Rutgers Business School
Newark and New Brunswick

Buyer Targeting Optimization

A Unified Customer Segmentation Perspective

Yuan Qu

Background

- Marketing: STP Model
  - **Segmentation**: divide customers into groups of people with common characteristics and needs.
  - **Targeting**: identifies promising customers and allocates marketing resources to increase profits.
  - Positioning
- Usually, accomplished by independent steps.

Background

Rutgers Business School
Newark and New Brunswick

- For Data Mining…
    - Segmentation → **Clustering**
    - Targeting → **Classification**
- Is it possible to optimize the buyer targeting performance by constructing these tasks in a more integrated way?
- Key idea: use clustering solution to boost the **classification** performance.

Data Sets

Rutgers Business School
Newark and New Brunswick

Data	Size	Pos_class	Neg_class
Synthetic data	500	200	300
Real — Product A	8,315	1,680	6,635
Real — Product B	22,160	4,232	17,828

- Synthetic data
    - (x, y, z): $(x, y) \in R^2, z \in \{-1, +1\}$
- Real business data
    - 49 integer attributes

Data Sets

Rutgers Business School
Newark and New Brunswick

Demographics	Values
Company Size	Small Business, Enterprise, Unknown
Industry	Heavy Hitters, Potentials
Job Title	IT Staff, IT Manager, Executive, Researcher, Non-IT, Unknown

Behaviors	Values
Event	Corporate Event, Trade Show, Conference, Webinar, Seminar, Technology Preview
Offer	Official Website, Direct Mail, Email, Call Center, Search Engine, Web Advertising (third party), Social Media
Product	Product Download, Product Free Trial, Product Renewal, Product Activation, Product Training
Activity	Subscribe, Unsubscribe, Active, Inactive

RUTGERS
Rutgers Business School
Newark and New Brunswick

Problem Definition

- Inputs: dataset X with N rows and D discrete attributes, and label attribute y. $(X, y) \subset R^{N \times D} \times \{-1, +1\}$

- Group the customers into K clusters: $\{S_1, S_2 \ldots S_K\}$

- learn the classifier C_k for each segment S_k

- Minimize the total loss:

$$J(S, C) = \frac{1}{N} \sum_{i=1}^{N} loss(x_i, y_i \mid C_{l_i}) = \frac{1}{N} \sum_{k=1}^{K} \sum_{x_i \in S_K} loss(x_i, y_i \mid C_k)$$

Problem Definition

- Minimize the total loss:

$$J(S, C) = \frac{1}{N}\sum_{i=1}^{N} loss(x_i, y_i \mid C_{l_i}) = \frac{1}{N}\sum_{k=1}^{K}\sum_{x_i \in S_K} loss(x_i, y_i \mid C_k)$$

- Loss functions:
  - SVM loss: $loss(x, y \mid C) = \max\{0, 1 - y \cdot C(x)\}$
  - Logit regression: $loss(x, y \mid C) = \log(1 + \exp(-y \cdot C(x)))$

Algorithm

Rutgers Business School
Newark and New Brunswick

- K-classifiers segmentation

Algorithm 1 K-Classifiers Segmentation Algorithm.

Input: X, Y, K, $loss$.
Output: S, C.

for $n = 1, \cdots, N$ **do**
  $\ell_n \leftarrow rand\{1, \cdots, K\}$.
repeat
  #Update step:
  for $k = 1, \cdots, K$ **do**
    Learn C_k based on $\{x_n, y_n | n \in S_k\}$.
  #Assignment step:
  for $n = 1, \cdots, N$ **do**
    $\ell_n \leftarrow \arg\min_k \text{loss}(x_n, y_n | C_k)$.
until Convergence.

Weakness:

Profile Inconsistent

Algorithm

RUTGERS
Rutgers Business School
Newark and New Brunswick

- Profile-consistent algorithm

Algorithm 2 Profile-Consistent Algorithm.

Input: $X, Y, K, M, loss$. X — data, Y — labels, K — clusters, M — sub-regions

Output: S, C. S — Clusters set, C — Classifiers set

Randomly select M seed points to construct the Voronoi decomposition $T_1, \cdots, T_M$. T — M sub-regions, generated by K-means

for $m = 1, \cdots, M$ **do**

  $\ell_m^t \leftarrow rand\{1, \cdots, K\}$. l — segment assignment; assign randomly.

repeat

  #Update step:

  for $k = 1, \cdots, K$ **do** Learn classifiers based on each cluster

    Learn C_k based on $\{x_n, y_n | n \in S_k\}$.

