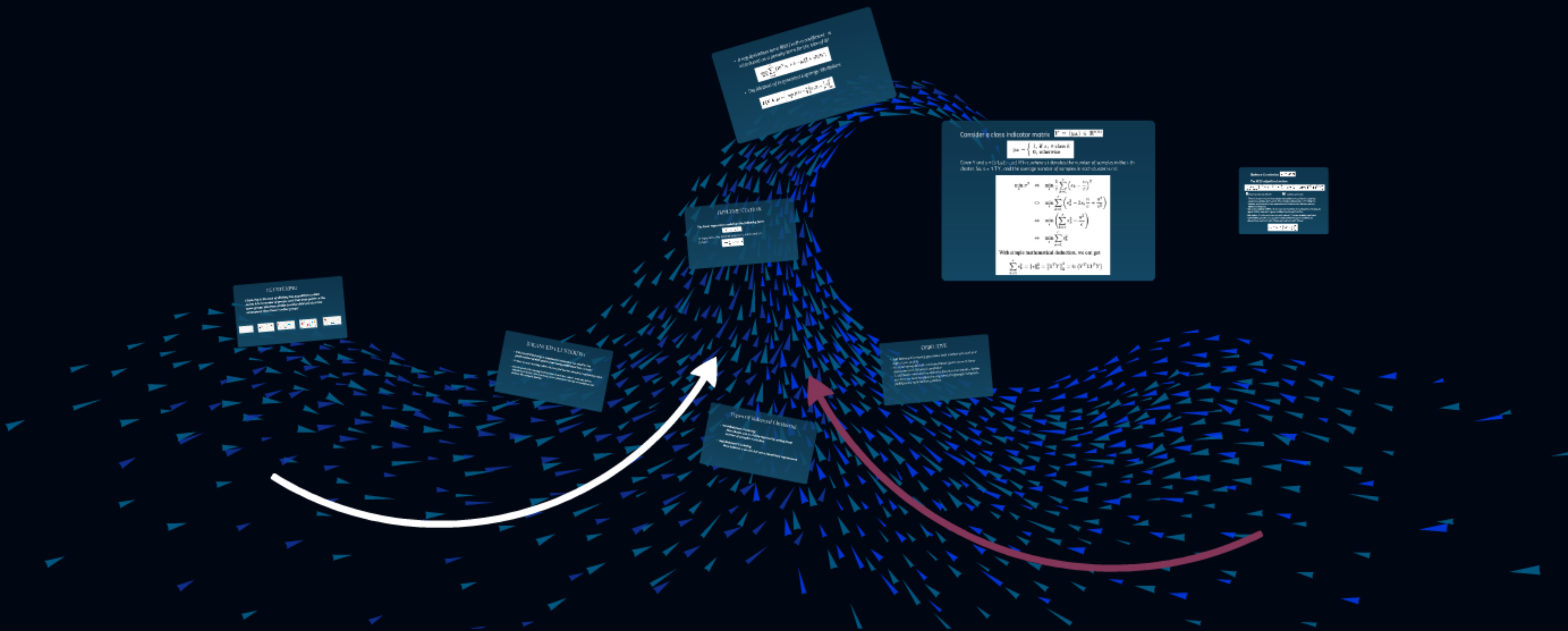
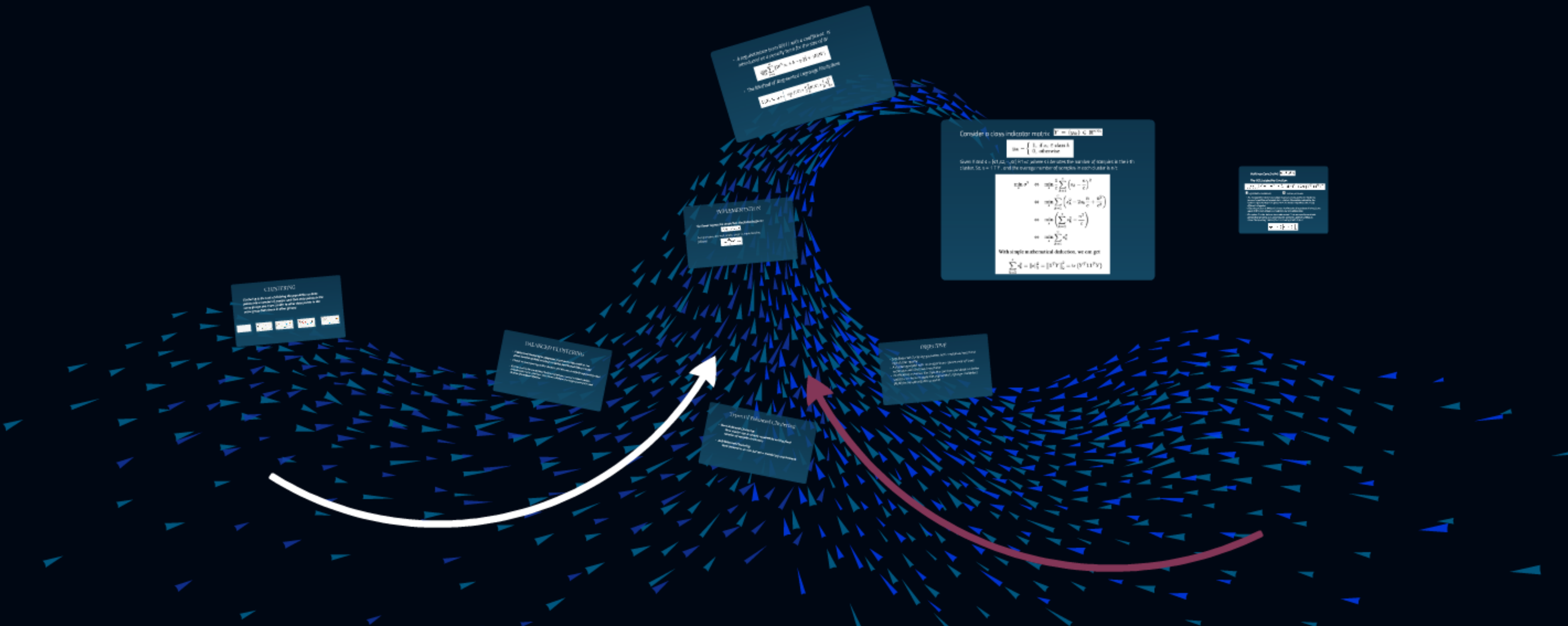


