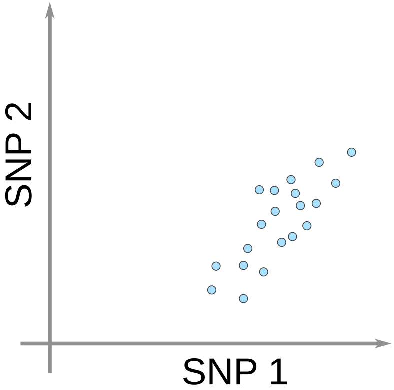
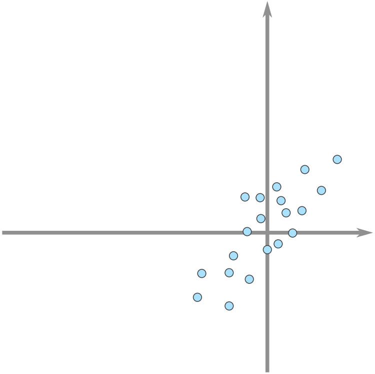
April 22, 2021

Principal Component Analysis



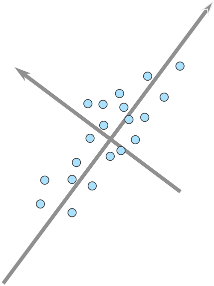
- Raw SNP data $\mathbf{X}$; x_{ij}

Principal Component Analysis



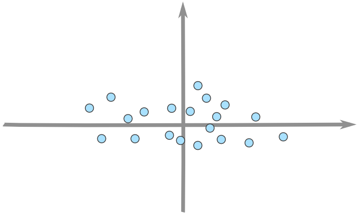
- Raw SNP data $\mathbf{X}$; x_{ij}
- Centering
  $\mathbf{Y} = \mathbf{CX}$; $y_{ij} = x_{ij} - \mu_j$

Principal Component Analysis



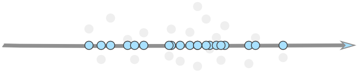
- Raw SNP data $\mathbf{X}$; x_{ij}
- Centering
  $\mathbf{Y} = \mathbf{CX}$; $y_{ij} = x_{ij} - \mu_j$
- Rotation $\mathbf{Y} = \underbrace{\mathbf{P}}_{\text{PCs}} \underbrace{\mathbf{L}}_{\text{Rotation}}$

Principal Component Analysis



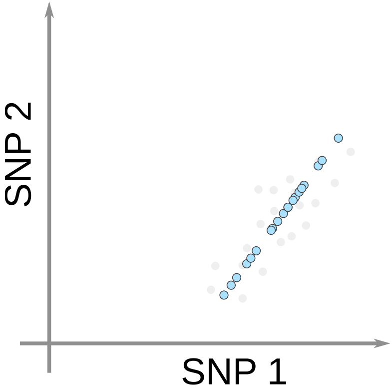
- Raw SNP data $\mathbf{X}$; x_{ij}
- Centering
  $\mathbf{Y} = \mathbf{CX}$; $y_{ij} = x_{ij} - \mu_j$
- Rotation $\mathbf{Y} = \mathbf{PL}$

Principal Component Analysis



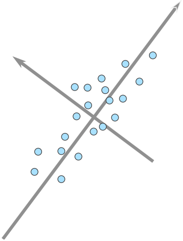
- Raw SNP data $\mathbf{X}$; x_{ij}
- Centering
  $\mathbf{Y} = \mathbf{CX}$; $y_{ij} = x_{ij} - \mu_j$
- Rotation $\mathbf{Y} = \mathbf{PL}$
- Truncation $\hat{\mathbf{P}} = \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \hline \\ \vdots \\ \mathbf{p}_n \end{pmatrix}$

Principal Component Analysis



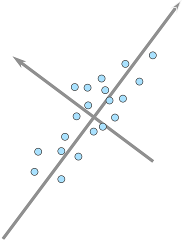
- Raw SNP data $\mathbf{X}$; x_{ij}
- Centering
  $\mathbf{Y} = \mathbf{CX}$; $y_{ij} = x_{ij} - \mu_j$
- Rotation $\mathbf{Y} = \mathbf{PL}$
- Truncation $\hat{\mathbf{P}} = \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \hline \\ \vdots \\ \mathbf{p}_n \end{pmatrix}$
- Approximation $\hat{\mathbf{Y}} = \hat{\mathbf{P}}\hat{\mathbf{L}}$

