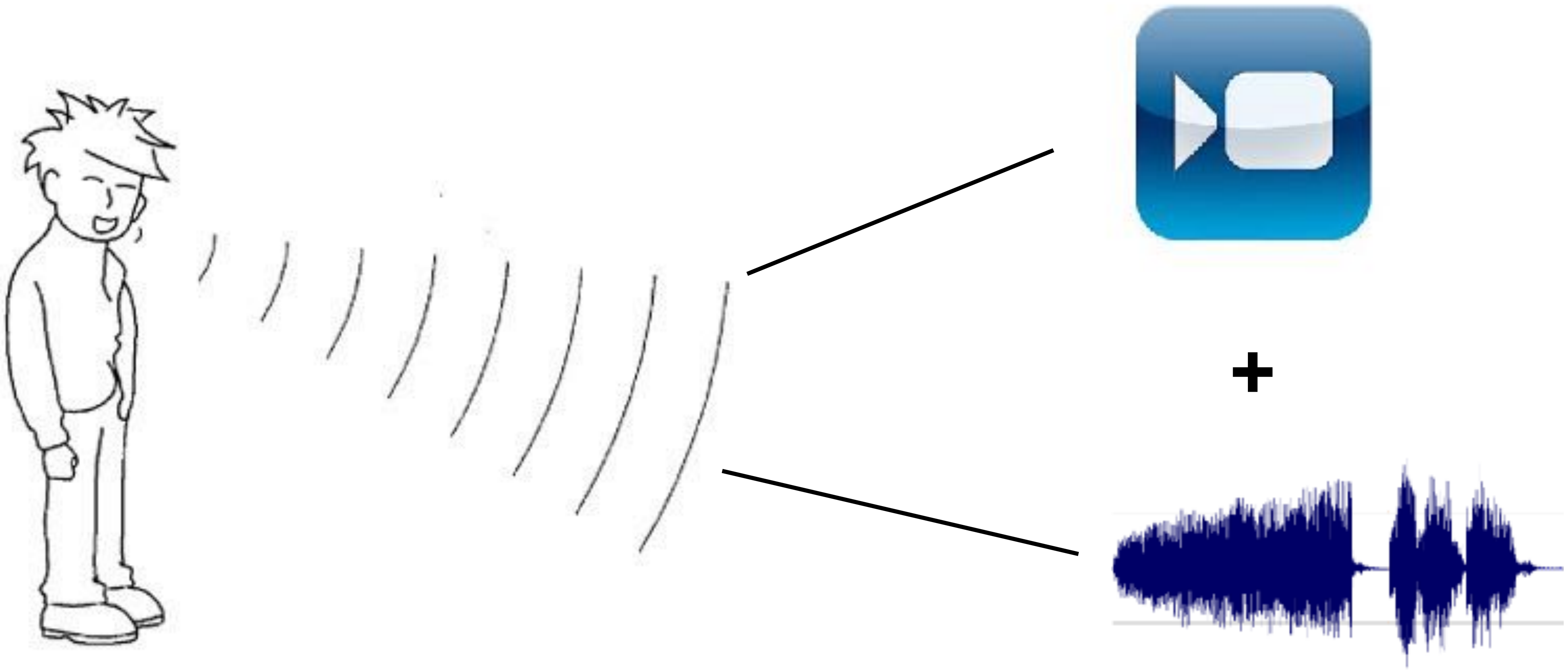
Machine Learning for Data Science (CS4786)
Lecture 6