Balanced Clustering with Least Square Regression

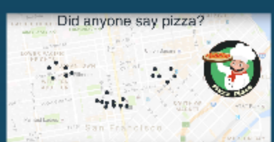


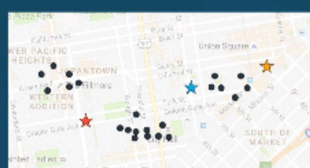
Balanced Clustering with Least Square Regression

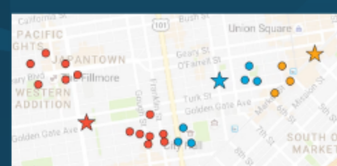
Prezi

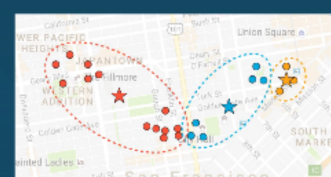
CLUSTERING

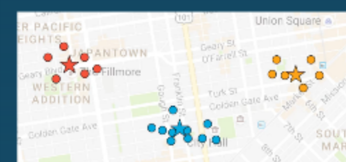
Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups





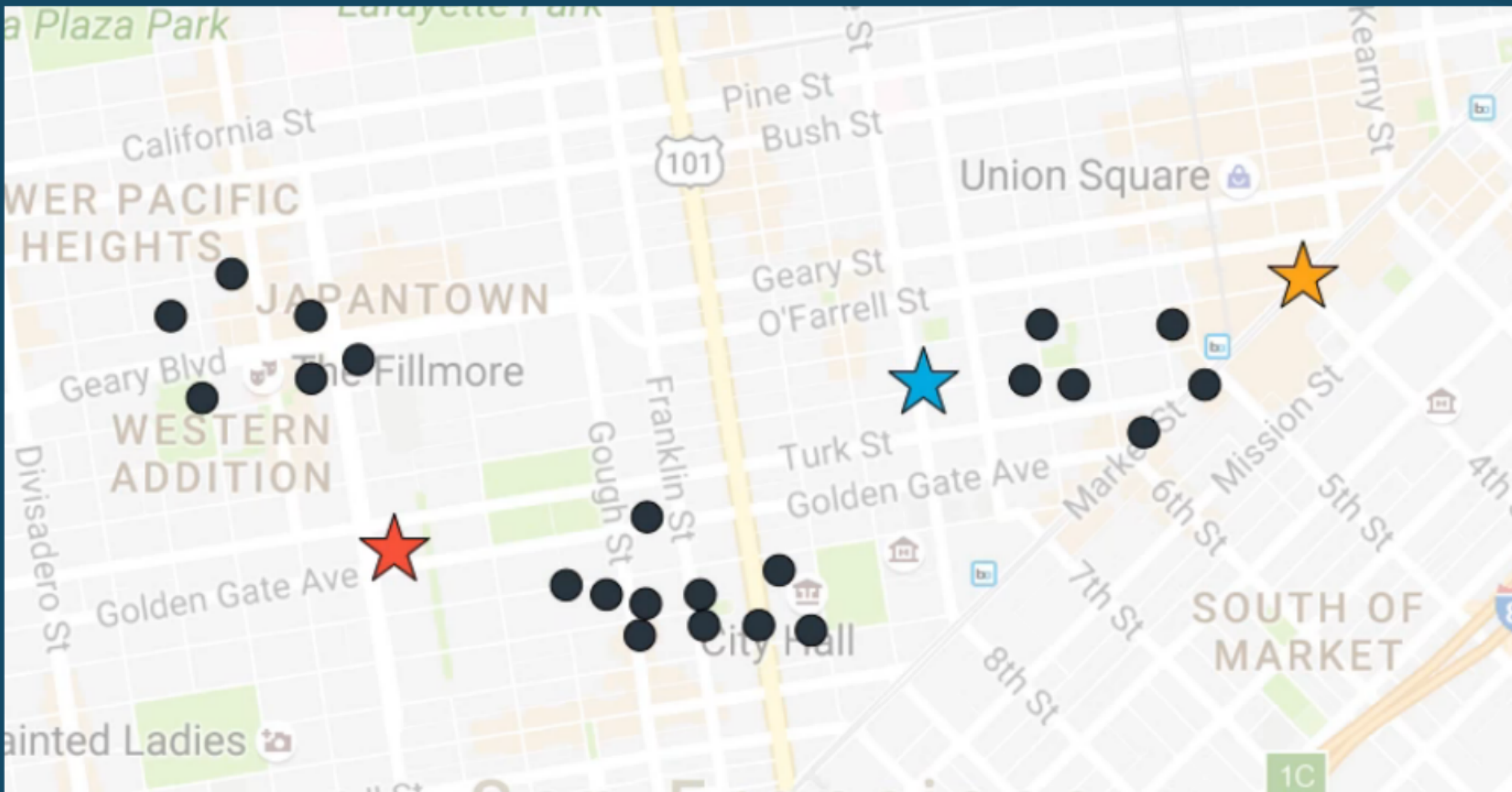




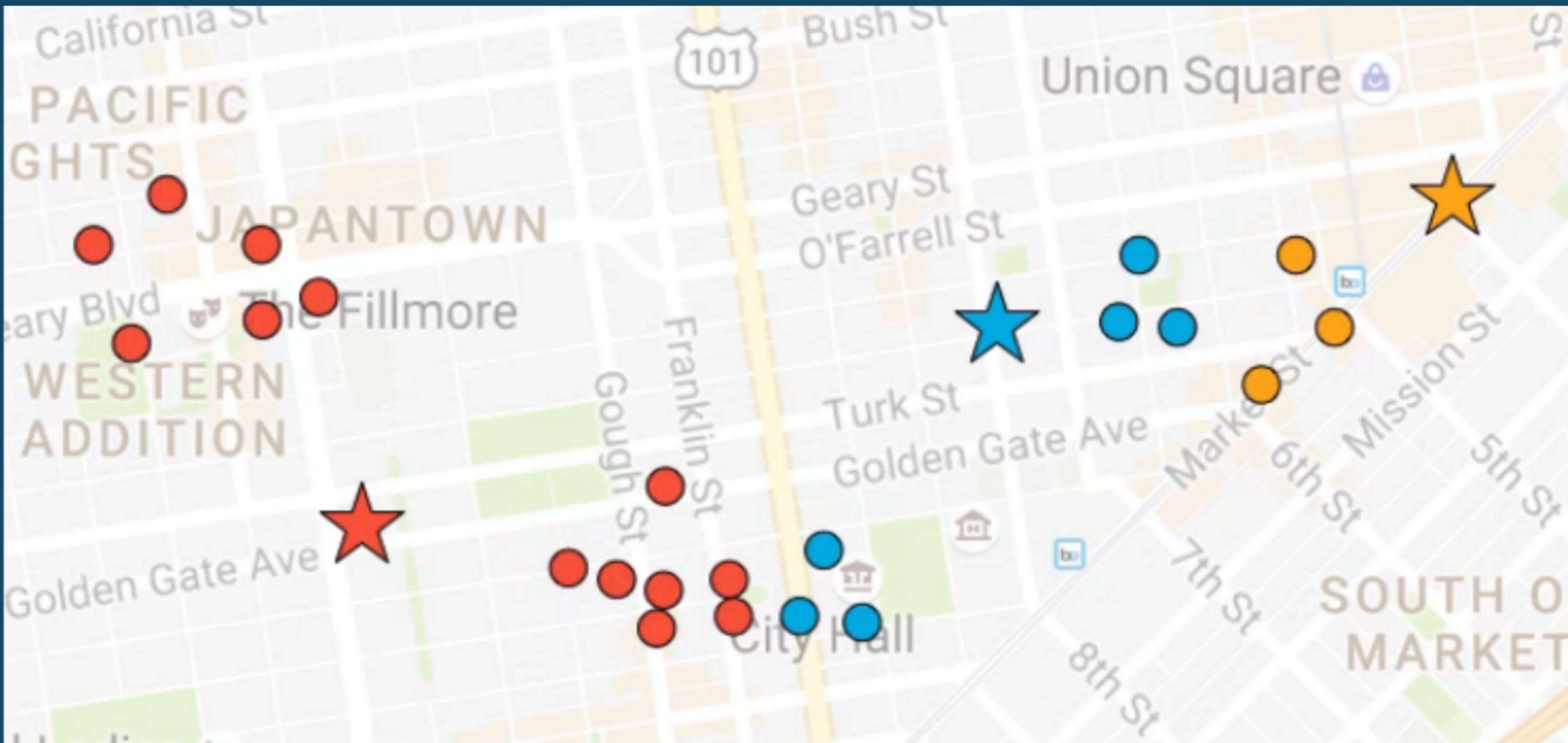


Prezi

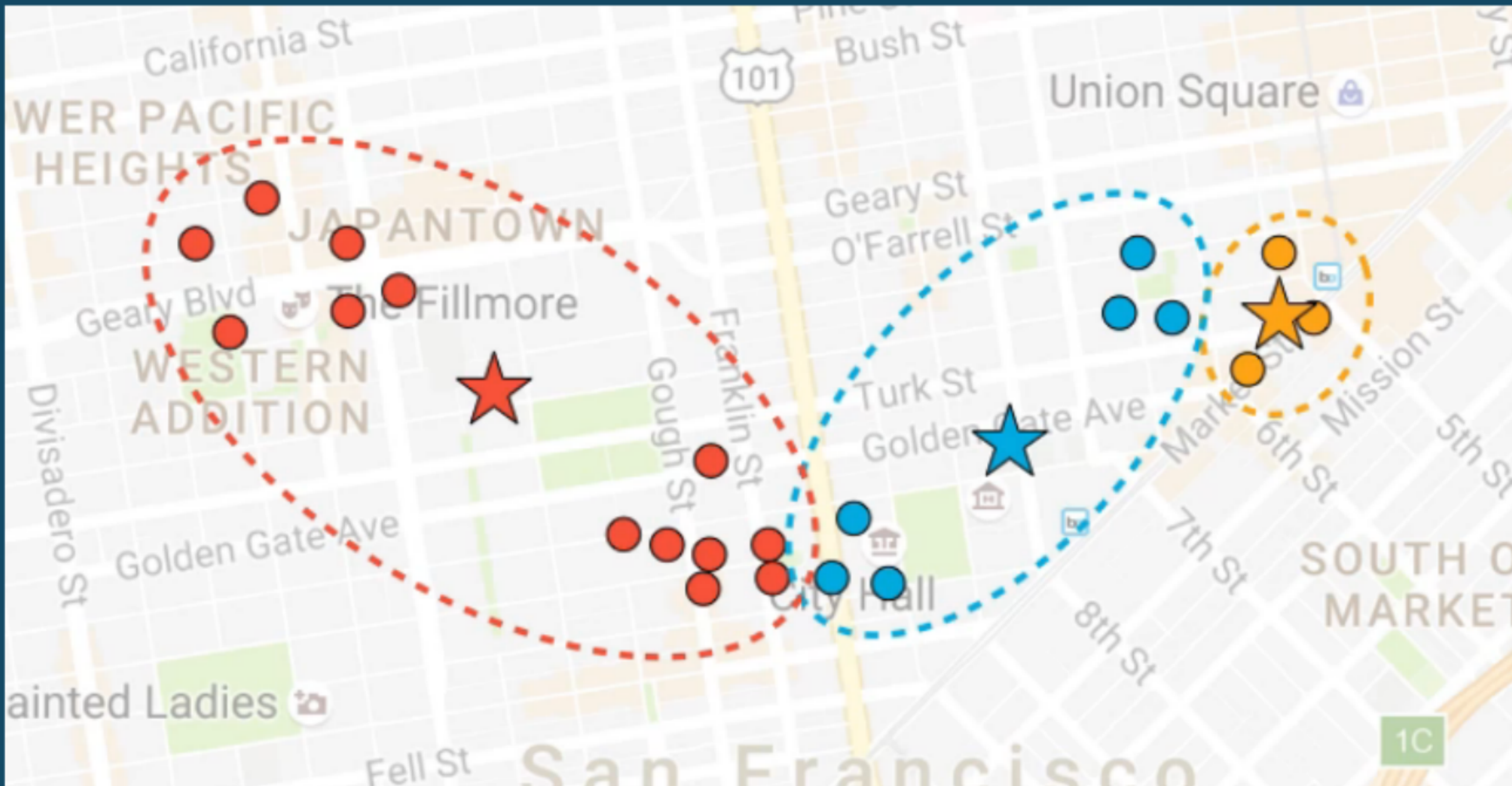
Did anyone say pizza?


