PRanking with Ranking

Koby Crammer

Technion – Israel Institute of Technology

Based on joint work with Yoram Singer at the Hebrew University of Jerusalem

Problem

Machine Prediction	★★	★★★★★	★	★★	★★★★
User's Rating	★★★★★	★★	★	★★★★★	★★★
Ranking Loss	3	3	0	3	1

Ranking – Formal Description

Problem Setting

- Instances $\mathbf{x} \in \chi$
- Labels $y \in \{1, 2, \ldots, k\}$
- Structure $1 \prec 2 \prec 3 \prec 4 \prec 5$
- Ranking rule $f : \chi \rightarrow \{1, 2, \ldots, k\}$

x_1 is preferred over x_2

$\Updownarrow$

$$f(\mathbf{x}_1) \succ f(\mathbf{x}_2)$$

Online Framework

- Algorithm works in rounds
- On each round the ranking algorithm :
  - Gets an input instance
  - Outputs a rank as prediction
  - Receives the correct rank-value
  - Computes loss
  - Updates the rank-prediction rule

Goal

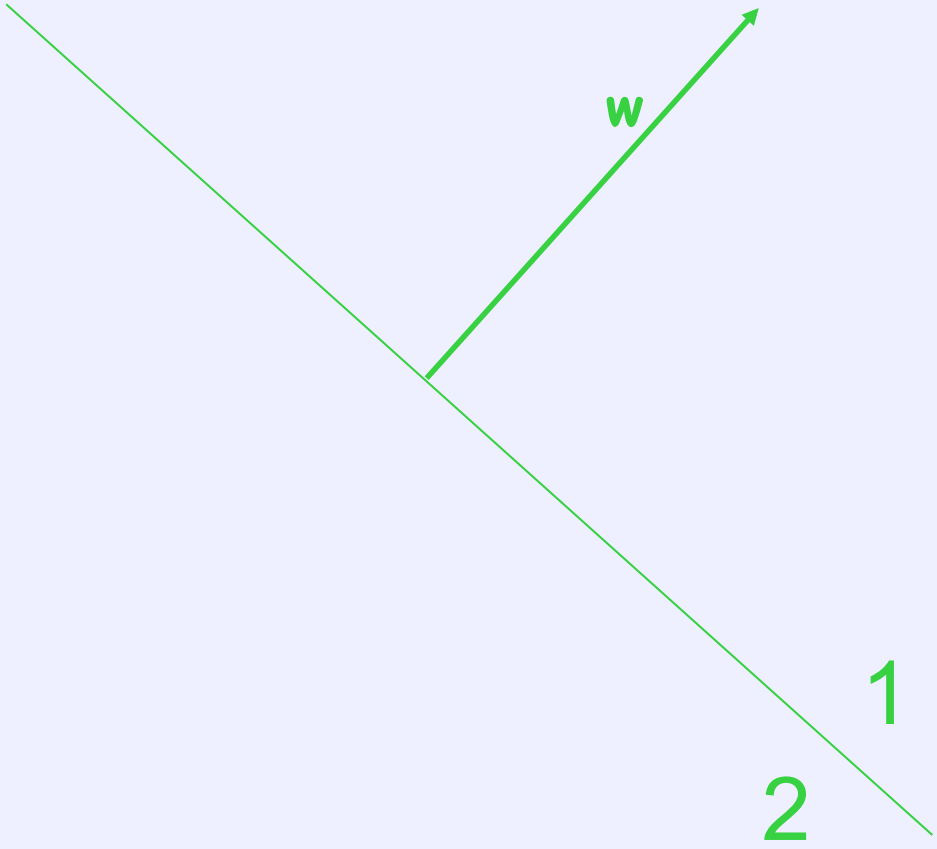
- Algorithms Loss $$L_t = \sum_{i=1}^{t} \left| y_i - \hat{y}_i \right|$$
- Loss of a fixed function $$L_t(f) = \sum_{i=1}^{t} \left| y_i - f(x_i) \right|$$
- Regret $$L_t - \inf_{f \in F} L_t(f)$$
- No statistical assumptions over data
- The algorithm should do well irrespectively of specific sequence of inputs and target labels

Background

Binary Classification

The Perceptron Algorithm

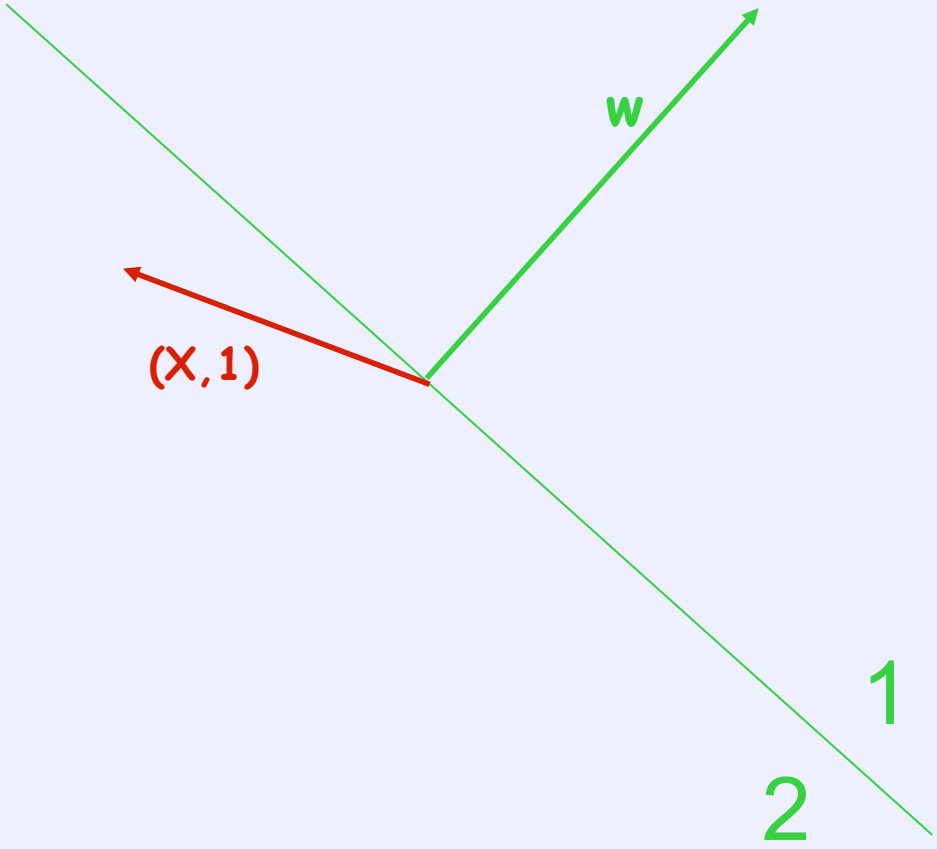
Rosenblatt, 1958



- Hyperplane **w**

The Perceptron Algorithm

Rosenblatt, 1958

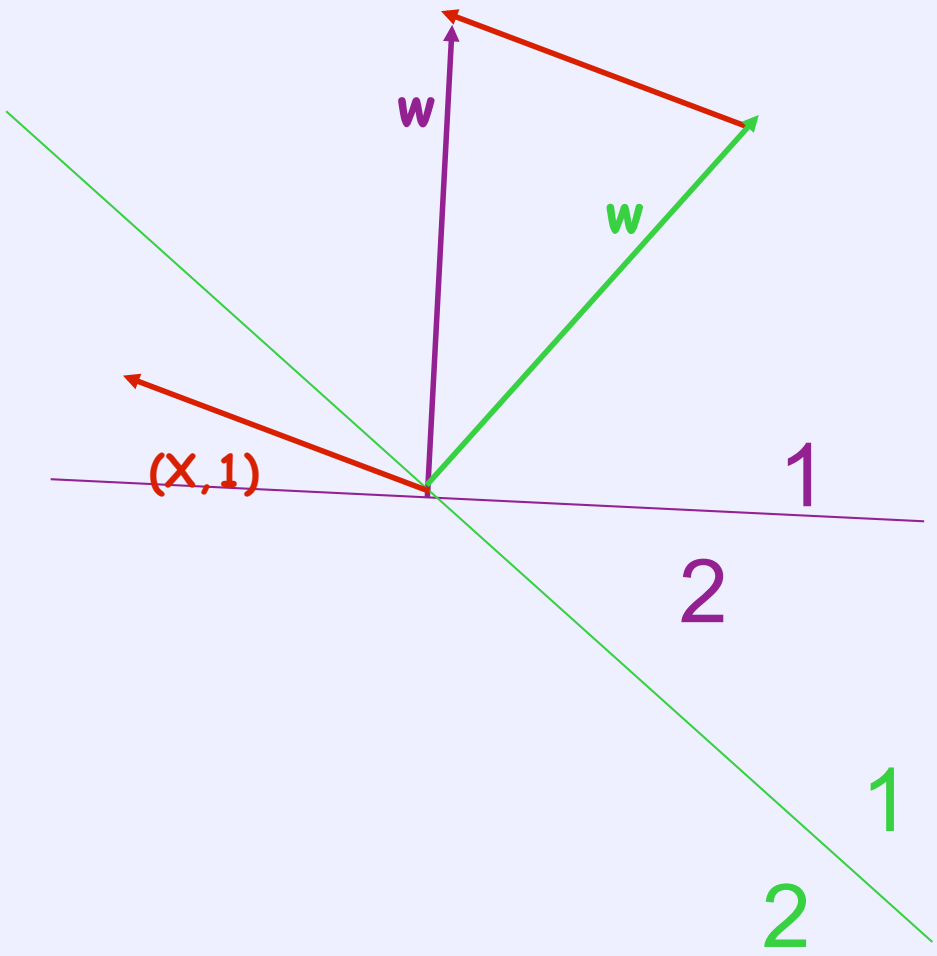


- Hyperplane $\mathbf{w}$
- Get new instance $\mathbf{x}$
- Classify $\mathbf{x}$:

$$\text{sign}(\mathbf{w} \cdot \mathbf{x})$$

The Perceptron Algorithm

Rosenblatt, 1958



- Hyperplane $\mathbf{w}$
- Get new instance $\mathbf{x}$
- Classify $\mathbf{x}$:

  $$\text{sign}(\mathbf{w} \cdot \mathbf{x})$$

- Update (in case of a mistake)

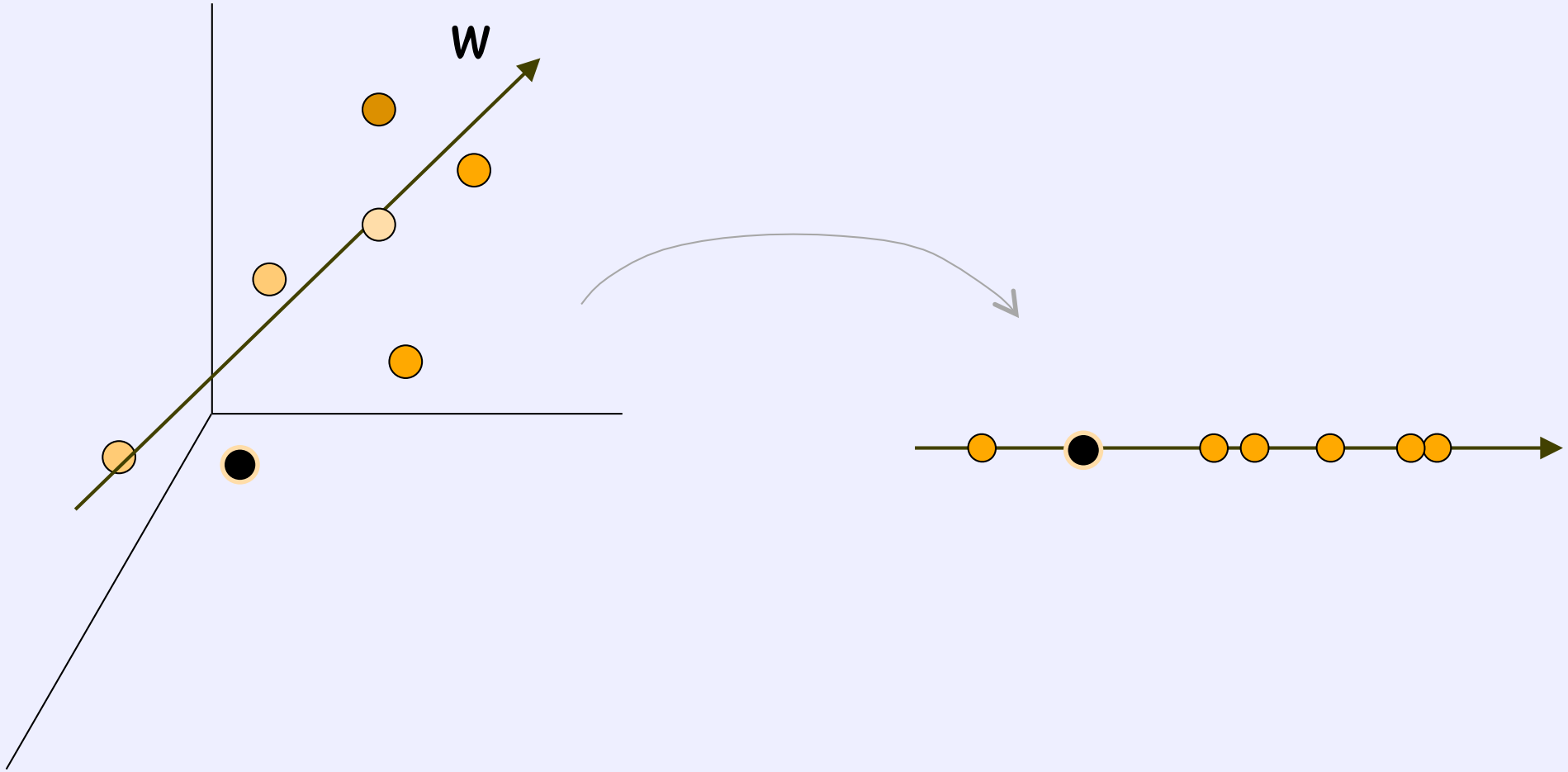
  $$\mathbf{w} \leftarrow \mathbf{w} + \mathbf{x}$$