Canonical Correlation Analysis

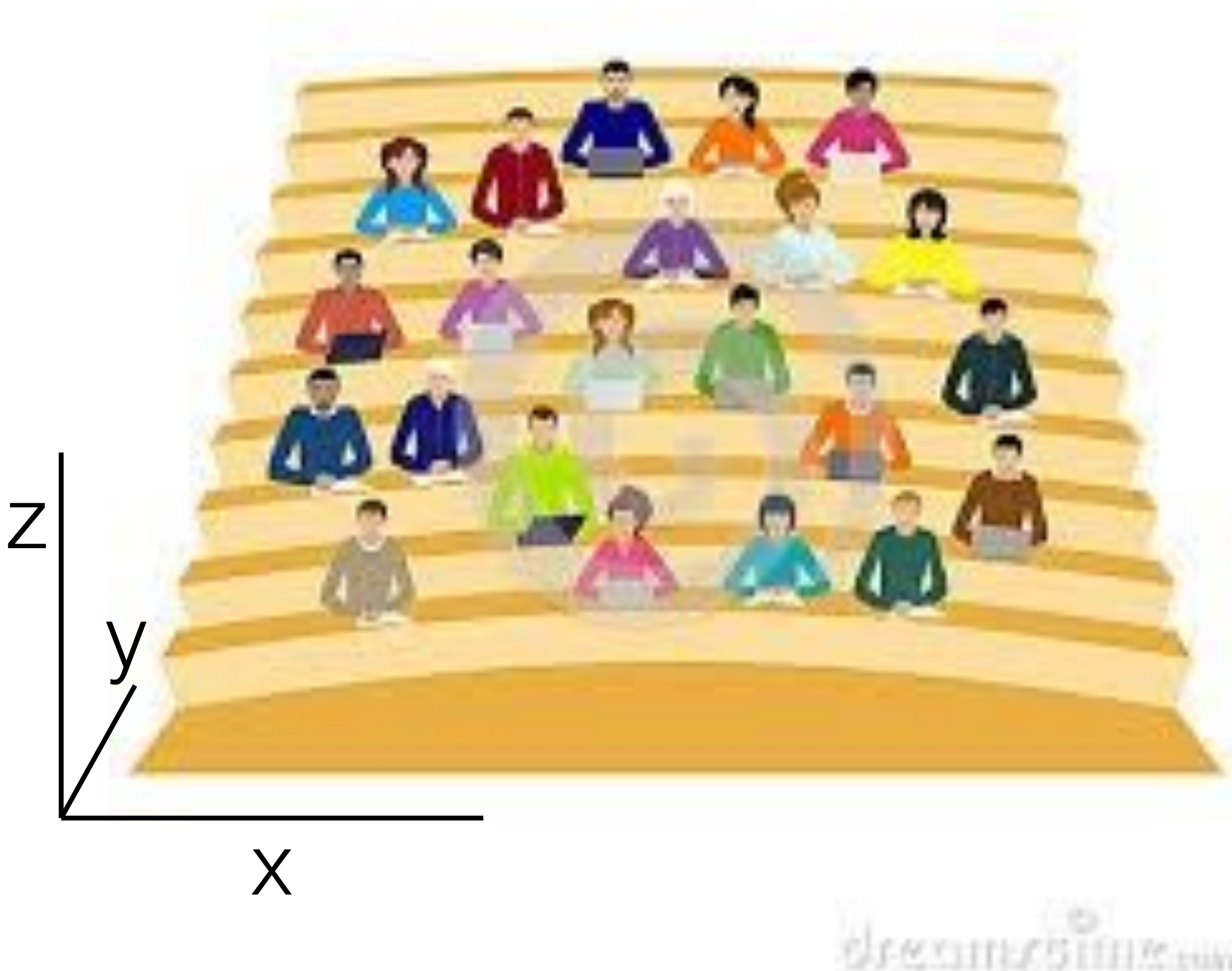
Example I: Speech Recognition



- Audio might have background sounds uncorrelated with video
- Video might have lighting changes uncorrelated with audio
- Redundant information between two views: the speech

Canonical Correlation Analysis

Canonical Correlation Analysis



Age

\+ Gender

Candies per week

Favorite Cartoon

Two View Dimensionality Reduction

- Data comes in pairs $(\mathbf{x}_1, \mathbf{x}'_1), \ldots, (\mathbf{x}_n, \mathbf{x}'_n)$ where $\mathbf{x}_t$'s are d dimensional and $\mathbf{x}'_t$'s are d' dimensional

Two View Dimensionality Reduction

- Data comes in pairs $(\mathbf{x}_1, \mathbf{x}'_1), \ldots, (\mathbf{x}_n, \mathbf{x}'_n)$ where $\mathbf{x}_t$'s are d dimensional and $\mathbf{x}'_t$'s are d' dimensional
- Goal: Compress say view one into $\mathbf{y}_1, \ldots, \mathbf{y}_n$, that are K dimensional vectors
  - Retain information redundant between the two views
  - Eliminate "noise" specific to only one of the views

Example II: Combining Feature Extractions

- Method A and Method B are both equally good feature extraction techniques

Example II: Combining Feature Extractions

- Method A and Method B are both equally good feature extraction techniques
- Concatenating the two features blindly yields large dimensional feature vector with redundancy

Example II: Combining Feature Extractions

- Method A and Method B are both equally good feature extraction techniques
- Concatenating the two features blindly yields large dimensional feature vector with redundancy
- Applying techniques like CCA extracts the key information between the two methods

Example II: Combining Feature Extractions

- Method A and Method B are both equally good feature extraction techniques
- Concatenating the two features blindly yields large dimensional feature vector with redundancy
- Applying techniques like CCA extracts the key information between the two methods
- Removes extra unwanted information

How do we get the right direction? (single dimension K =1)

How do we get the right direction? (single dimension K =1)

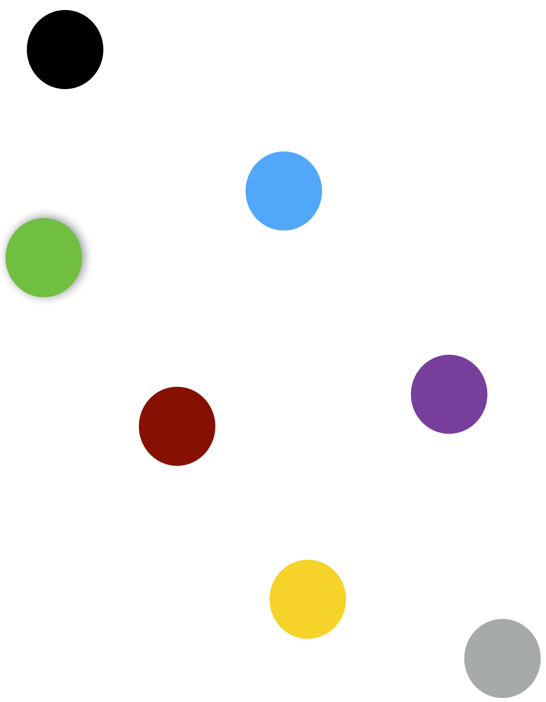
Age

\+ Gender

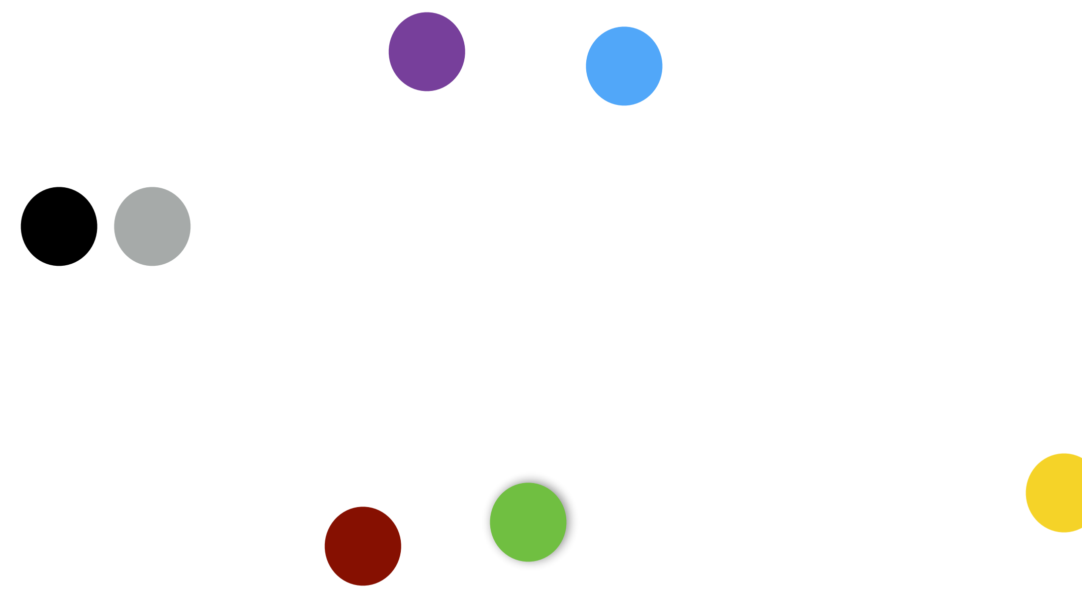
Candies per week

Favorite Cartoon

Which Direction to Pick?