Plaza Park
California St
WER PACIFIC
HEIGHTS
JAPANTOWN
Geary Blvd
The Fillmore
WESTERN
ADDITION
Divisadero St
Golden Gate Ave
ainted Ladies
Gough St
Franklin St
101
Pine St
Bush St
Geary St
O'Farrell St
Turk St
Golden Gate Ave
City Hall
Union Square
Kearny St
Market St
Mission St
6th St
5th St
4th St
7th St
8th St
SOUTH OF
MARKET
1C

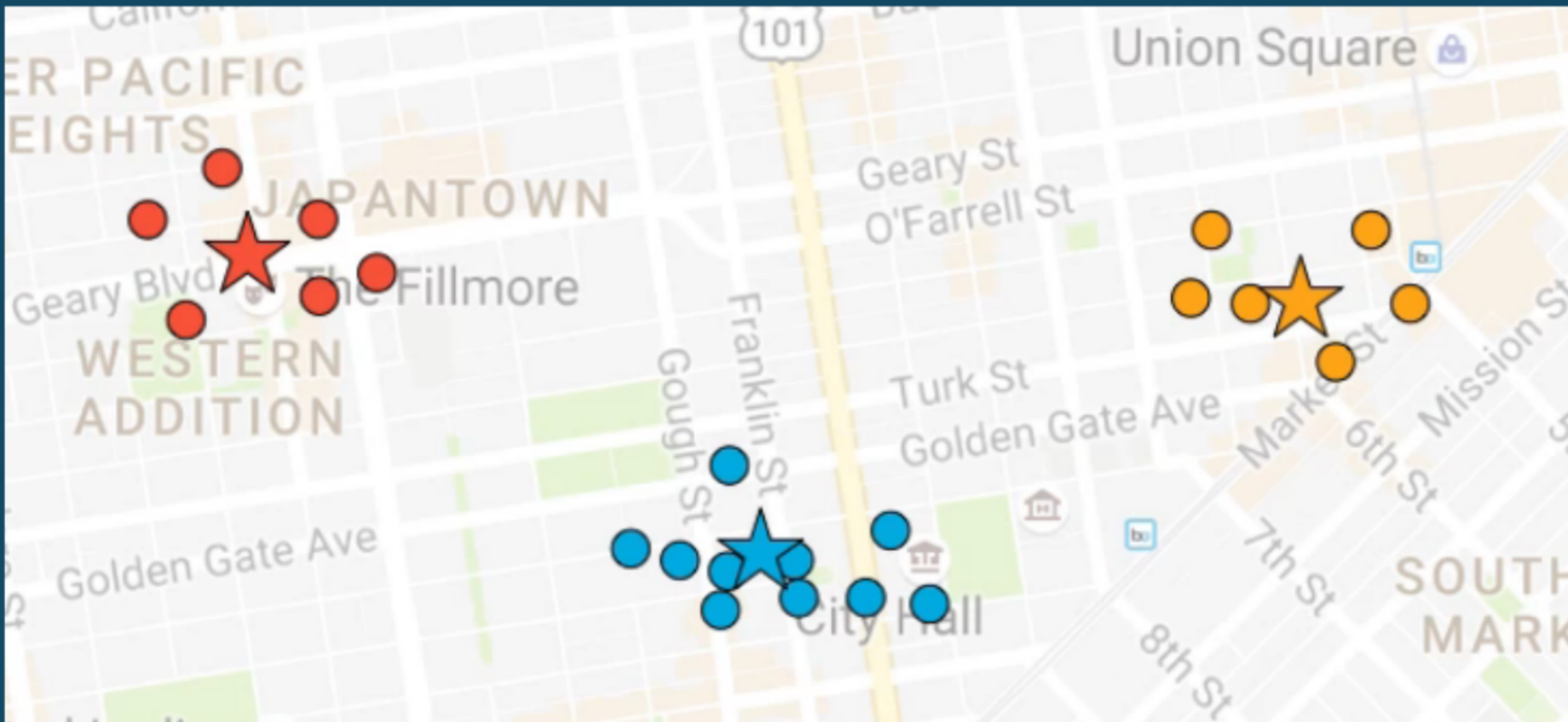
Prezi


California St
Bush St
101
Union Square
PACIFIC
GHTS
JAPANTOWN
Geary St
O'Farrell St
eary Blvd
The Fillmore
WESTERN
ADDITION
Franklin St
Gough St
Turk St
Golden Gate Ave
Market St
Mission St
6th St
5th St
7th St
8th St
Golden Gate Ave
City Hall
SOUTH OF
MARKET

Prezi


California St
Bush St
101
Union Square
WER PACIFIC
HEIGHTS
JAPANTOWN
Geary St
O'Farrell St
Geary Blvd
The Fillmore
WESTERN
ADDITION
Gough St
Franklin St
Turk St
Golden Gate Ave
Market St
6th St
Mission St
5th St
Divisadero St
Golden Gate Ave
City Hall
7th St
8th St
SOUTH O
MARKET
ainted Ladies
Fell St
San Francisco
1C

Prezi


Union Square
101
ER PACIFIC
EIGHTS
JAPANTOWN
Geary St
O'Farrell St
Geary Blvd
The Fillmore
WESTERN
ADDITION
Gough St
Franklin St
Turk St
Golden Gate Ave
Market St
Mission St
6th St
7th St
8th St
Golden Gate Ave
City Hall
SOUTH
MARK

Prezi

BALANCED CLUSTERING

- *A Balanced Clustering is supposed to prevent a too small or too great number of data points from being partitioned into a cluster*
- *It tends to avoid forming outlier clusters, and thus has beneficial regularizing effect*
- *Can be used in the energy load balance of wireless sensor networks where unbalanced cluster structure may cause unbalanced energy consumption and shorten the network lifetime*

Types Of Balanced Clustering

- ***Hard-Balanced Clustering:***
  Here cluster size is strictly required by setting fixed number of samples in clusters

- ***Soft-Balanced Clustering:***