A Function Class for Ranking

Our Approach to Ranking

- Project

$$x \rightarrow w \cdot x$$



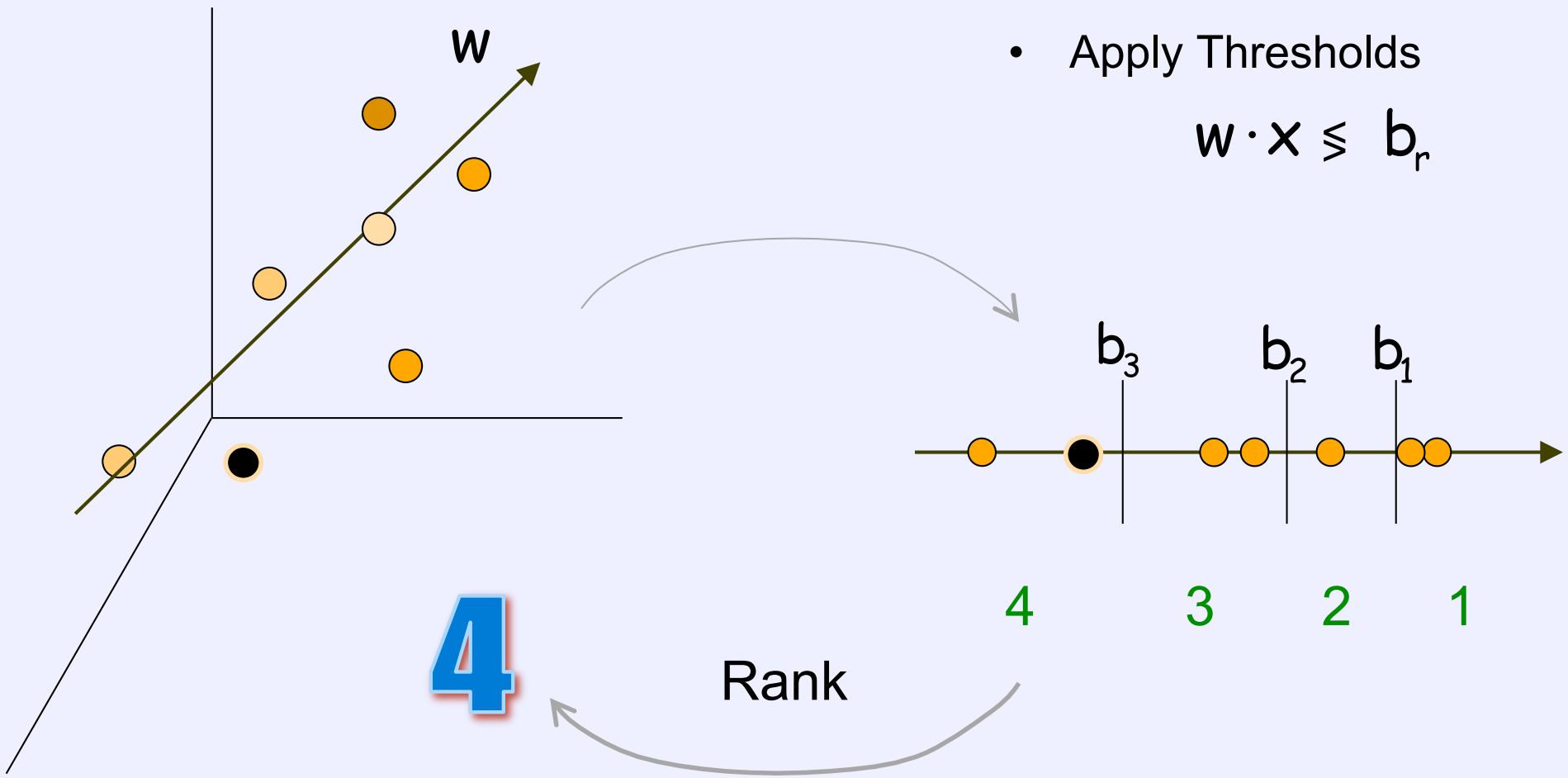
Our Approach to Ranking

- Project

$$x \rightarrow w \cdot x$$

- Apply Thresholds

$$w \cdot x \lessgtr b_r$$



Update of a Specific Algorithm

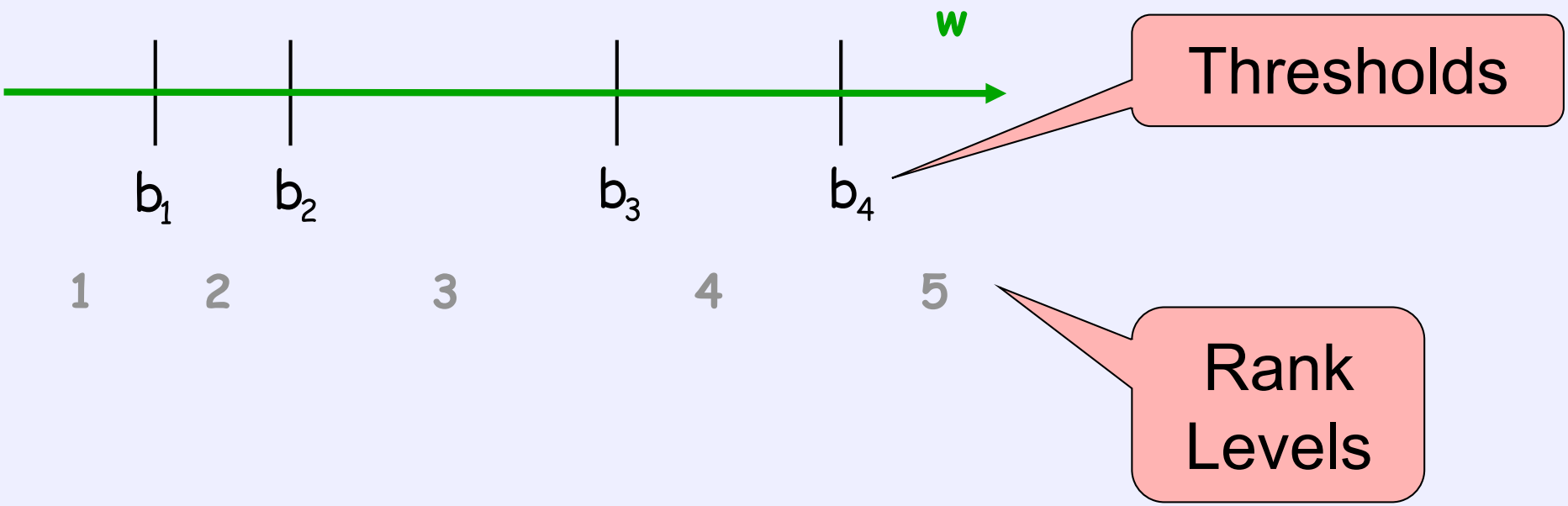
~~FIX~~ it if its not Broken

Least change as possible

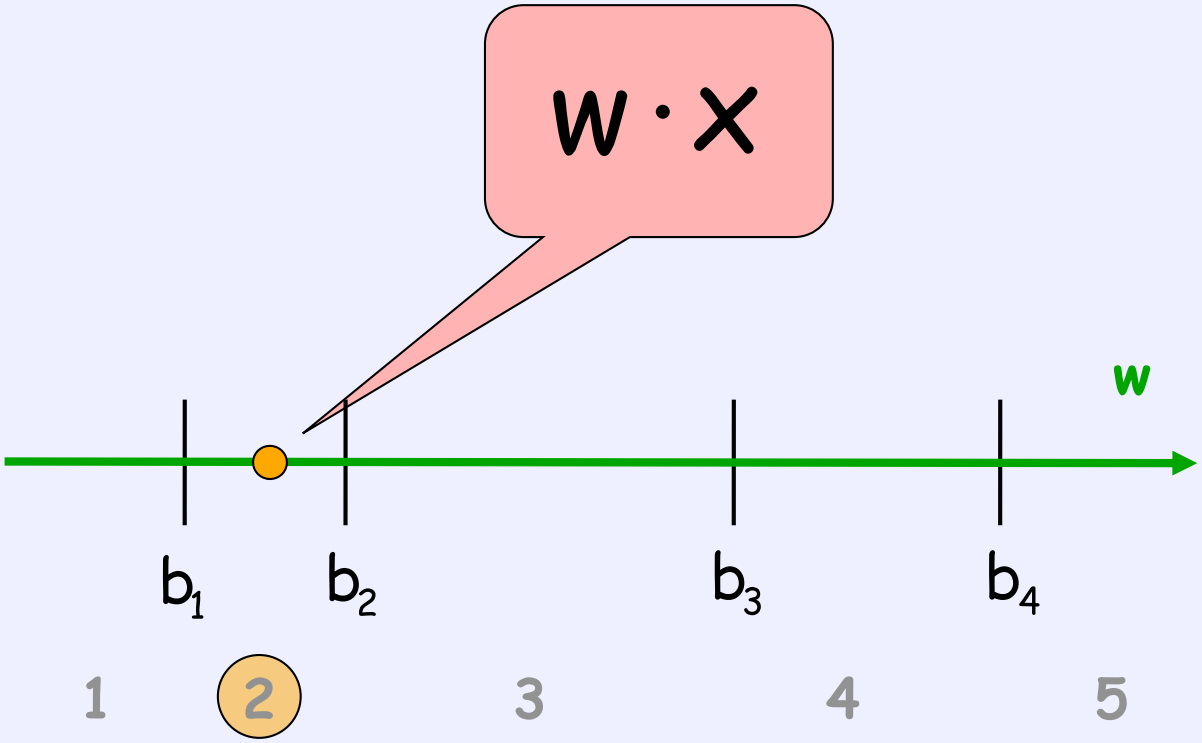
One step at a time

PRank

- Direction **w**,
  Thresholds $\mathbf{b}_{k-1}, \ldots, \mathbf{b}_1$

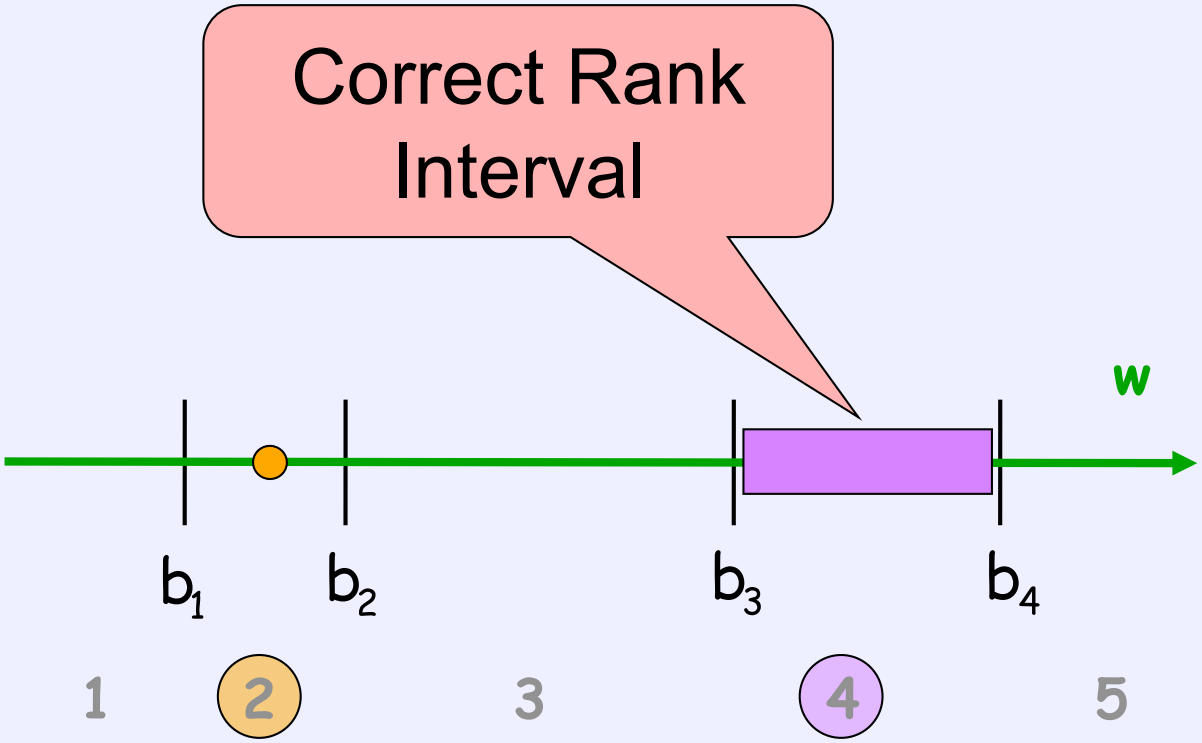


PRank



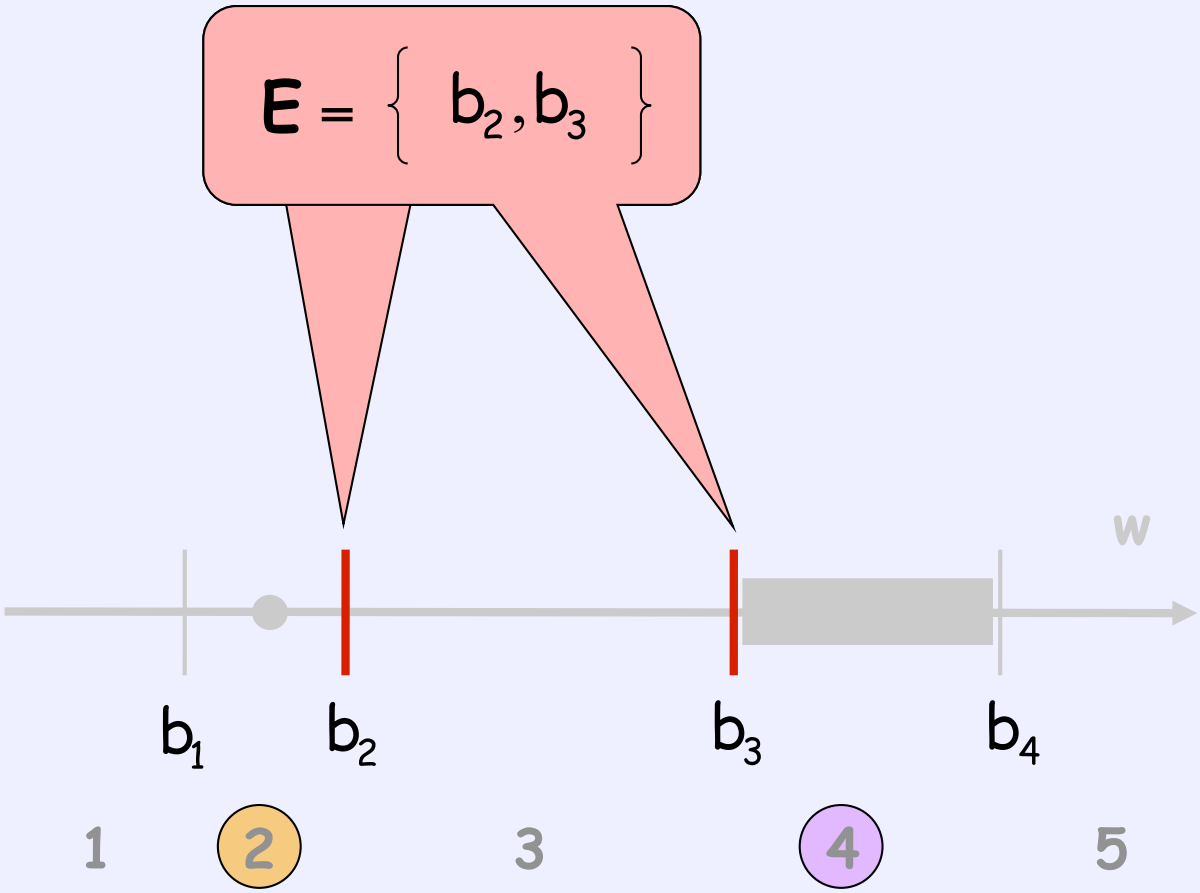
- Direction **w**,
  Thresholds $\mathbf{b}_{k-1}, \ldots, \mathbf{b}_1$