How to find PCs



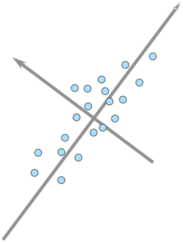
- Singular Value Decomposition:
  $\mathbf{Y} = (\mathbf{UD})\mathbf{L} = \mathbf{PL}$

How to find PCs



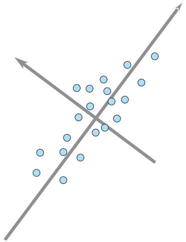
- Singular Value Decomposition: $\mathbf{Y} = (\mathbf{UD})\mathbf{L} = \mathbf{PL}$
- Eigendecomposition of $\mathbf{YY}^T$: $\mathbf{YY}^T = \mathbf{UD}^2\mathbf{U}^T = \mathbf{PP}^T$

How to find PCs



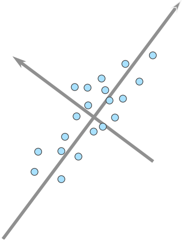
- Singular Value Decomposition: $\mathbf{Y} = (\mathbf{UD})\mathbf{L} = \mathbf{PL}$
- Eigendecomposition of $\mathbf{YY}^T$: $\mathbf{YY}^T = \mathbf{UD}^2\mathbf{U}^T = \mathbf{PP}^T$
- y_{ij}

How to find PCs



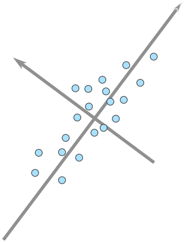
- Singular Value Decomposition: $\mathbf{Y} = (\mathbf{UD})\mathbf{L} = \mathbf{PL}$
- Eigendecomposition of $\mathbf{YY}^T$: $\mathbf{YY}^T = \mathbf{UD}^2\mathbf{U}^T = \mathbf{PP}^T$
- $y_{ij} = \sum_l (x_{il} - \mu_l)(x_{jl} - \mu_l)$

How to find PCs



- Singular Value Decomposition: $\mathbf{Y} = (\mathbf{UD})\mathbf{L} = \mathbf{PL}$
- Eigendecomposition of $\mathbf{YY}^T$: $\mathbf{YY}^T = \mathbf{UD}^2\mathbf{U}^T = \mathbf{PP}^T$
- $y_{ij} = \sum_l (x_{il} - \mu_l)(x_{jl} - \mu_l)$
- $y_{ij} = F_3(\boldsymbol{\mu}; \mathbf{X}_i, \mathbf{X}_j)$

How to find PCs

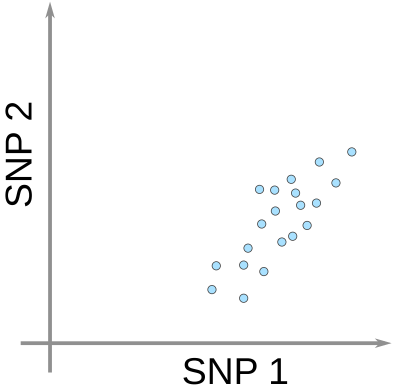


- Singular Value Decomposition:
  $\mathbf{Y} = (\mathbf{UD})\mathbf{L} = \mathbf{PL}$
- Eigendecomposition of $\mathbf{YY}^T$:
  $\mathbf{YY}^T = \mathbf{UD}^2\mathbf{U}^T = \mathbf{PP}^T$
- $y_{ij} = \sum_l (x_{il} - \mu_l)(x_{jl} - \mu_l)$
- $y_{ij} = F_3(\boldsymbol{\mu}; \mathbf{X}_i, \mathbf{X}_j)$