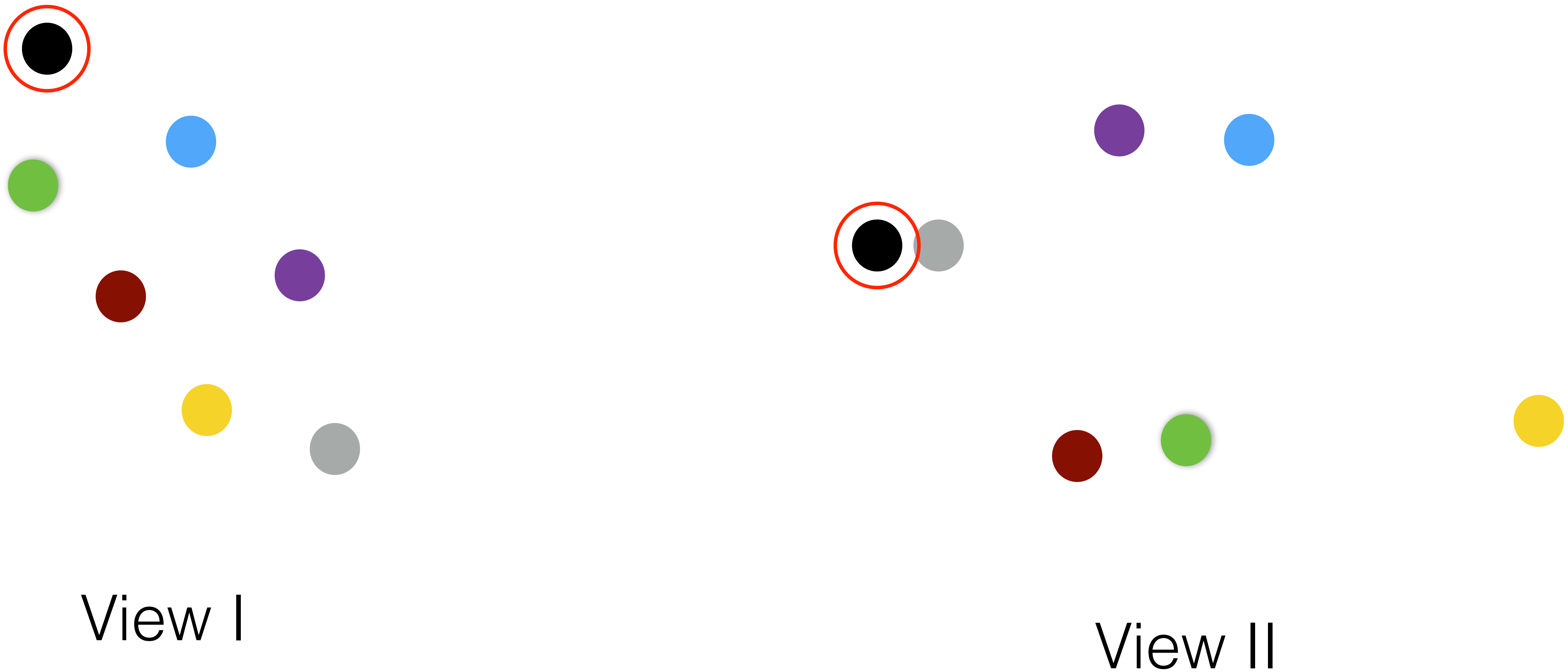


View I

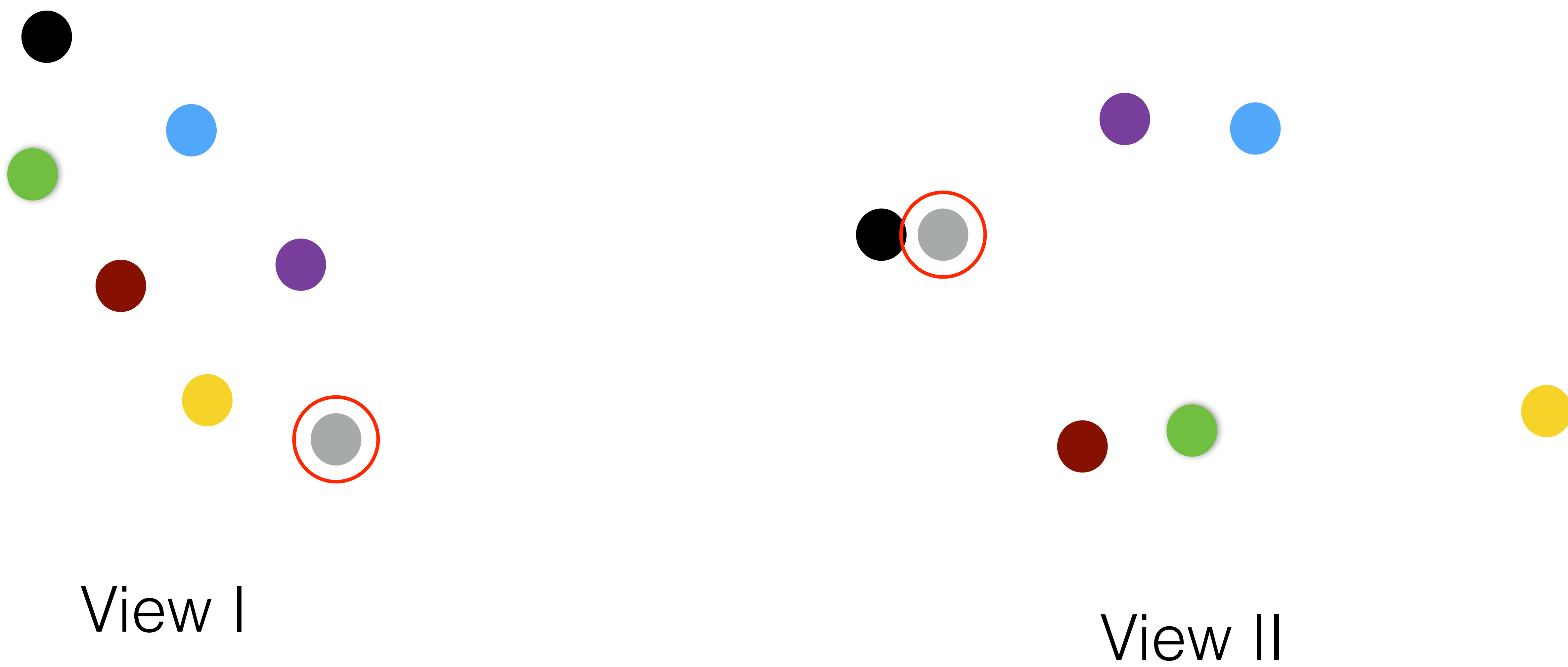