- Rank a new instance **x**

PRank



- Direction **w**, Thresholds $\mathbf{b}_{k-1}, \ldots, \mathbf{b}_1$
- Rank a new instance **x**
- Get the correct rank **y**

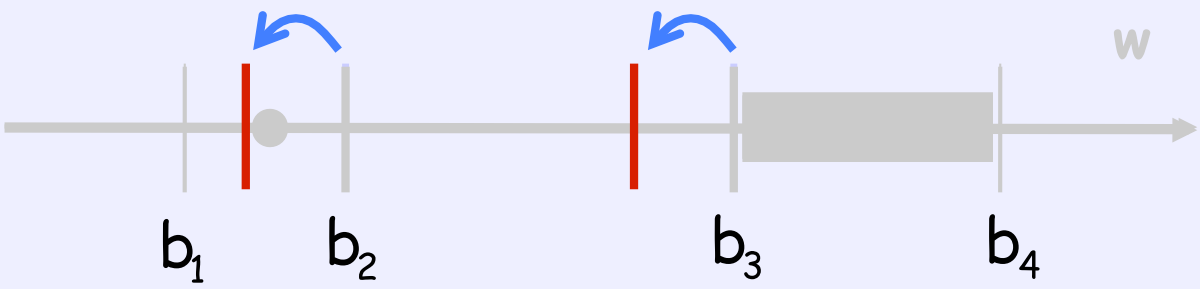
PRank



- Direction **w**, Thresholds $\mathbf{b}_{k-1}, \ldots, \mathbf{b}_1$
- Rank a new instance **x**
- Get the correct rank **y**
- Compute Error-Set **E**

PRank – Update

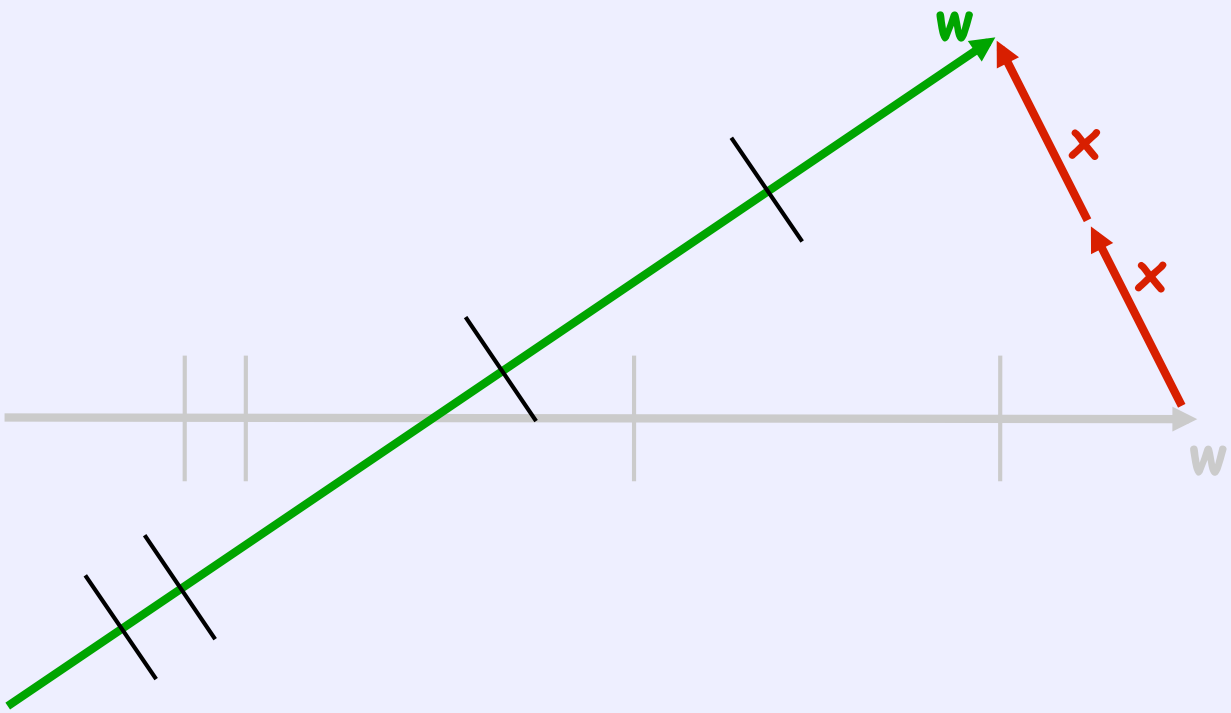
$$\mathbf{E} = \{ b_2, b_3 \}$$



- Direction **w**,
  Thresholds $\mathbf{b}_{k-1}, \ldots, \mathbf{b}_1$
- Rank a new instance **x**
- Get the correct rank **y**
- Compute Error-Set **E**
- Update :
  - $b_r \leftarrow b_r - 1 \quad r \in \mathbf{E}$