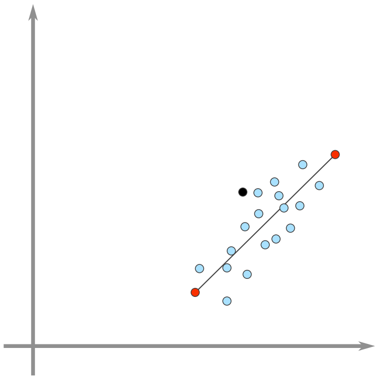
Observation

PCA is equivalent to outgroup-F_3-analysis with sample mean as outgroup

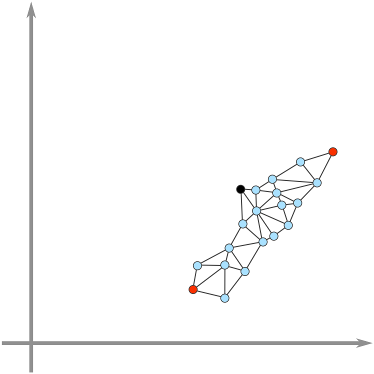
(metric) Multi-Dimensional Scaling (MDS)



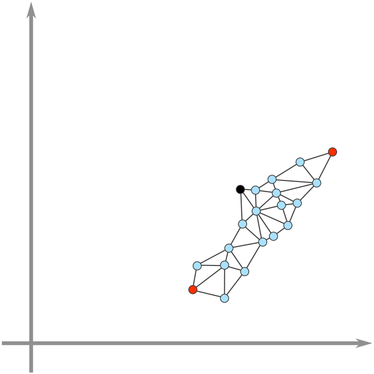
(metric) Multi-Dimensional Scaling (MDS)

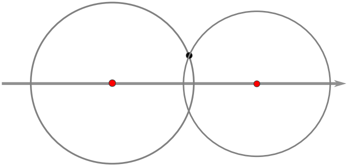


(metric) Multi-Dimensional Scaling (MDS)

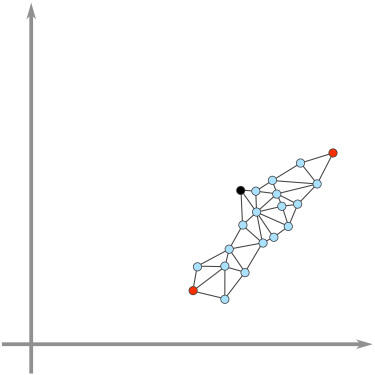


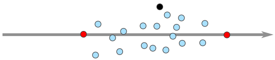
(metric) Multi-Dimensional Scaling (MDS)



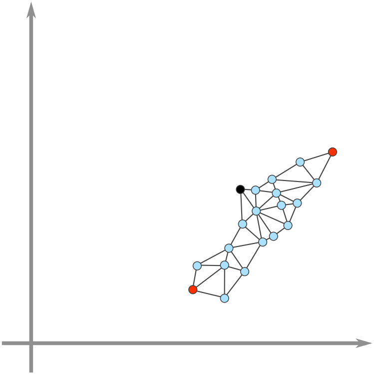


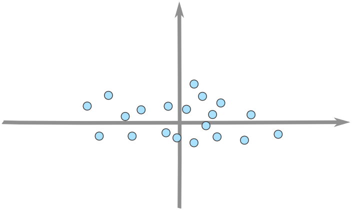
(metric) Multi-Dimensional Scaling (MDS)



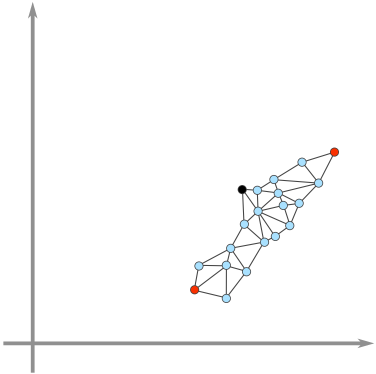


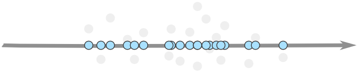
(metric) Multi-Dimensional Scaling (MDS)





(metric) Multi-Dimensional Scaling (MDS)





PCA is MDS on $\mathbf{F}_2$

- PCA is decomposition of Covariance matrix: $\mathbf{Y}\mathbf{Y}^T$

PCA is MDS on $\mathbf{F}_2$

- PCA is decomposition of Covariance matrix: $\mathbf{YY}^T$
- Consider $\mathbf{F}_2$; $f_{ij} = F_2(X_i, X_j) = X_i^2 + X_j^2 - 2X_iX_j$