  #Assignment step:

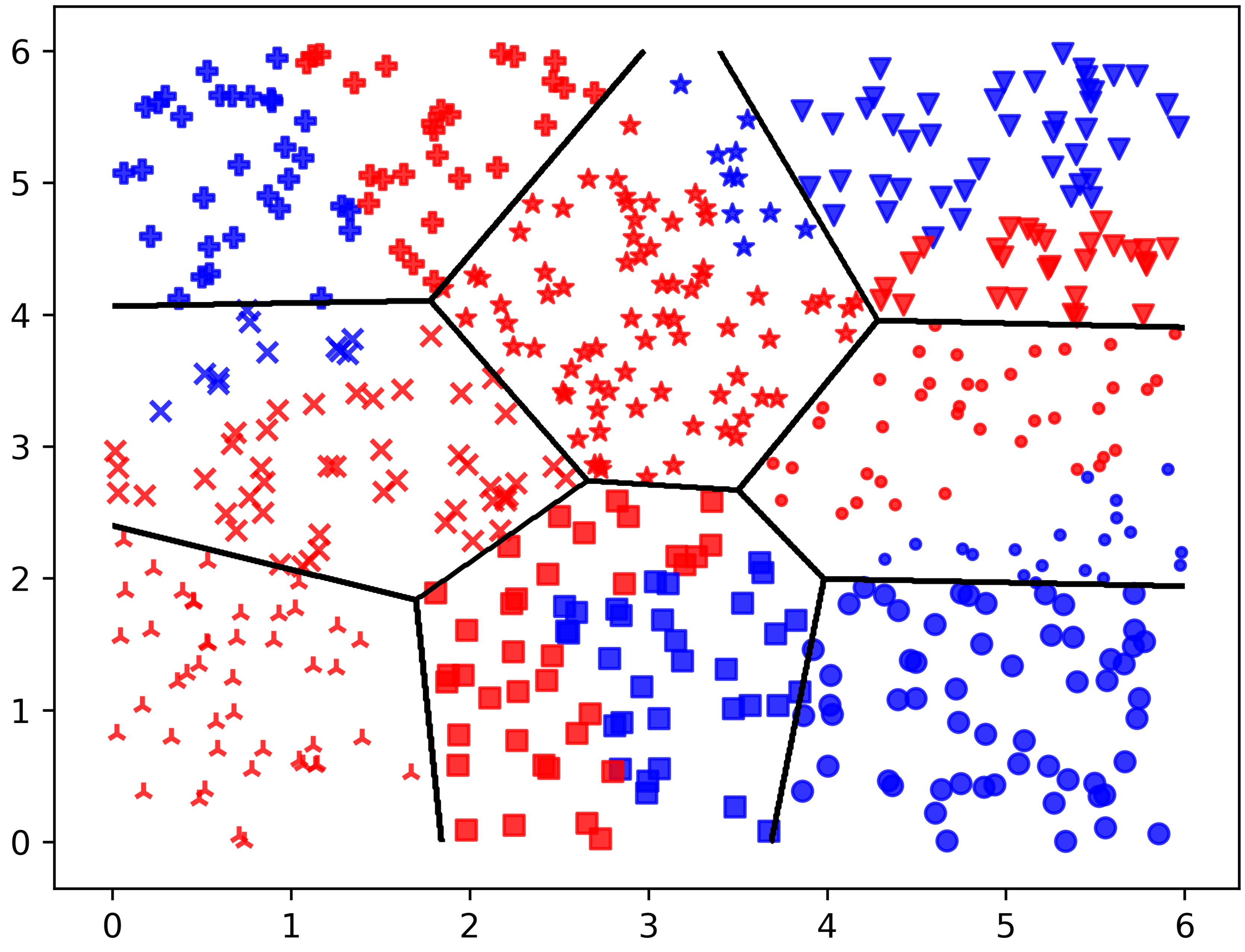
  for $m = 1, \cdots, M$ **do** Re-assign sub-regions to cluster based on the loss on corresponding classifiers

    $\ell_m^t \leftarrow \arg\min_k \sum_{n \in T_m} \text{loss}(x_n, y_n | C_k)$.

until Convergence.