PRank – Update

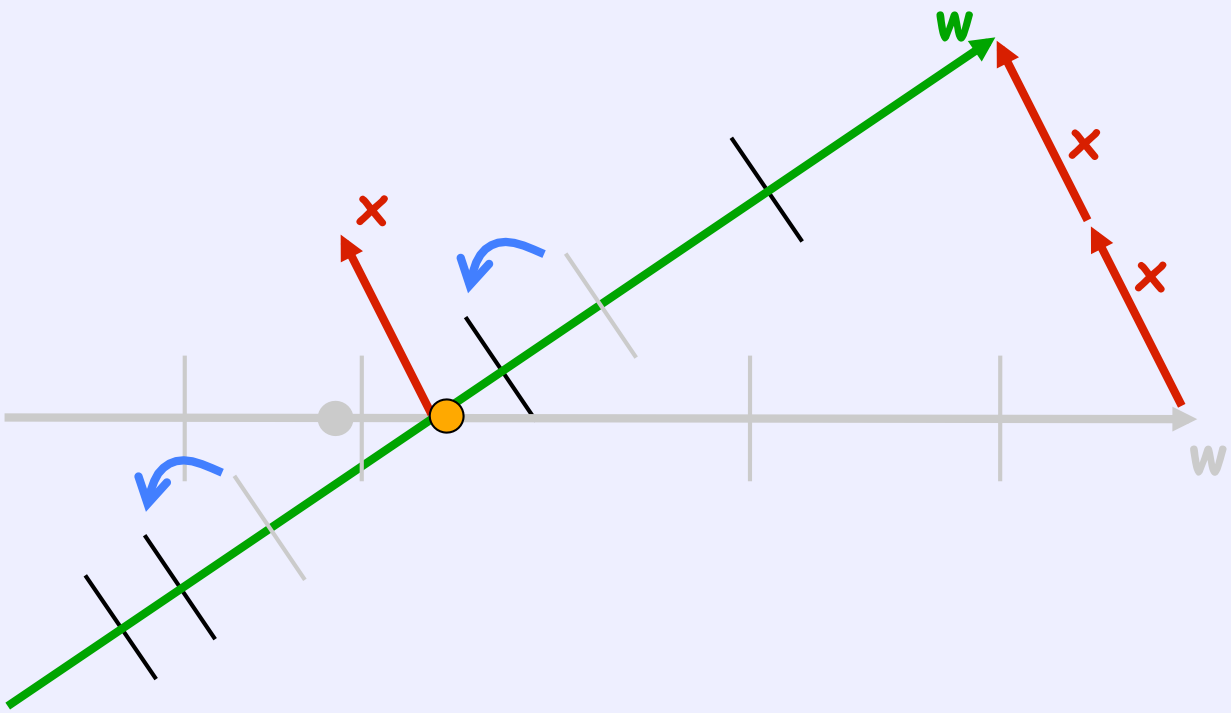
$$\mathbf{E} = \{ b_2, b_3 \}$$



- Direction **w**, Thresholds $\mathbf{b}_{k-1}, \ldots, \mathbf{b}_1$
- Rank a new instance **x**
- Get the correct rank **y**
- Compute Error-Set **E**
- Update :
  - $b_r \leftarrow b_r - 1 \quad r \in \mathbf{E}$
  - $\mathbf{w} \leftarrow \mathbf{w} + |\mathbf{E}|\mathbf{x}$

PRank – Summary of Update

$$\mathbf{E} = \{ b_2, b_3 \}$$



- Direction **w**, Thresholds $\mathbf{b}_{k-1}, \ldots, \mathbf{b}_1$
- Rank a new instance **x**
- Get the correct rank **y**
- Compute Error-Set **E**
- Update :
  - $b_r \leftarrow b_r - 1 \quad r \in \mathbf{E}$
  - $\mathbf{w} \leftarrow \mathbf{w} + |\mathbf{E}|\mathbf{x}$

The PRank Algorithm

Maintain $\mathbf{w}, b_1, \ldots, b_{k-1}$

Get an instance $\mathbf{x}$

Predict :
$\hat{y} = 1 + \#\{r : \mathbf{w} \cdot \mathbf{x} < b_r\}$

Get the true rank y

Compute Error set :
$E = \{r : y \le r < \hat{y}\}$

$E \neq \Phi$?