  Here balance is an aim but not a mandatory requirement.

OBJECTIVE

- *Soft Balanced Clustering-guarantee both a balanced result and high cluster quality*
- *A clustering model with minimized least square error of linear regression with Balanced constraint*
- *To efficiently minimize the objective function and obtain a better solution, we have to apply the augmented Lagrange multipliers (ALM) for the optimization problem*

IMPLEMENTATION

The linear regression model has the following form:

$$f(x) = x^T w + b$$

In regression, the estimation error is minimized as follows:

$$\min \sum_{i=1}^{n} e(f(x_i), y_i)$$

- *A regularization term R(W) with a coefficient   is introduced as a penalty term for the size of W*

$$\min_{W,b} \sum_{i=1}^{n} \|W^T x_i + b - y_i\|_2^2 + \gamma R(W)$$

- *The Method of Augmented Lagrange Multipliers*

$$L(X, \Lambda, \mu) = \min_X f(X) + \frac{\mu}{2} \left\| H(X) + \frac{1}{\mu}\Lambda \right\|_F^2$$

Consider a class indicator matrix $Y = (y_{ik}) \in \mathbb{R}^{n\times c}$,

$$y_{ik} = \begin{cases} 1, & \text{if } x_i \in \text{class } k \\ 0, & \text{otherwise} \end{cases}$$

Given Y and s = [s1,s2,···,sc] R1×c ,where s i denotes the number of samples in the i-th cluster. So, s = 1 T Y , and the average number of samples in each cluster is n/c

$$\begin{aligned} \min_s \sigma^2 \quad &\Leftrightarrow \quad \min_s \frac{1}{c} \sum_{k=1}^{c} \left(s_k - \frac{n}{c}\right)^2 \\ &\Leftrightarrow \quad \min_s \sum_{k=1}^{c} \left(s_k^2 - 2s_k \frac{n}{c} + \frac{n^2}{c^2}\right) \\ &\Leftrightarrow \quad \min_s \left(\sum_{k=1}^{c} s_k^2 - \frac{n^2}{c}\right) \\ &\Leftrightarrow \quad \min_s \sum_{k=1}^{c} s_k^2 \end{aligned}$$

With simple mathematical deduction, we can get

$$\sum_{k=1}^{c} s_k^2 = \|s\|_2^2 = \left\|\mathbf{1}^T Y\right\|_2^2 = \operatorname{tr}\left(Y^T \mathbf{1}\mathbf{1}^T Y\right)$$