Sub-region Generate

Rutgers Business School
Newark and New Brunswick



Blue: y=1

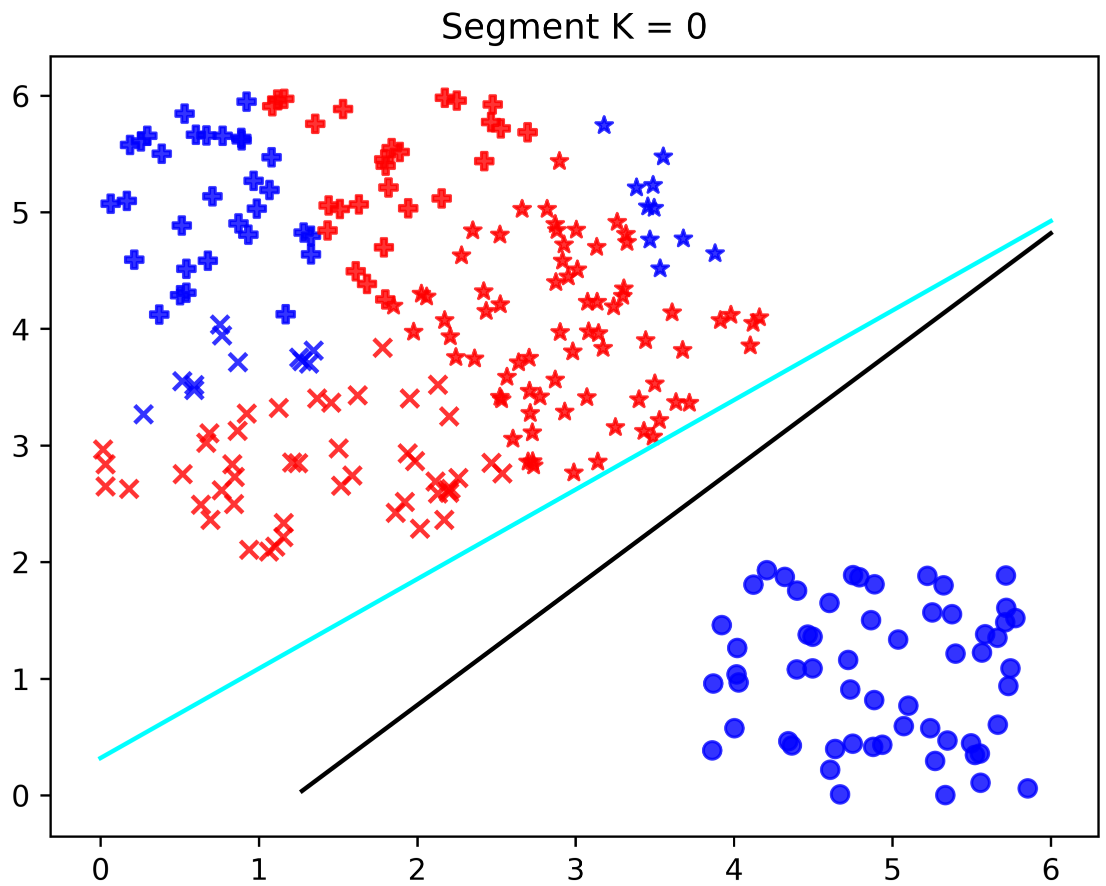
Red: y=-1

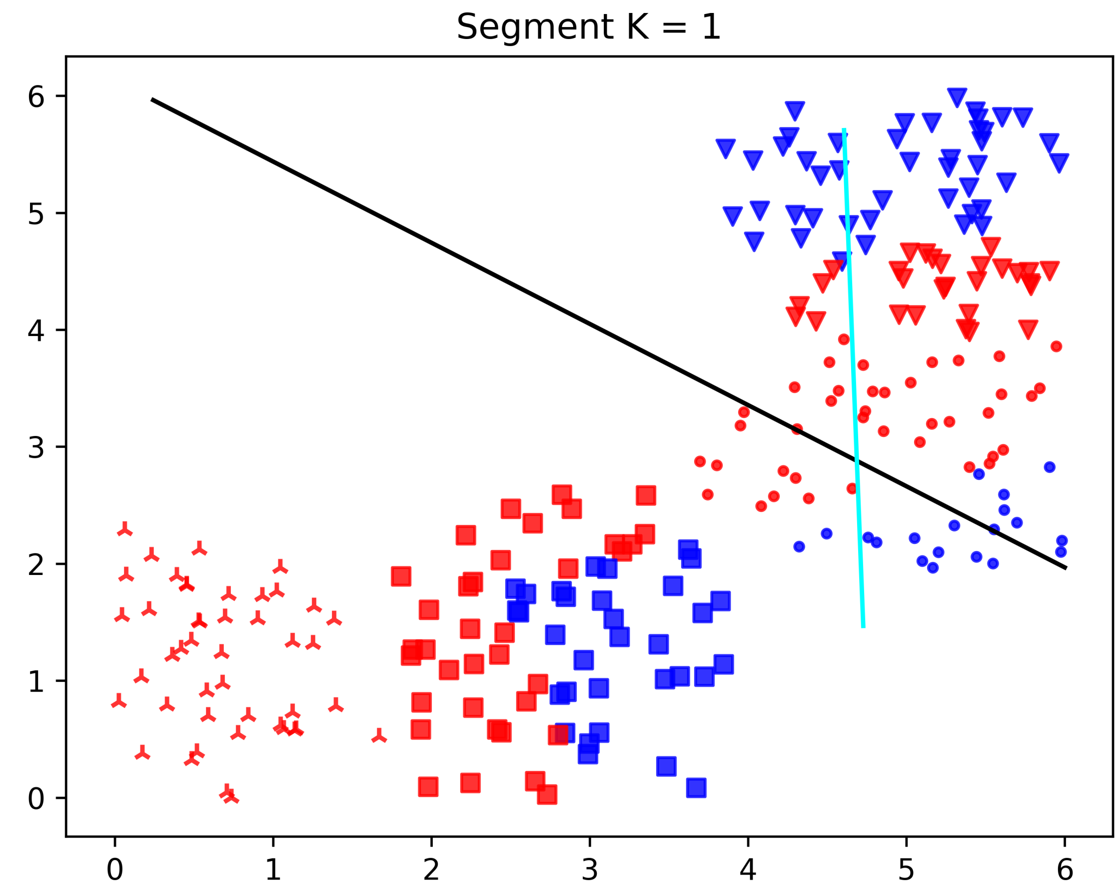
border lines:

separate lines among subregions

Initial Training

Rutgers
Rutgers Business School
Newark and New Brunswick





Black line: SVM

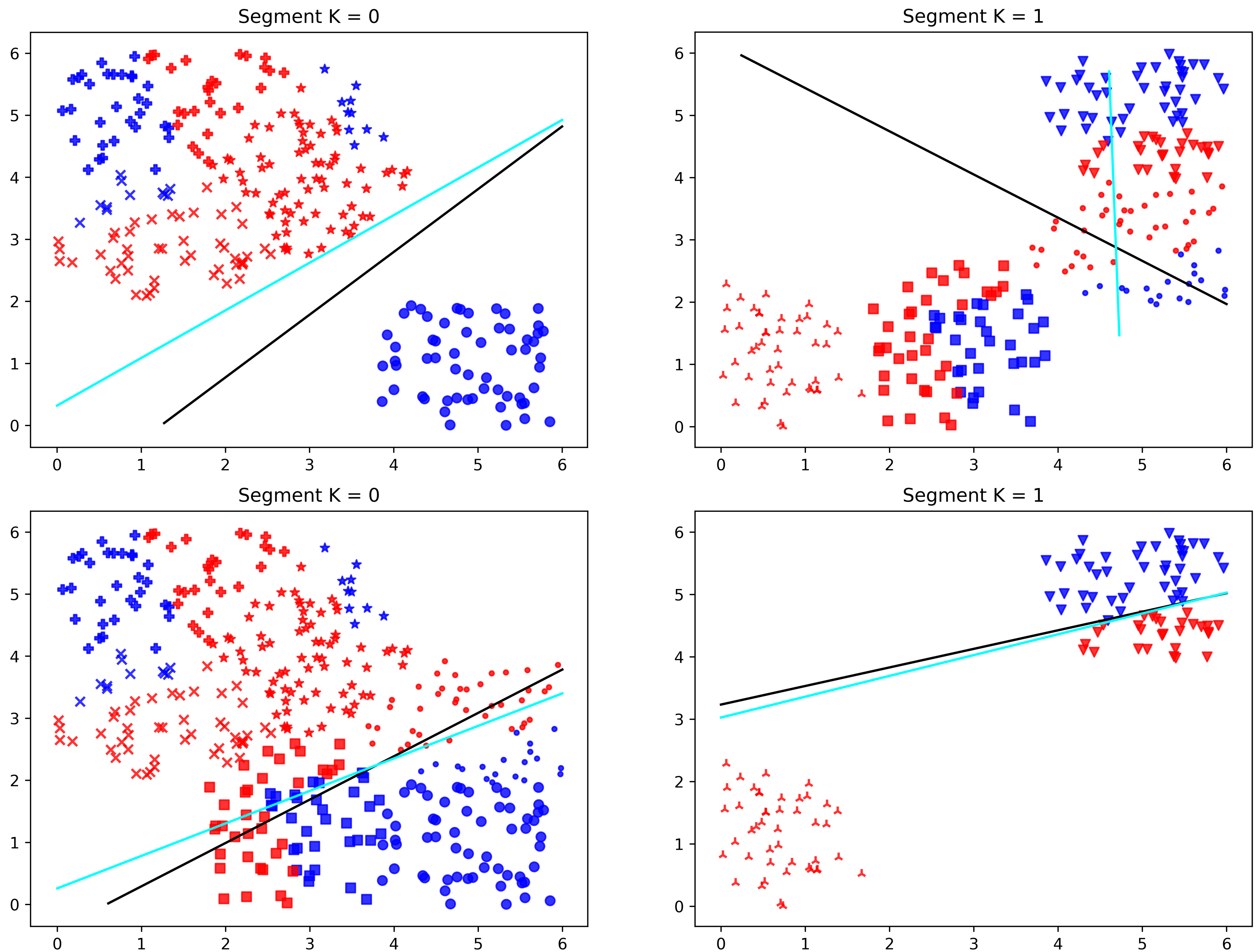
Cyan line: logit regression

First Round Update

Rutgers Business School
Newark and New Brunswick

Step 0:

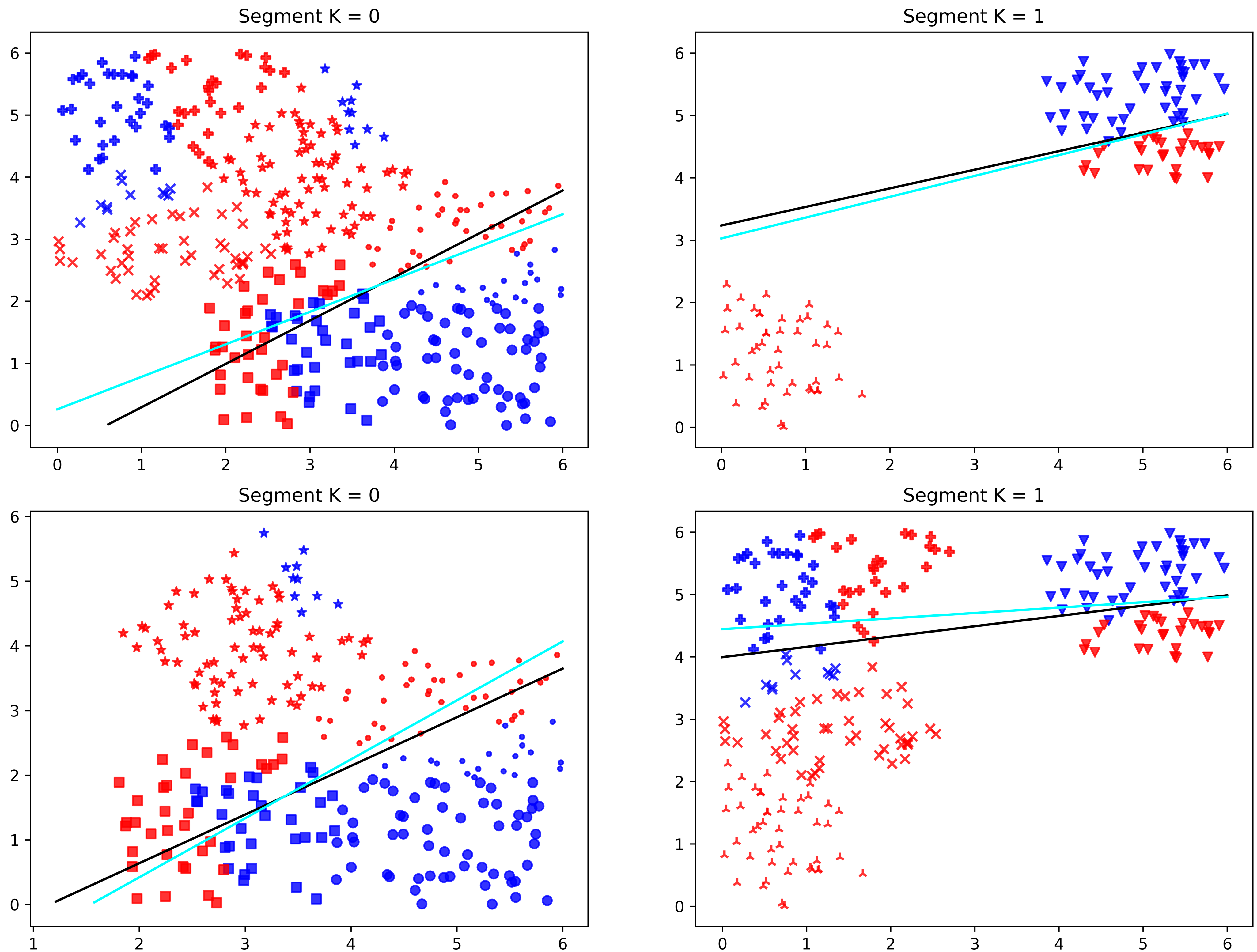
Step 1:



Second Round Update

Step 1:

Step 2:
(converge)