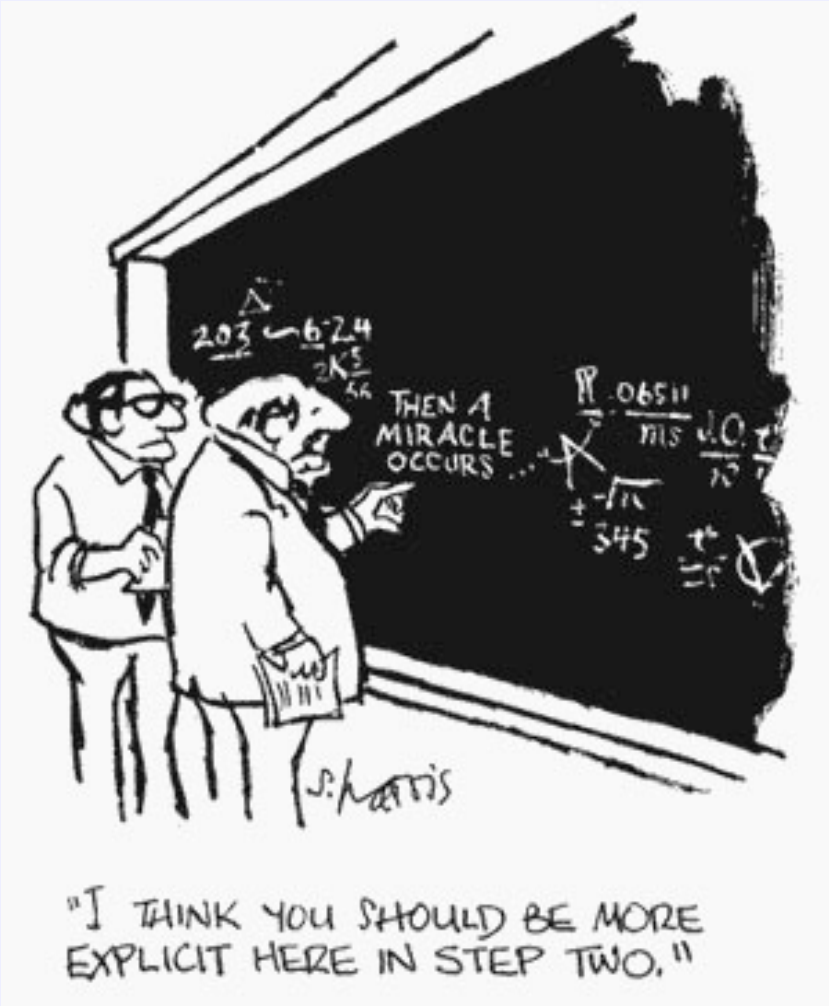
No

Yes

Update

$\mathbf{w} \leftarrow \mathbf{w} + |E|\mathbf{x}$

$b_r \leftarrow b_r - 1 \quad r \in E$

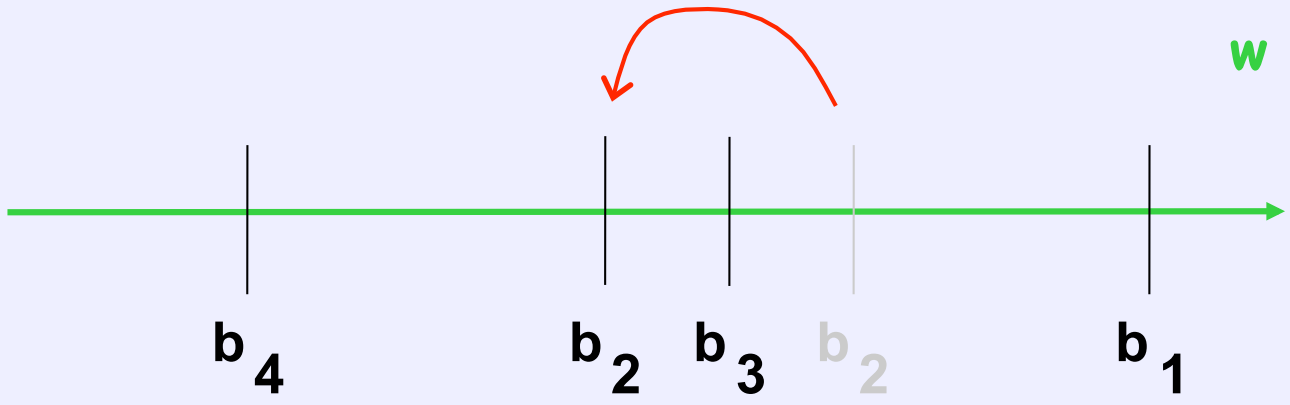


Analysis

Two Lemmas

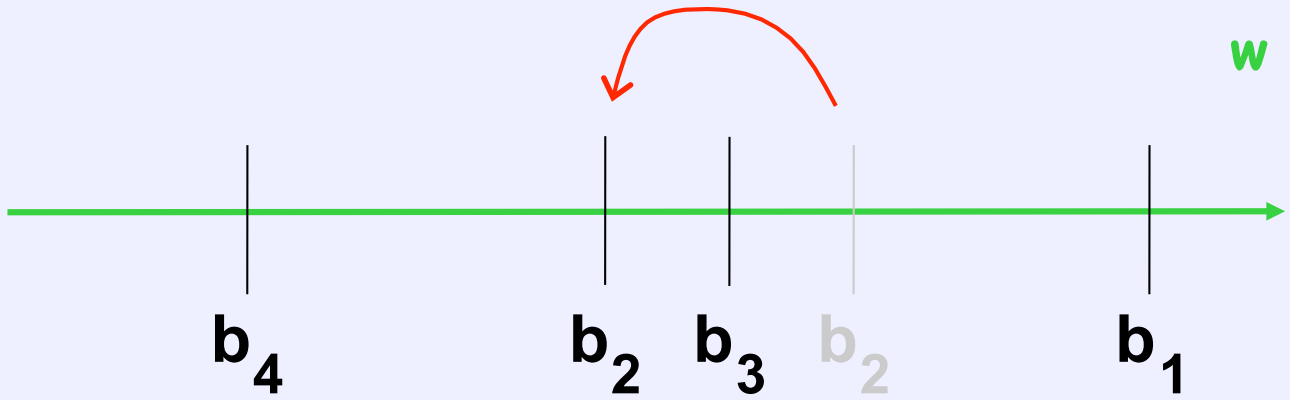
Consistency

- Can the following happen?



Consistency

- Can the following happen? No



- The order of the thresholds is preserved after each round of PRank : $b_1 \leq \ldots \leq b_{k-1}$.

Regret Bound

Given :

- Arbitrary input sequence

Easy Case:

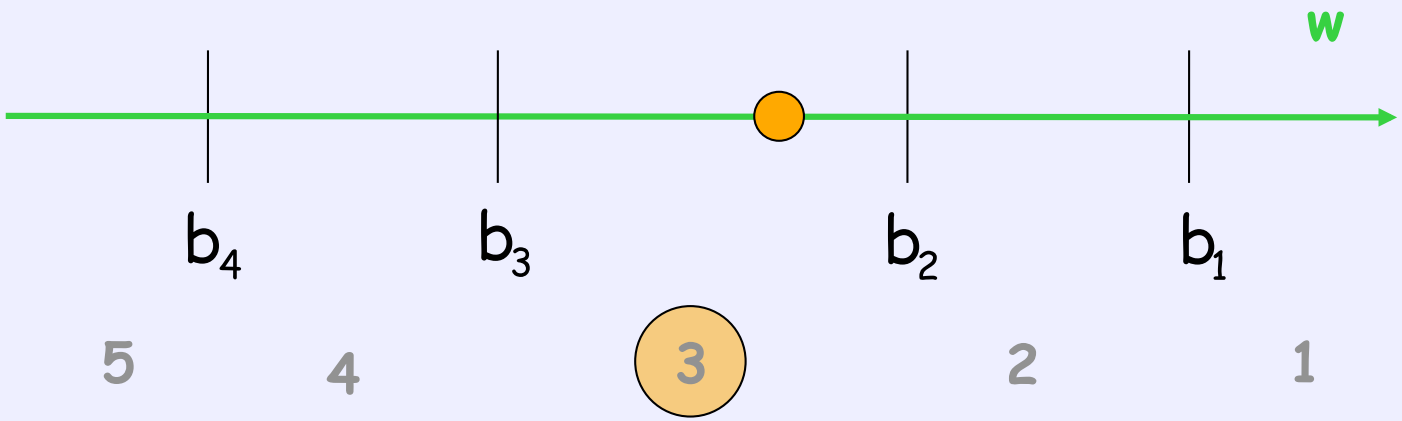
- Assume there exists a model that ranks all the input instances correctly
  - The total loss the algorithm suffers is bounded

Hard Case:

- In general
  - A "regret" $L_t - \inf_{f \in F} \tilde{L}_t(f)$ is bounded

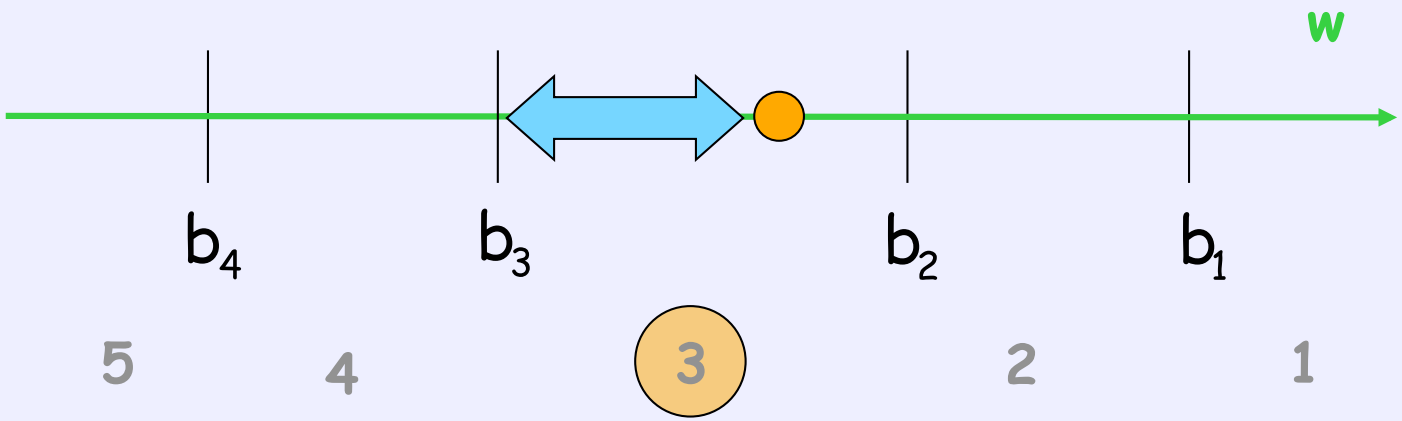
Ranking Margin

$$\text{Margin}(\mathbf{x},y) = \min\left\{ \min_{r \geq y}\{\mathbf{w}\cdot\mathbf{x} - b_r\},\ \min_{r<y}\{b_r - \mathbf{w}\cdot\mathbf{x}\} \right\}$$



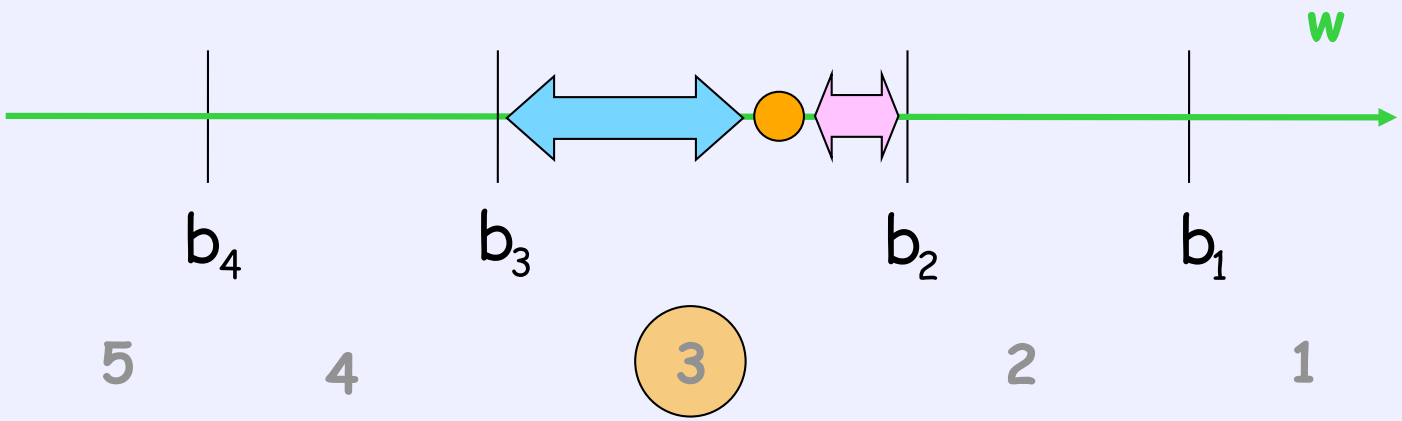
Ranking Margin

$$\text{Margin}(\mathbf{x},y) = \min\left\{ \min_{r \geq y}\{\mathbf{w}\cdot\mathbf{x} - b_r\},\ \min_{r<y}\{b_r - \mathbf{w}\cdot\mathbf{x}\} \right\}$$