View II

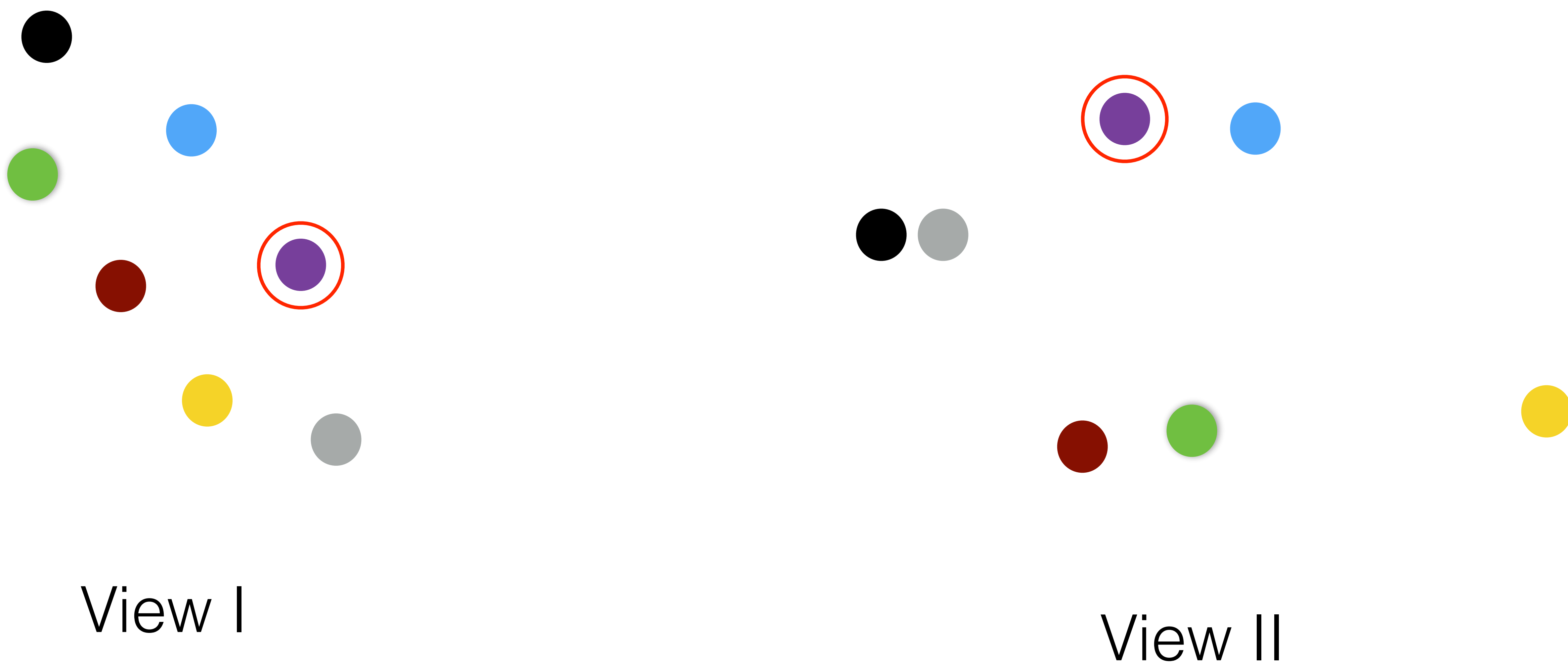
Which Direction to Pick?