Performances

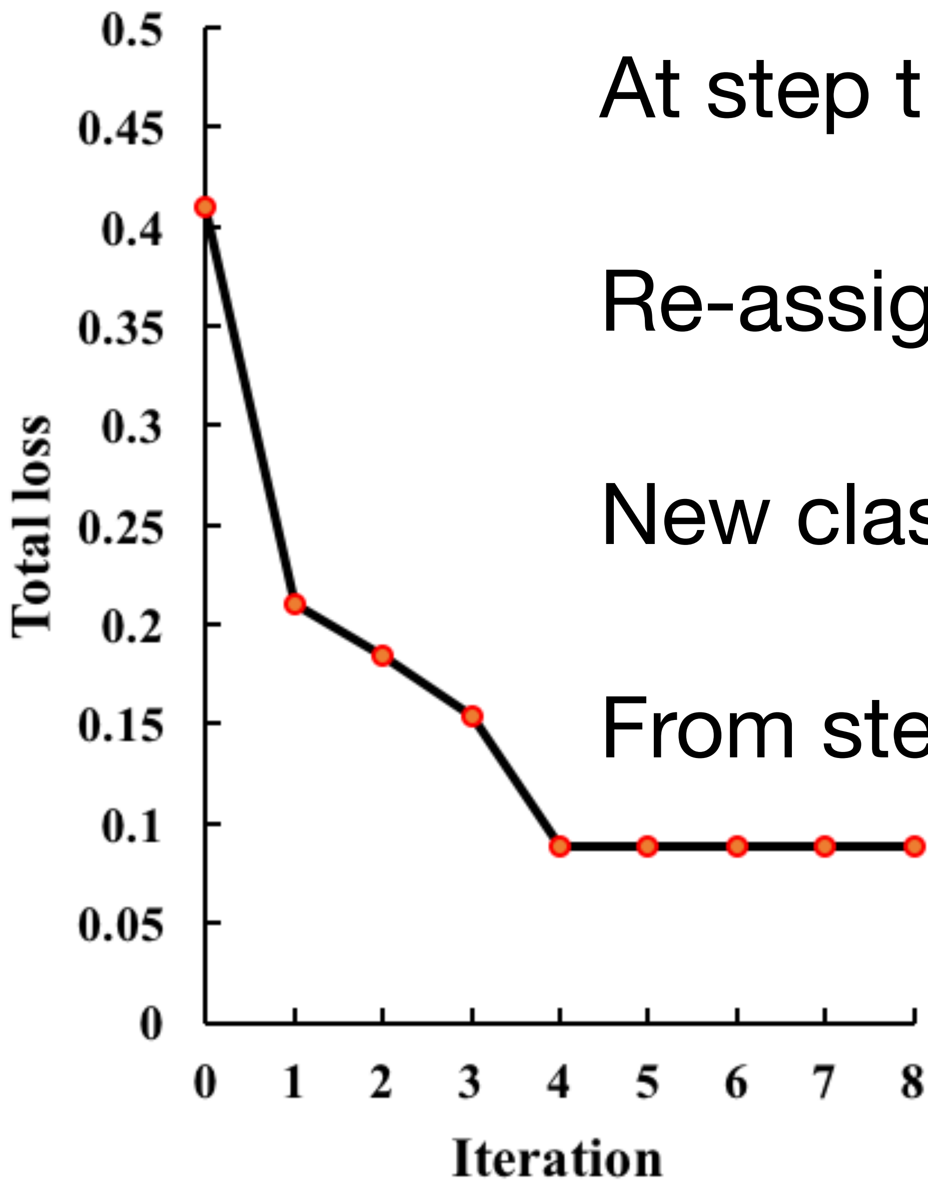
RUTGERS
Rutgers Business School
Newark and New Brunswick