Ranking Margin

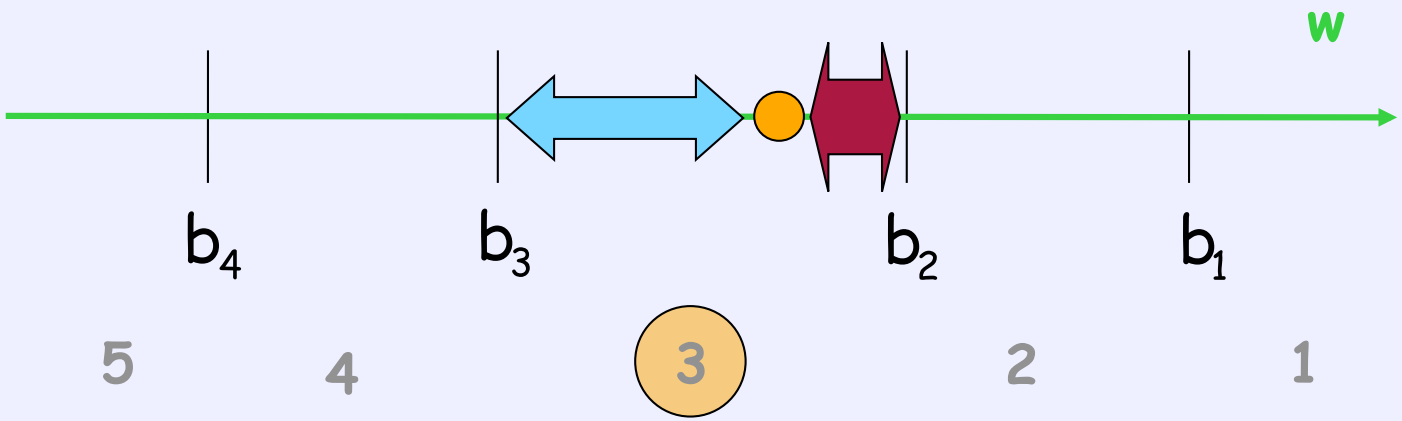
$$\text{Margin}(\mathbf{x},y) = \min\left\{ \min_{r \geq y}\{\mathbf{w}\cdot\mathbf{x} - b_r\},\ \min_{r<y}\{b_r - \mathbf{w}\cdot\mathbf{x}\} \right\}$$



Ranking Margin

$$\text{Margin}(\mathbf{x},y) = \min\left\{ \min_{r \geq y}\{\mathbf{w}\cdot\mathbf{x} - b_r\},\ \min_{r<y}\{b_r - \mathbf{w}\cdot\mathbf{x}\} \right\}$$

$$\text{Margin}\left((\mathbf{x}^1, y^1), \ldots, (\mathbf{x}^T, y^T)\right) = \min \text{Margin}\left(\mathbf{x}^t, y^t\right)$$



Mistake Bound

Given :

- Input sequence $(\mathbf{x}^1, y^1), \ldots, (\mathbf{x}^T, y^T)$,
- Norm of instances is bounded $\|\mathbf{x}^t\|^2 \le R^2$
- Ranked correctly by a normalized ranker
  $\|\mathbf{w}\|^2 + b_1^2 + \ldots + b_{k-1}^2 = 1$ with Margin>0

Then :

Number of Mistakes PRank Makes $\le (k-1)\dfrac{R^2+1}{\text{Margin}^2}$

Exploit Structure

	Loss	Range	Structure
Classification	$\langle y \neq \hat{y} \rangle$	$y \in \{1,2,\ldots,k\}$	None Under Constraint
Regression	$(y-\hat{y})^2$	$y \in \Re$	Metric Over Constraint
Ranking	$\lvert y-\hat{y} \rvert$	$y \in \{1,2,\ldots,k\}$	Order

Other Approaches

- Treat Ranking as Classification or Regression

  E.g. Basu, Hirsh, Cohen 1998

- Reduce a ranking problem into a classification problem over pair of examples

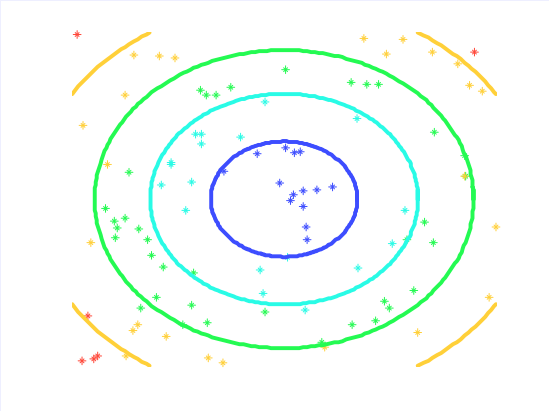
  E.g. Freund, Lyer, Schapire, Singer 1998
  Herbrich, Graepel, Obermayer 2000

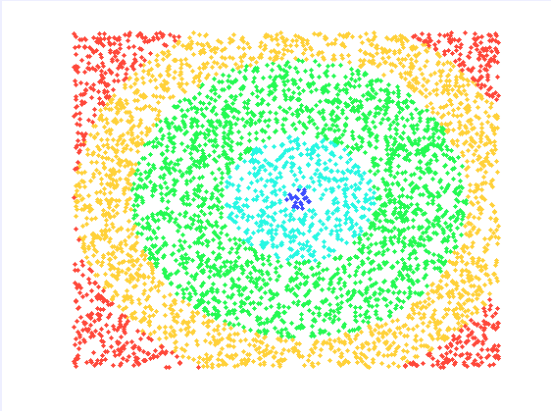
  – Not simple to combine preferences predictions over pairs into a singe consistent ordering
  – No simple adaptation for online settings

Empirical Study

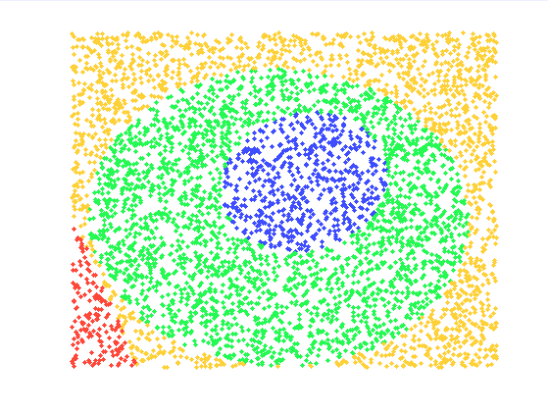
An Illustration

- Five concentric ellipses
- Training set of 50 points
- Three approaches
  - Pranking
  - Classification
  - Regression