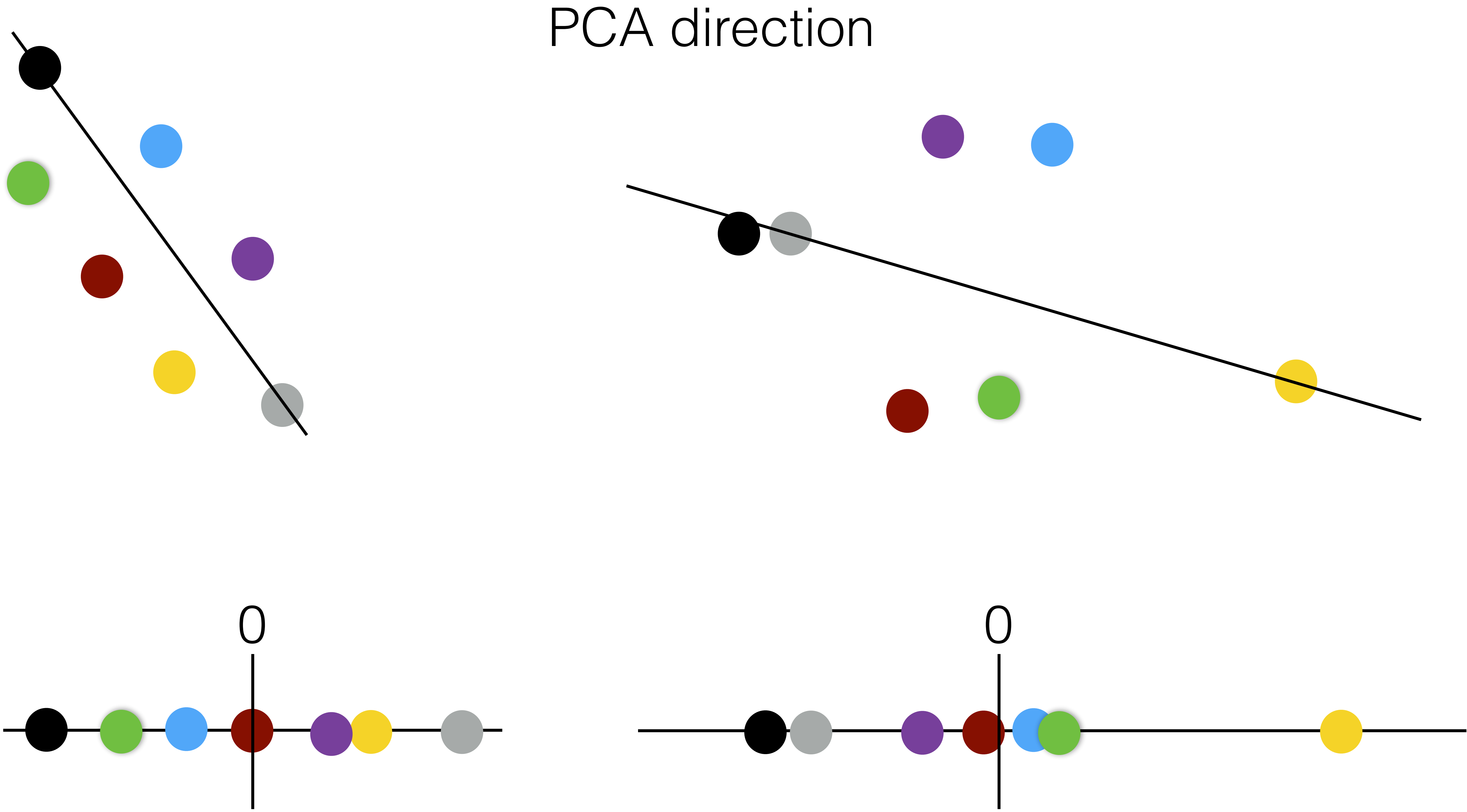
Which Direction to Pick?



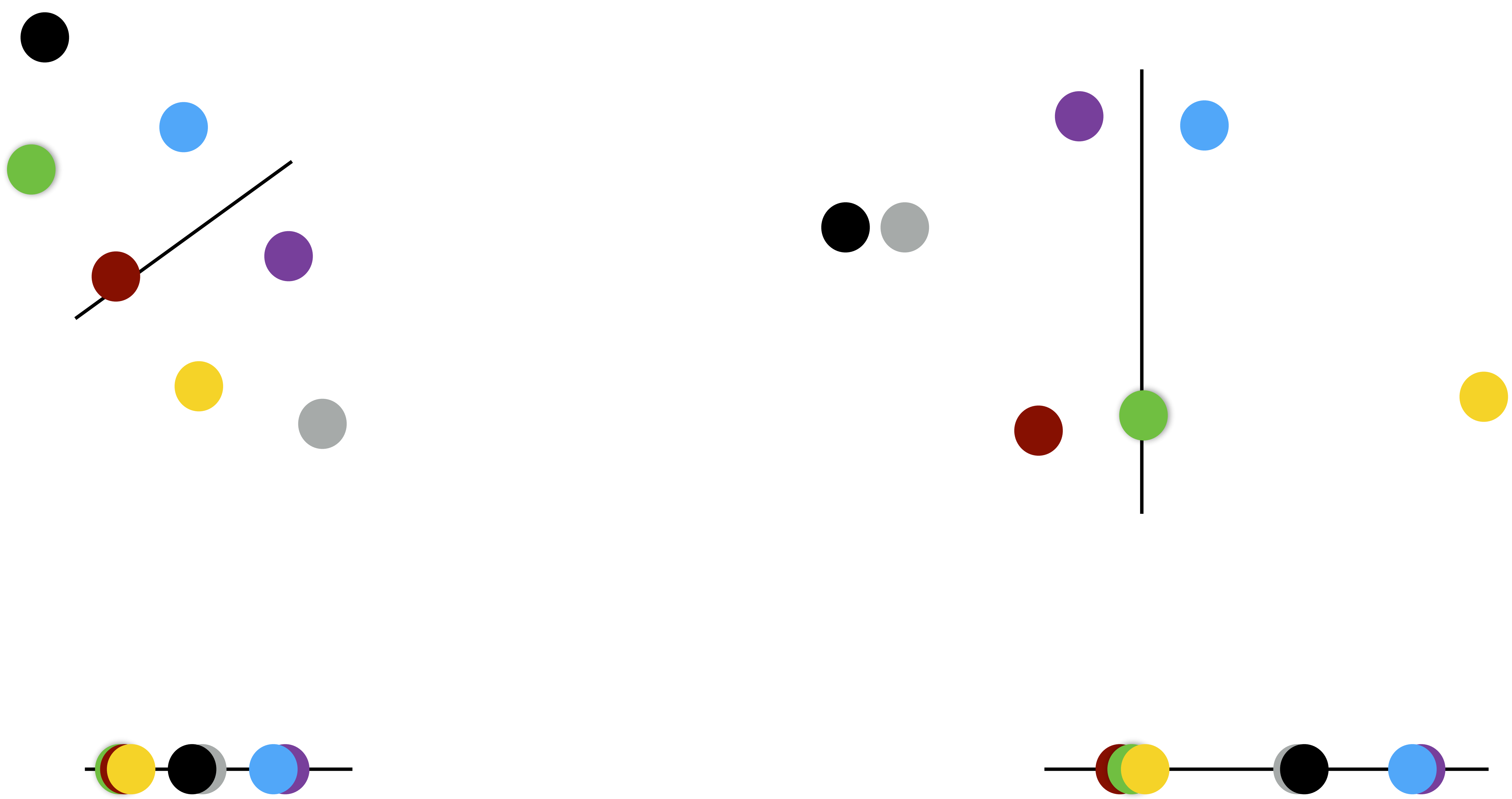
Which Direction to Pick?