Balance Constraint: $\operatorname{tr}\left(Y^T \mathbf{1}\mathbf{1}^T Y\right)$

The BCLS objective function:

$$\min_{W,b,Y\in Ind} \left\|X^T W + \mathbf{1}b^T - Y\right\|_F^2 + \gamma \|W\|_F^2 + \lambda \operatorname{tr}\left(Y^T \mathbf{1}\mathbf{1}^T Y\right)$$

γ-regularization parameter

λ -balance parameter

- ***The minimization of the least square regression term guides the clustering process to partition data points into c clusters. Meanwhile, minimizing the balance regularization term guarantees the balanced partitioning among different categories.***
- ***Since the problem is NP-hard, it is very hard to solve it in polynomial time.So, we apply ALM to help obtain good solutions for the optimization***
- ***We replace Y in the balance term with matrix Z into an equality constraint optimization problem and approximately obtain the optimal solution by alternatively solving Y with Z fixed and solving Z with Y fixed.***

$$\min_X f(X) + \frac{\mu}{2}\left\|H(X) + \frac{1}{\mu}\Lambda\right\|_F^2$$

We get the final optimization problem for our BCLS method

$$\min_{W,b,Y\in Ind,Z} \left\|X^T W + \mathbf{1}b^T - Y\right\|_F^2 + \gamma\|W\|_F^2 \; + \lambda \operatorname{tr}\left(Z^T \mathbf{1}\mathbf{1}^T Z\right) + \frac{\mu}{2}\left\|Y - Z + \frac{1}{\mu}\Lambda\right\|_F^2$$

μ - *positive scalar and its value increases slightly after each iteration*

Λ -- *estimate of the Lagrange multiplier*

- ***The estimation accuracy improves at every step of iteration.***
- ***It can be solved by alternatively updating the four variables W , b, Y and Z***

1) With Y and Z fixed

$$\min_{W,b} \left\|X^T W + \mathbf{1}b^T - Y\right\|_F^2 + \gamma\|W\|_F^2$$

2) simply by setting the derivatives of the objective function with respect to W and b to zeros, we have

$$\begin{cases} W = \left(XX^T + \gamma I_d\right)^{-1} XY \\ b = \dfrac{1}{n} Y^T \mathbf{1} \end{cases}$$

3) With W , b and Y

$$\min_{Z} \lambda \operatorname{tr}\left(Z^T \mathbf{1}\mathbf{1}^T Z\right) + \frac{\mu}{2}\left\|Y - Z + \frac{1}{\mu}\Lambda\right\|_F^2$$

4) Z can be obtained by setting the derivative of the objective function with respect to Z to zero

$$Z = \left(\mu I_n + 2\lambda \mathbf{1}\mathbf{1}^T\right)^{-1}(\mu Y + \Lambda)$$

5) With W, b and Z fixed, we can transform BCLS eq into:

$$\min_{Y \in Ind} \|Y - V\|_F^2 + const.$$

where const. denotes a constant and V = (v ik) R n×c , and

$$V = \frac{2}{2+\mu}\left(X^T W + \mathbf{1}b^T\right) + \frac{1}{2+\mu}(\mu Z - \Lambda)$$

6) Each element of Y = (y ik) R n×c is binary and the sum of each row is 1

We can obtain the solution as follows:

$$y_{ik} = \begin{cases} 1, & \text{if } k = \arg\max_k \{v_{ik}\}_{k=1}^{c} \\ 0, & \text{otherwise} \end{cases}$$

Prezi

Algorithm of the BCLS method

Input: Centered dataset $X = [x_1, x_2, \cdots, x_n] \in \mathbb{R}^{d\times n}$, the number of clusters c, and parameters γ, λ, and $\rho(\rho > 0)$.
Output: Assignment matrix $Y \in \mathbb{R}^{n\times c}$.
Initialize Y randomly, initialize μ, and set $\Lambda = \mathbf{0} \in \mathbb{R}^{n\times c}$.
Repeat

1. Obtain W and b by solving the problem in Eq. (14) with the solution in Eq. (15);
2. Obtain Z by solving Eq. (16) with the solution in Eq. (17);
3. Obtain Y by Eq. (19) and Eq. (21);
4. **Update** Λ**:** $\Lambda^{(t+1)} = \Lambda^{(t)} + \mu^{(t)}(Y - Z)$;
5. **Update** μ**:** $\mu^{(t+1)} = \rho\mu^{(t)}$.