PCA is MDS on $\mathbf{F}_2$

- PCA is decomposition of Covariance matrix: $\mathbf{Y}\mathbf{Y}^T$
- Consider $\mathbf{F}_2$; $f_{ij} = F_2(X_i, X_j) = X_i^2 + X_j^2 - 2X_iX_j$
- MDS is Eigendecomposition of $-\frac{1}{2}\mathbf{C}\mathbf{F}_2\mathbf{C}$

PCA is MDS on $\mathbf{F}_2$

- PCA is decomposition of Covariance matrix: $\mathbf{YY}^T$
- Consider $\mathbf{F}_2$; $f_{ij} = F_2(X_i, X_j) = X_i^2 + X_j^2 - 2X_iX_j$
- MDS is Eigendecomposition of $-\frac{1}{2}\mathbf{CF}_2\mathbf{C}$
- $\mathbf{CF}_2\mathbf{C} = \underbrace{\mathbf{CX}_i^2\mathbf{C}}_{0} + \underbrace{\mathbf{CX}_i^2\mathbf{C}}_{0} - 2\underbrace{\mathbf{CXX}^T\mathbf{C}}_{\mathbf{YY}^T}$

PCA is MDS on $\mathbf{F}_2$

- PCA is decomposition of Covariance matrix: $\mathbf{YY}^T$
- Consider $\mathbf{F}_2$; $f_{ij} = F_2(X_i, X_j) = X_i^2 + X_j^2 - 2X_iX_j$
- MDS is Eigendecomposition of $-\frac{1}{2}\mathbf{CF}_2\mathbf{C}$
- $\mathbf{CF}_2\mathbf{C} = \underbrace{\mathbf{CX}_i^2\mathbf{C}}_{0} + \underbrace{\mathbf{CX}_i^2\mathbf{C}}_{0} - 2\underbrace{\mathbf{CXX}^T\mathbf{C}}_{\mathbf{YY}^T}$

Observation

PCA is equivalent to MDS on $\mathbf{F}_2$

PCA is MDS on Outgroup$\mathbf{F}_3$

- PCA is decomposition of Covariance matrix: $\mathbf{YY}^T$

PCA is MDS on Outgroup$\mathbf{F}_3$

- PCA is decomposition of Covariance matrix: $\mathbf{YY}^T$
- Consider $\mathbf{F}_3(O)$; $f_{ij} = F_3(O; X_i, X_j) = O^2 - OX_i - OX_j + X_iX_j$

PCA is MDS on Outgroup$\mathbf{F}_3$

- PCA is decomposition of Covariance matrix: $\mathbf{YY}^T$
- Consider $\mathbf{F}_3(O)$; $f_{ij} = F_3(O; X_i, X_j) = O^2 - OX_i - OX_j + X_iX_j$

PCA is MDS on Outgroup$\mathbf{F}_3$

- PCA is decomposition of Covariance matrix: $\mathbf{Y}\mathbf{Y}^T$
- Consider $\mathbf{F}_3(O)$; $f_{ij} = F_3(O; X_i, X_j) = O^2 - OX_i - OX_j + X_iX_j$
- $\mathbf{C}\mathbf{F}_3\mathbf{C} = \underbrace{\mathbf{C}O^2\mathbf{C}}_{0} - \underbrace{\mathbf{C}O\mathbf{X}_i\mathbf{C}}_{0} - \underbrace{\mathbf{C}O\mathbf{X}_j\mathbf{C}}_{0} + \underbrace{\mathbf{C}\mathbf{X}\mathbf{X}^T\mathbf{C}}_{\mathbf{Y}\mathbf{Y}^T}$

PCA is MDS on Outgroup$\mathbf{F}_3$

- PCA is decomposition of Covariance matrix: $\mathbf{YY}^T$
- Consider $\mathbf{F}_3(O)$; $f_{ij} = F_3(O; X_i, X_j) = O^2 - OX_i - OX_j + X_iX_j$
- $\mathbf{CF}_3\mathbf{C} = \underbrace{\mathbf{C}O^2\mathbf{C}}_{0} - \underbrace{\mathbf{C}O\mathbf{X}_i\mathbf{C}}_{0} - \underbrace{\mathbf{C}O\mathbf{X}_j\mathbf{C}}_{0} + \underbrace{\mathbf{CXX}^T\mathbf{C}}_{\mathbf{YY}^T}$

Observation

Decomposition of *any* centered F_3-matrix is equivalent to PCA.

0-diagonal MDS