Which Direction to Pick?



Which Direction to Pick?



Direction has large covariance

How do we pick the right direction to project to?

Pick the directions in each view so that resultant projections have high covariance.

Maximizing Correlation Coefficient

- Say $\mathbf{w}_1$ and $\mathbf{v}_1$ are the directions we choose to project in views 1 and 2 respectively we want these directions to maximize,

$$\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{y}_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}_t[1]\right)\cdot\left(\mathbf{y}'_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}'_t[1]\right)$$

where $\mathbf{y}_t[1]=\mathbf{w}_1^\top\mathbf{x}_t$ and $\mathbf{y}'_t[1]=\mathbf{v}_1^\top\mathbf{x}'_t$

This should work right?!?!

How do we get the right direction? (single dimension K =1)

How do we get the right direction? (single dimension K =1)

Age

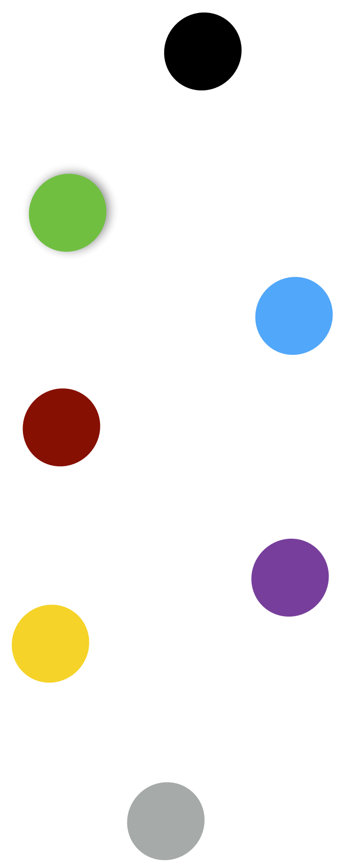
\+ Gender

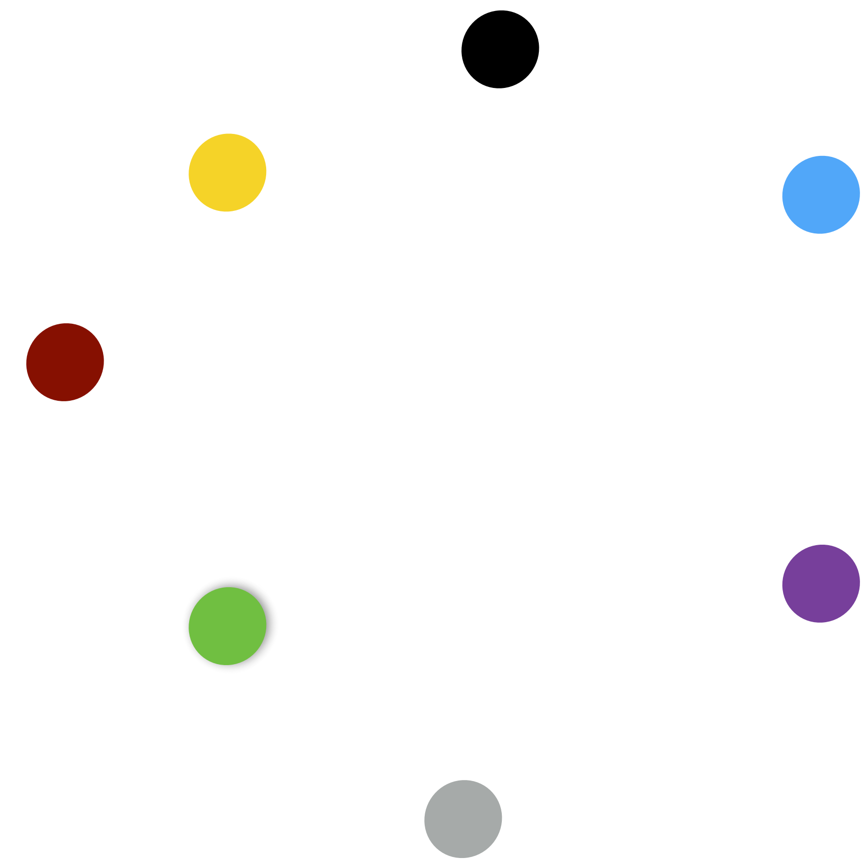
Candies per week

Favorite Cartoon