Until $Y - Z \to \mathbf{0}$ or Λ remains unchanged
Return Y.

$$\min_{W,b} \|X^T W + \mathbf{1}b^T - Y\|_F^2 + \gamma \|W\|_F^2 \quad (14)$$

$$\begin{cases} W = (XX^T + \gamma I_d)^{-1} XY \\ b = \frac{1}{n} Y^T \mathbf{1} \end{cases} \quad (15)$$

$$\min_{Z} \lambda \mathrm{tr}\left(Z^T \mathbf{1}\mathbf{1}^T Z\right) + \frac{\mu}{2} \left\| Y - Z + \frac{1}{\mu}\Lambda \right\|_F^2 \quad (16)$$

$$Z = \left(\mu I_n + 2\lambda \mathbf{1}\mathbf{1}^T\right)^{-1} (\mu Y + \Lambda) \quad (17)$$

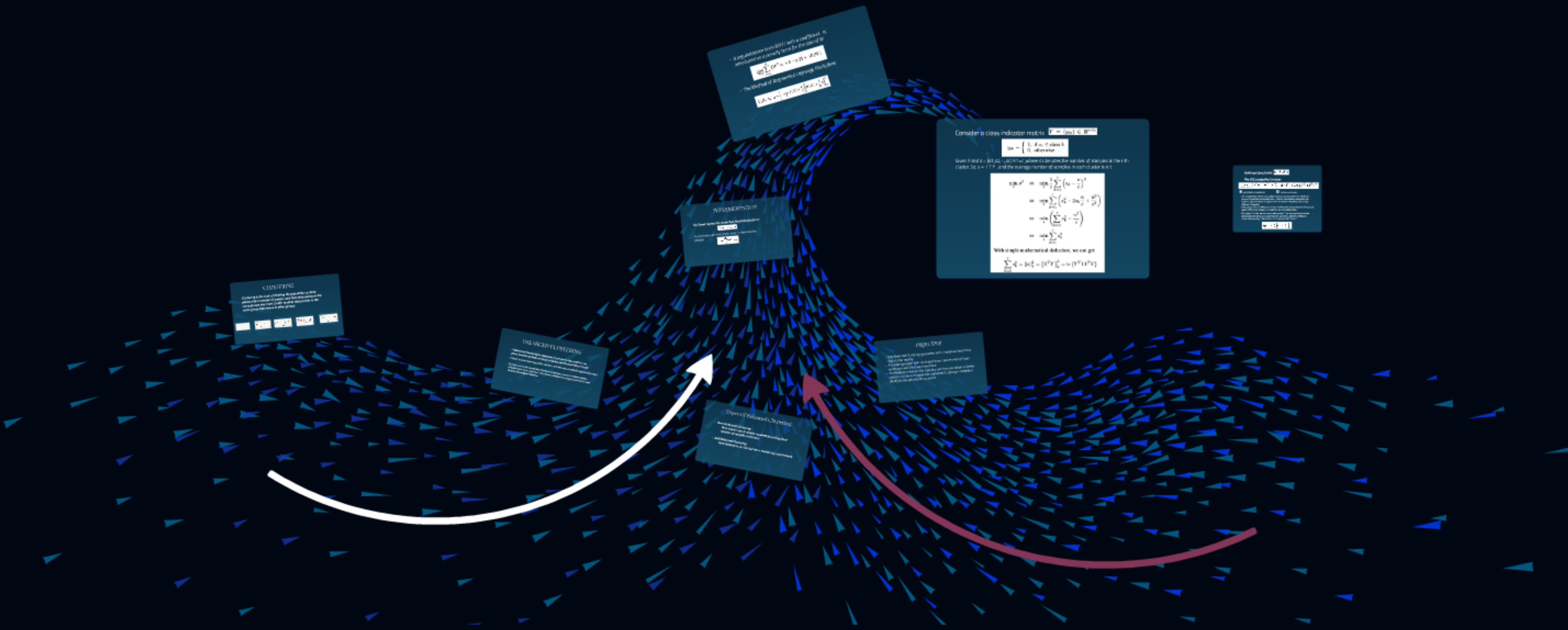
$$V = \frac{2}{2+\mu}(X^T W + \mathbf{1}b^T) + \frac{1}{2+\mu}(\mu Z - \Lambda) \quad (19)$$

$$y_{ik} = \begin{cases} 1, & \text{if } k = \arg\max_k \{v_{ik}\}_{k=1}^c \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

THANK YOU!!!

https://github.com/juhi-09/SMAI-BCLS.git

Team:Breaking Bad
Juhi Gupta(20172031)
Nikita Agrawal(20172093)
Tanu Rathi(20172125)
Tejus Agarwal(20172066)



Balanced Clustering with Least Square Regression