		$Accuracy$	$Precision$	$Recall$	$F-measure$	AP	AUC
Synthetic (K=2, M=8)	LR	0.6466 ± 0.07	0.7295 ± 0.05	0.7492 ± 0.14	0.7325 ± 0.07	0.8226 ± 0.03	0.5964 ± 0.08
	SW_{LR}	0.8800 ± 0.06	0.9441 ± 0.05	0.8744 ± 0.08	0.9052 ± 0.05	0.9510 ± 0.03	0.8831 ± 0.06
	KC_{LR}	0.5296 ± 0.16	0.6861 ± 0.15	0.5279 ± 0.16	0.5918 ± 0.15	0.7636 ± 0.10	0.5299 ± 0.17
	$\mathbf{PC_{LR}}$	$\mathbf{0.9499} \pm 0.05$	$\mathbf{0.9651} \pm 0.03$	$\mathbf{0.9600} \pm 0.05$	$\mathbf{0.9618} \pm 0.04$	$\mathbf{0.9759} \pm 0.02$	$\mathbf{0.9450} \pm 0.05$
	SVM	0.7039 ± 0.08	0.8272 ± 0.14	0.7692 ± 0.18	0.7686 ± 0.07	0.8751 ± 0.05	0.6746 ± 0.13
	SW_{SVM}	0.7935 ± 0.10	0.9018 ± 0.12	0.8092 ± 0.16	0.8345 ± 0.10	0.9188 ± 0.05	0.7846 ± 0.14
	KC_{SVM}	0.5873 ± 0.10	0.8630 ± 0.13	0.4834 ± 0.17	0.5897 ± 0.14	0.8444 ± 0.05	0.6367 ± 0.09
	$\mathbf{PC_{SVM}}$	$\mathbf{0.9268} \pm 0.06$	$\mathbf{0.9688} \pm 0.03$	$\mathbf{0.9200} \pm 0.07$	$\mathbf{0.9422} \pm 0.05$	$\mathbf{0.9710} \pm 0.02$	$\mathbf{0.9300} \pm 0.05$
		$Accuracy$	$Precision$	$Recall$	$F-measure$	AP	AUC
Prod-A (K=5, M=60)	LR	0.8922 ± 0.01	0.7737 ± 0.02	0.6536 ± 0.03	0.7081 ± 0.03	0.7483 ± 0.02	0.8028 ± 0.01
	SW_{LR}	0.8942 ± 0.01	0.7798 ± 0.03	0.6554 ± 0.03	0.7132 ± 0.02	0.7530 ± 0.02	$\mathbf{0.8055} \pm 0.01$
	KC_{LR}	0.5547 ± 0.13	0.2774 ± 0.20	0.5406 ± 0.06	0.3471 ± 0.14	0.4549 ± 0.12	0.5494 ± 0.10
	$\mathbf{PC_{LR}}$	$\mathbf{0.8954} \pm 0.01$	$\mathbf{0.7875} \pm 0.05$	$\mathbf{0.6578} \pm 0.04$	$\mathbf{0.7146} \pm 0.03$	$\mathbf{0.7559} \pm 0.03$	0.8054 ± 0.02
	SVM	0.8117 ± 0.05	0.5482 ± 0.13	0.5495 ± 0.09	0.5414 ± 0.09	0.5939 ± 0.08	0.7133 ± 0.05
	SW_{SVM}	0.8255 ± 0.04	0.5813 ± 0.12	$\mathbf{0.5875} \pm 0.10$	0.5738 ± 0.08	0.6257 ± 0.07	$\mathbf{0.7363} \pm 0.04$
	KC_{SVM}	0.5399 ± 0.03	0.2222 ± 0.02	0.5194 ± 0.07	0.3108 ± 0.03	0.4188 ± 0.04	0.5322 ± 0.03
	$\mathbf{PC_{SVM}}$	$\mathbf{0.8504} \pm 0.01$	$\mathbf{0.6635} \pm 0.04$	0.5165 ± 0.09	$\mathbf{0.5766} \pm 0.06$	$\mathbf{0.6383} \pm 0.04$	0.7252 ± 0.04
		$Accuracy$	$Precision$	$Recall$	$F-measure$	AP	AUC
Prod-B (K=3, M=150)	LR	0.9277 ± 0.01	0.8555 ± 0.01	0.7682 ± 0.05	0.8087 ± 0.03	0.8350 ± 0.02	0.8679 ± 0.02
	SW_{LR}	0.9281 ± 0.01	$\mathbf{0.8557} \pm 0.01$	0.7705 ± 0.05	0.8102 ± 0.03	0.8360 ± 0.02	0.8690 ± 0.02
	KC_{LR}	0.6053 ± 0.12	0.3120 ± 0.19	0.5737 ± 0.04	0.3864 ± 0.12	0.4855 ± 0.11	0.5935 ± 0.08
	$\mathbf{PC_{LR}}$	$\mathbf{0.9292} \pm 0.00$	0.8556 ± 0.01	$\mathbf{0.7771} \pm 0.02$	$\mathbf{0.8143} \pm 0.01$	$\mathbf{0.8386} \pm 0.01$	$\mathbf{0.8721} \pm 0.01$
	SVM	0.8551 ± 0.02	0.6313 ± 0.07	$\mathbf{0.7042} \pm 0.07$	0.6610 ± 0.04	0.6973 ± 0.03	0.7985 ± 0.03
	SW_{SVM}	0.8534 ± 0.03	0.6285 ± 0.07	0.6830 ± 0.09	0.6500 ± 0.06	0.6875 ± 0.05	0.7895 ± 0.04
	KC_{SVM}	0.5346 ± 0.07	0.2259 ± 0.03	0.5347 ± 0.09	0.3157 ± 0.04	0.4268 ± 0.05	0.5346 ± 0.05
	$\mathbf{PC_{SVM}}$	$\mathbf{0.8961} \pm 0.01$	$\mathbf{0.7736} \pm 0.04$	0.6871 ± 0.09	$\mathbf{0.7230} \pm 0.05$	$\mathbf{0.7616} \pm 0.03$	$\mathbf{0.8177} \pm 0.04$

