PCA, which will contain a number of useful quantities

```
## [1] "sdev"     "rotation" "center"   "scale"    "x"
```

Fitting the PCA

We use software (Python, R, etc.) to fit a PCA, which will contain a number of useful quantities

```
## [1] "sdev"     "rotation" "center"   "scale"    "x"
```

The rotation value contains the principal component loadings

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
spicy	-0.32	-0.31	0.15	-0.10	0.21	0.00	0.29	-0.17	0.12	-0.77	0.00
heady	-0.35	-0.11	0.25	0.16	-0.21	-0.47	0.36	0.48	0.19	0.22	-0.23
fruity	0.34	0.15	-0.36	-0.17	0.26	-0.49	0.17	-0.21	-0.01	-0.07	-0.57
green	0.30	-0.15	0.62	0.27	0.36	0.31	0.05	-0.06	-0.04	0.14	-0.42
vanilla	-0.19	0.51	0.17	-0.28	-0.09	0.17	-0.29	0.40	-0.26	-0.32	-0.38
floral	0.34	-0.20	-0.27	0.07	-0.17	0.28	-0.13	0.39	0.63	-0.22	-0.18
woody	-0.25	-0.37	-0.14	-0.59	0.48	0.15	-0.10	0.22	0.04	0.35	-0.05
citrus	0.33	-0.18	0.38	-0.18	0.07	-0.54	-0.51	0.14	0.04	-0.17	0.28
marine	0.32	-0.08	0.27	-0.61	-0.51	0.12	0.39	-0.13	-0.02	0.06	0.01
greedy	-0.09	0.58	0.23	-0.16	0.26	-0.02	0.09	-0.17	0.65	0.11	0.20
oriental	-0.35	-0.18	0.08	-0.04	-0.35	-0.05	-0.47	-0.51	0.25	0.12	-0.39

Visualize

How can we visualize?

Visualize

How can we visualize?

- Representing the data set itself requires 11 dimensions.

Visualize

How can we visualize?

- Representing the data set itself requires 11 dimensions.
- Representing all pairwise structure requires $\binom{55}{2} = 55$ pairwise scatterplots

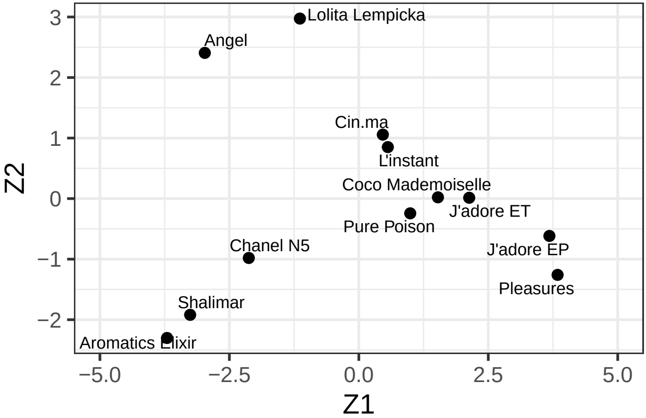
Visualize

How can we visualize?

- Representing the data set itself requires 11 dimensions.
- Representing all pairwise structure requires $\binom{55}{2} = 55$ pairwise scatterplots

We can use principal components to focus our attention on small dimensional representation which describes most of the structure.

Scatterplot



Interpretation

Effectively interpreting principal the loading vector for principal components usually requires domain knowledge. But we can try!

Interpretation

Effectively interpreting principal the loading vector for principal components usually requires domain knowledge. But we can try!

What does Z_1 represent? (i.e for what values of x is Z_1 large? small?)

```
##     spicy     heady    fruity     green   vanilla    floral     woody    citrus
##    -0.324    -0.352     0.340     0.304    -0.192     0.344    -0.252     0.330
##    marine    greedy  oriental
##     0.322    -0.085    -0.353
```

Interpretation

Effectively interpreting principal the loading vector for principal components usually requires domain knowledge. But we can try!

What does Z_1 represent? (i.e for what values of x is Z_1 large? small?)

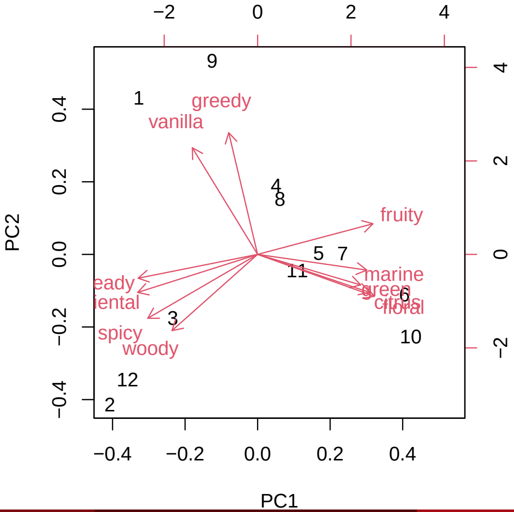
```
##     spicy     heady    fruity     green   vanilla    floral     woody    citrus
##    -0.324    -0.352     0.340     0.304    -0.192     0.344    -0.252     0.330
##    marine    greedy  oriental
##     0.322    -0.085    -0.353
```

What does Z_2 represent?

```
##     spicy     heady    fruity     green   vanilla    floral     woody    citrus
##    -0.307    -0.114     0.147    -0.147     0.512    -0.201    -0.366    -0.183
##    marine    greedy  oriental
##    -0.075     0.584    -0.182
```

Another Visualization

We can create a **biplot**, which shows the location of each observation in the first 2 principal components, along arrows indicating the *loading* vectors.



Scree Plot

The scree plot can be used to find the "elbow"

- In this case, 3 principal components might be optimal