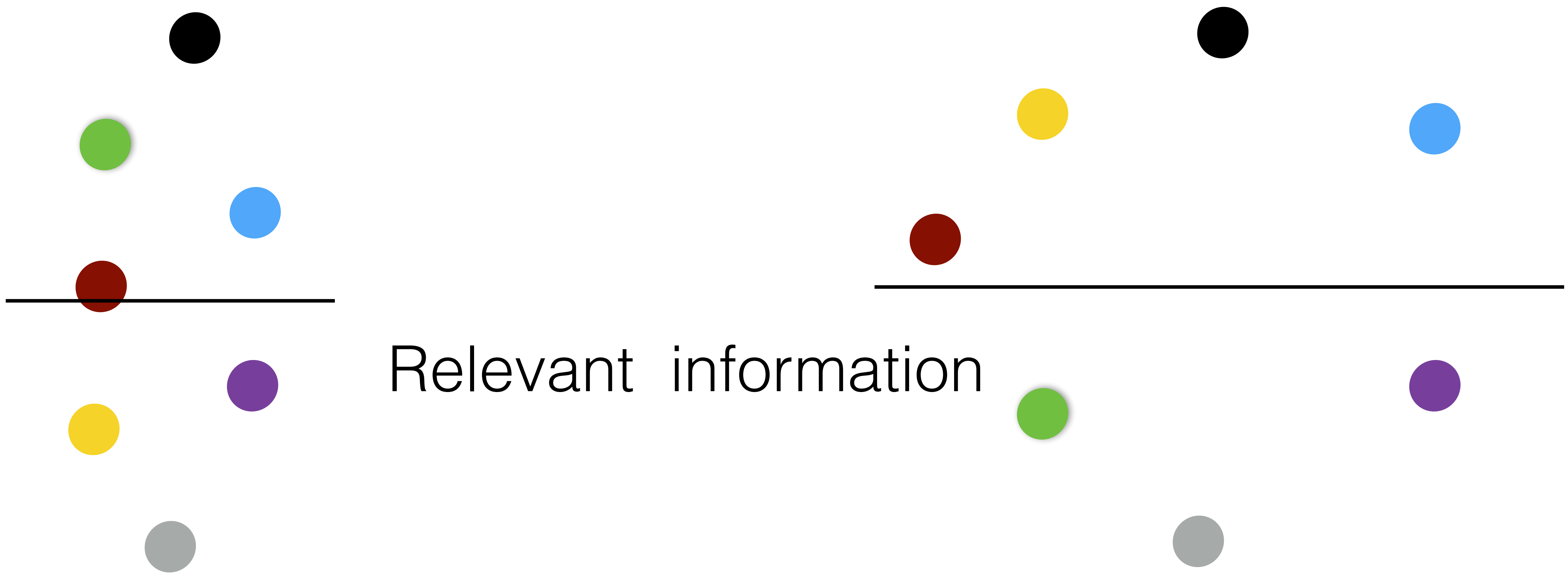
Why not Maximize Covariance

Why not Maximize Covariance

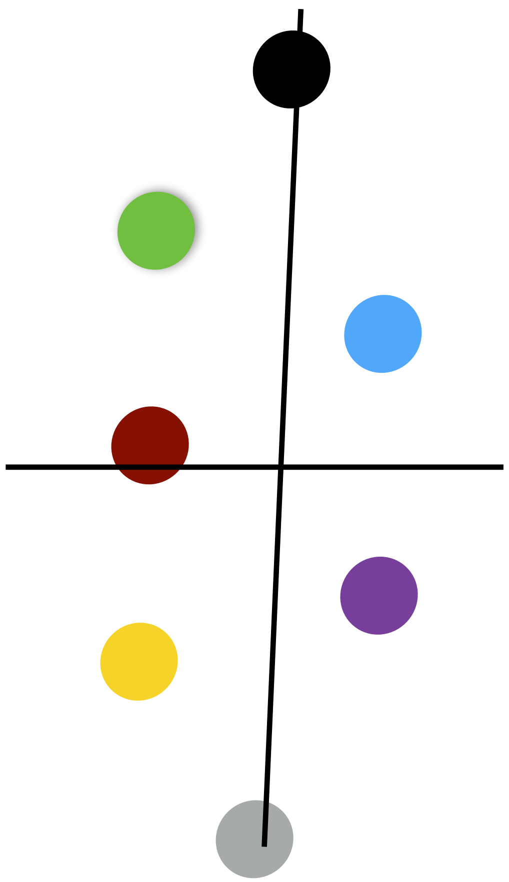




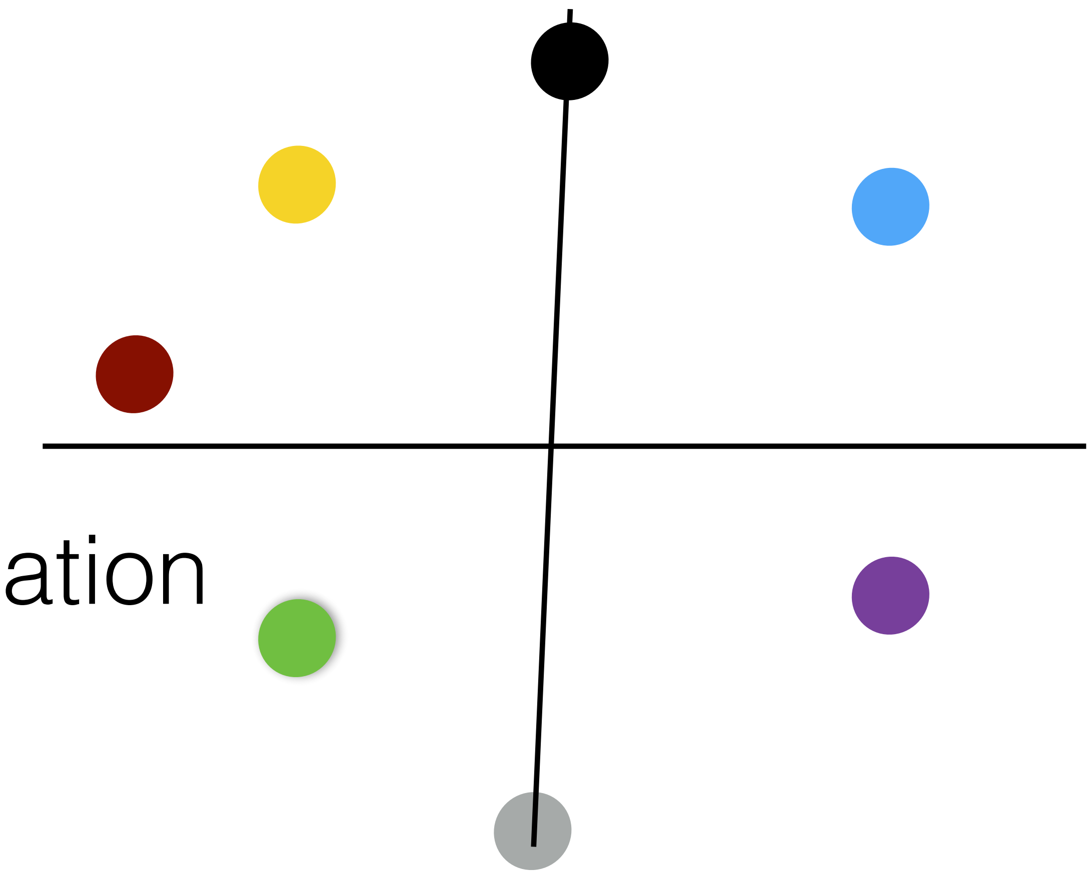
Why not Maximize Covariance



Why not Maximize Covariance



Relevant information



Say covariance in some coordinate just happens to be > 0

Scaling up this coordinate we can blow up covariance

Maximizing Correlation Coefficient

- Say $\mathbf{w}_1$ and $\mathbf{v}_1$ are the directions we choose to project in views 1 and 2 respectively we want these directions to maximize,

$$\frac{\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{y}_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}_t[1]\right)\cdot\left(\mathbf{y}'_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}'_t[1]\right)}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{y}_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}_t[1]\right)^2}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{y}'_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}'_t[1]\right)^2}}$$