LR: logit regression, SW: segment wised step-by-step method

Converge Analysis

RUTGERS
Rutgers Business School
Newark and New Brunswick



At step t: $loss(C_i^t, S_i^t)$

Re-assignment: $loss(C_i^t, S_i^{t+1}) \leq loss(C_i^t, S_i^t)$

New classification: $loss(C_i^{t+1}, S_i^{t+1}) \leq loss(C_i^t, S_i^{t+1})$

From step t to t+1: $loss(C_i^{t+1}, S_i^{t+1}) \leq loss(C_i^t, S_i^t)$

But, converge to what?

Convex Analysis

Rutgers Business School
Newark and New Brunswick

- Sub-region generation — K-means
  - Non-convex
  - For n data records, k clusters:

1. Choose k cluster centers randomly generated in a domain containing all the points
2. Assign each point to the closest cluster center
3. Recompute the cluster centers using the current cluster memberships
4. If a convergence criterion is met, stop; Otherwise go to step 2.

K-means

Rutgers
Rutgers Business School
Newark and New Brunswick

- Set the assignment matrix: $X = [x_{ij}] \in R^{N\times M}$; if x_i is assigned to T_j then $x_{ij} = 1$, otherwise $x_{ij} = 0$
- For a fixed cluster (sub-region) T_j, cluster center as the mean of all the points in the cluster:

$$\min_{c} \sum_{i=1}^{|T_j|} \|x_i^{(j)} - c_j\|^2 \Rightarrow c_j = \frac{1}{|T_j|} \sum_{i=1}^{|T_j|} x_i^{(j)}$$

$$\Rightarrow c_j = (\sum_{l=1}^{n} x_{lj} x_l) / (\sum_{l=1}^{n} x_{lj})$$

K-means

Rutgers Business School
Newark and New Brunswick

$$\min_{x_{ij}} \quad \sum_{j=1}^{M} \sum_{i=1}^{N} x_{ij} \| x_j - \frac{\sum_{l=1}^{n} x_{lj} x_l}{\sum_{l=1}^{n} x_{lj}} \|^2$$

$$s.t. \quad \sum_{j=1}^{M} x_{ij} = 1, \quad i = 1, \ldots N$$ each point is assigned to only one cluster

$$\sum_{i=1}^{N} x_{ij} \geq 1, \quad j = 1, \ldots M$$ each cluster contains at least one point

$$x_{ij} \in \{0,1\}, \quad i = 1, \ldots N, \ j = 1, \ldots M$$

Update Step

Rutgers
Rutgers Business School
Newark and New Brunswick

- Set the assignment matrix: $Z = [z_{jk}] \in R^{M \times K}$; if T_j is assigned to S_k then $x_{jk} = 1$, otherwise $x_{jk} = 0$
- For each segment S_k, a classifier C_k will be trained. If we choose logit regression, C_k could be represented by a sigmoid function $\sigma(\omega, x)$ where

$$\omega = \arg\min \sum_{j=1}^{M} z_{jk} \sum_{i=1}^{N} x_{ij} \log[1 + e^{-y \cdot \sigma(\omega^T x)}]$$

Update Step

Rutgers Business School
Newark and New Brunswick

$$\min_{x_{ij}, z_{jk}} \quad \sum_{k=1}^{K} \sum_{j=1}^{M} z_{jk} \sum_{i=1}^{N} x_{ij} \log[1 + e^{-y \cdot \sigma(\omega^T x)}]$$

$$s.t. \quad \omega = \arg\min \sum_{j=1}^{M} z_{jk} \sum_{i=1}^{N} x_{ij} \log[1 + e^{-y \cdot \sigma(\omega^T x)}]$$

$$\sum_{j=1}^{M} x_{ij} = 1,\ i = 1,\ldots N; \sum_{i=1}^{N} x_{ij} \geq 1,\ j = 1,\ldots M$$

$$\sum_{k=1}^{K} z_{jk} = 1,\ j = 1,\ldots M; \sum_{j=1}^{M} z_{jk} \geq 1,\ k = 1,\ldots K$$

$$x_{ij}, z_{jk} \in \{0,1\}, \quad i = 1,\ldots N,\ j = 1,\ldots M,\ k = 1,\ldots, K$$

Rutgers Business School
Newark and New Brunswick

Thanks