Basic Idea of CCA

- Normalize variance in chosen direction to be constant (say 1)
- Then maximize covariance
- This is same as maximizing "correlation coefficient"

Covariance Vs Correlation

- $\text{Covariance}(A, B) = \mathbb{E}[(A - \mathbb{E}[A]) \cdot (B - \mathbb{E}[B])]$

  Depends on the scale of A and B. If B is rescaled, covariance shifts.

- $\text{Corelation}(A, B) = \frac{\mathbb{E}[(A-\mathbb{E}[A]) \cdot (B-\mathbb{E}[B])]}{\sqrt{\text{Var}(A)}\sqrt{\text{Var}(B)}}$

  Scale free.

Maximizing Correlation Coefficient

- Say $\mathbf{w}_1$ and $\mathbf{v}_1$ are the directions we choose to project in views 1 and 2 respectively we want these directions to maximize,

$$\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{y}_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}_t[1]\right)\cdot\left(\mathbf{y}'_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}'_t[1]\right)$$

where $\mathbf{y}_t[1]=\mathbf{w}_1^\top\mathbf{x}_t$ and $\mathbf{y}'_t[1]=\mathbf{v}_1^\top\mathbf{x}'_t$

Maximizing Correlation Coefficient

- Say $\mathbf{w}_1$ and $\mathbf{v}_1$ are the directions we choose to project in views 1 and 2 respectively we want these directions to maximize,

$$\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{y}_t[1] - \frac{1}{n}\sum_{t=1}^{n}\mathbf{y}_t[1]\right)\cdot\left(\mathbf{y}'_t[1] - \frac{1}{n}\sum_{t=1}^{n}\mathbf{y}'_t[1]\right)$$

s.t. $\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{y}_t[1] - \frac{1}{n}\sum_{t=1}^{n}\mathbf{y}_t[1]\right)^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{y}'_t[1] - \frac{1}{n}\sum_{t=1}^{n}\mathbf{y}'_t[1]\right) = 1$

where $\mathbf{y}_t[1] = \mathbf{w}_1^\top \mathbf{x}_t$ and $\mathbf{y}'_t[1] = \mathbf{v}_1^\top \mathbf{x}'_t$

Canonical Correlation Analysis

- Hence we want to solve for projection vectors $\mathbf{w}_1$ and $\mathbf{v}_1$ that

$$\text{maximize } \frac{1}{n}\sum_{t=1}^{n} \mathbf{w}_1^\top(\mathbf{x}_t - \mu) \cdot \mathbf{v}_1^\top(\mathbf{x}_t' - \mu')$$

$$\text{subject to } \frac{1}{n}\sum_{t=1}^{n} (\mathbf{w}_1^\top(\mathbf{x}_t - \mu))^2 = \frac{1}{n}\sum_{t=1}^{n} (\mathbf{v}_1^\top(\mathbf{x}_t' - \mu'))^2 = 1$$

where $\mu = \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t$ and $\mu' = \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t'$

Canonical Correlation Analysis

- Hence we want to solve for projection vectors $\mathbf{w}_1$ and $\mathbf{v}_1$ that

$$\begin{aligned}\text{maximize } & \mathbf{w}_1^\top \Sigma_{1,2} \mathbf{v}_1 \\ \text{subject to } & \mathbf{w}_1^\top \Sigma_{1,1} \mathbf{w}_1 = \mathbf{v}_1^\top \Sigma_{2,2} \mathbf{v}_1 = 1\end{aligned}$$

Canonical Correlation Analysis

- Hence we want to solve for projection vectors $\mathbf{w}_1$ and $\mathbf{v}_1$ that

$$\begin{aligned} &\text{maximize } \mathbf{w}_1^\top \Sigma_{1,2} \mathbf{v}_1 \\ &\text{subject to } \mathbf{w}_1^\top \Sigma_{1,1} \mathbf{w}_1 = \mathbf{v}_1^\top \Sigma_{2,2} \mathbf{v}_1 = 1 \end{aligned}$$

$$\Sigma = \begin{matrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{matrix} = \mathrm{cov}\left(X \; X' \right)$$

Solution

V

Solution

$$W = \operatorname{eigs}\left(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, K\right)$$

$$V = \operatorname{eigs}\left(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, K\right)$$

CCA Algorithm

CCA Algorithm

1. $$X = \left(n \underset{d_1}{X_1}, \underset{d_2}{X_2} \right)$$

CCA Algorithm

1. $$X = \left({}^{n} \underbrace{X_1}_{d_1} , \underbrace{X_2}_{d_2} \right)$$

2. $$\Sigma = \begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} = \mathrm{cov}\left(X \right)$$

CCA Algorithm

1. $$X = \left(n \underset{d_1}{X_1} , \underset{d_2}{X_2} \right)$$

2. $$\Sigma = \begin{matrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{matrix} = \mathrm{cov}\left(X \right)$$

3. $$W = \mathrm{eigs}\left(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, K \right)$$

CCA Algorithm

1. $X = \left({}^{n}\underset{d_1}{\boxed{X_1}}, \underset{d_2}{\boxed{X_2}} \right)$

2. $\boxed{\Sigma} = \boxed{\begin{matrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{matrix}} = \mathrm{cov}\left(\boxed{X} \right)$

3. $\boxed{W} = \mathrm{eigs}\left(\boxed{\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}, K \right)$

4. $\boxed{Y_1} = \boxed{X_1 - \mu_1} \times \boxed{W}$

CCA Demo