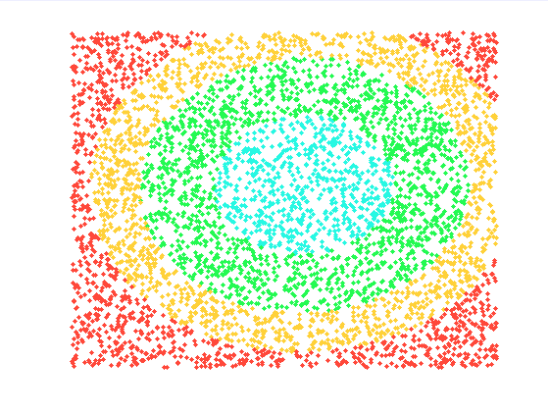




PRank
Ranking



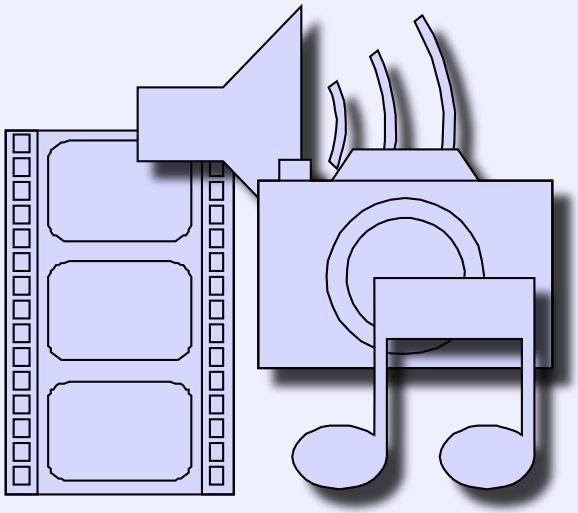
MC-Perceptron
Classification



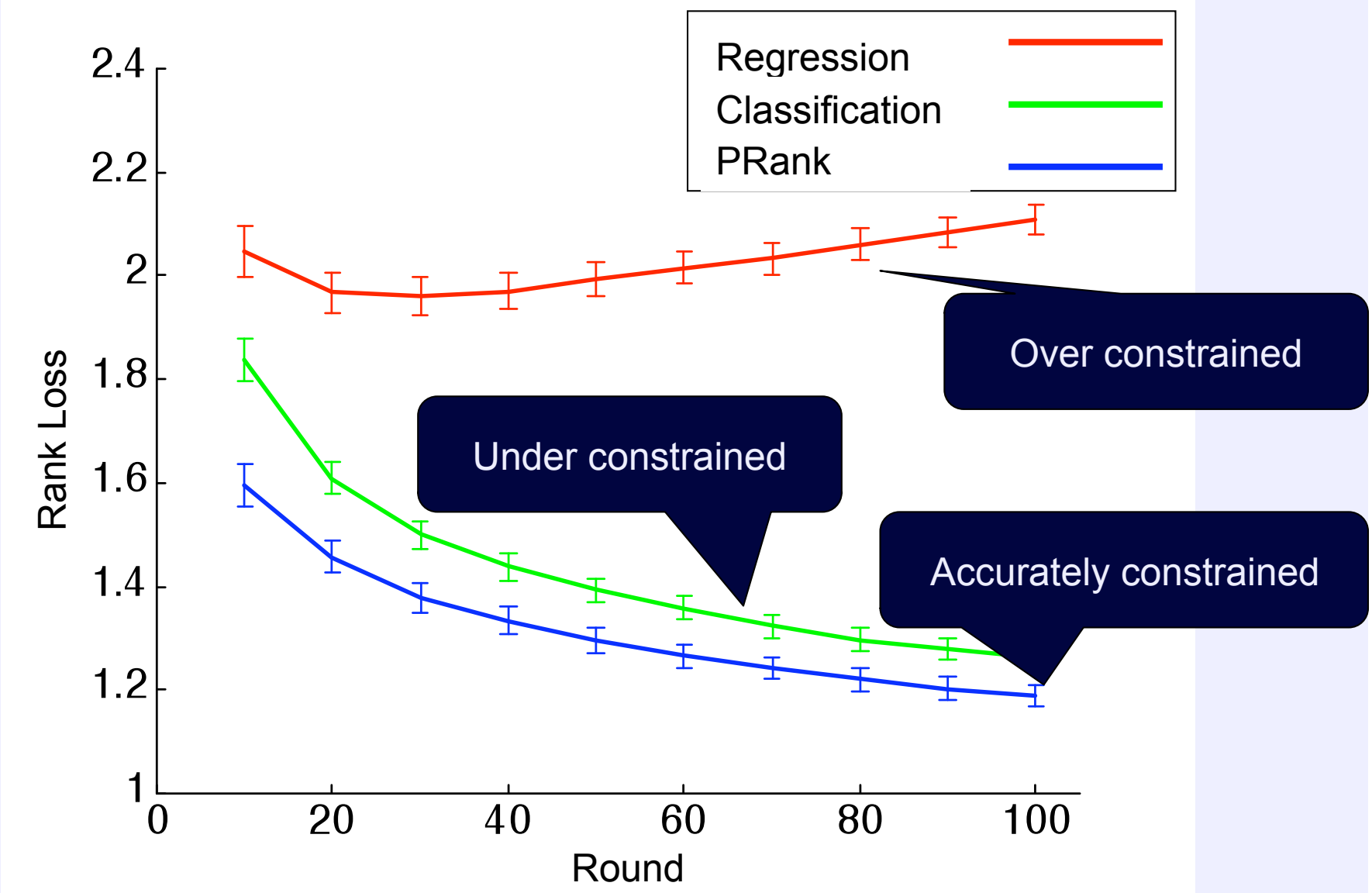
Widrow-Hoff
Regression

Each-Movie database

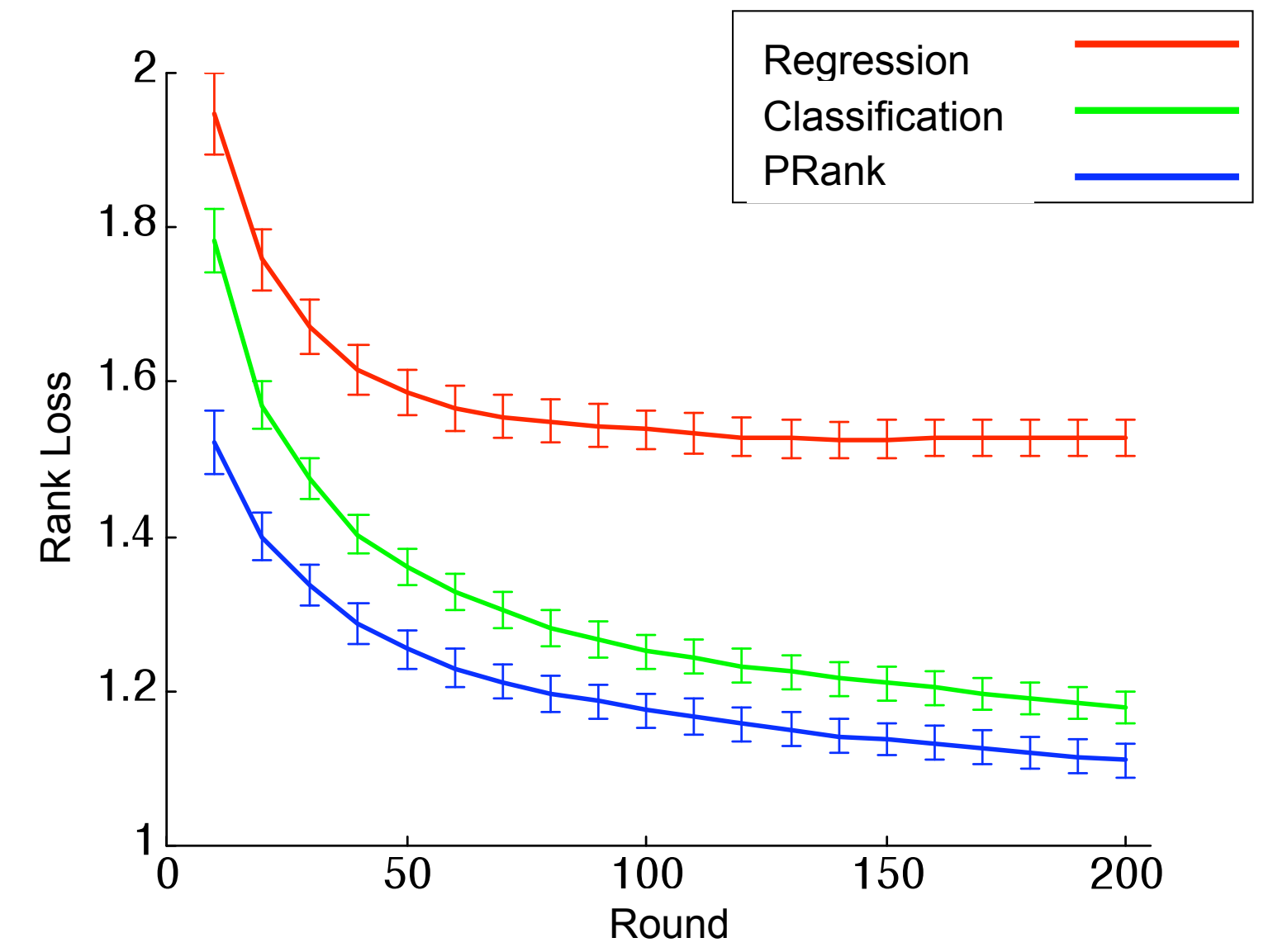
- 74424 registered Users
- 1648 listed Movies
- Users ranking of movies
- 7451 Users saw >100 movies
- 1801 Users saw >200 movies

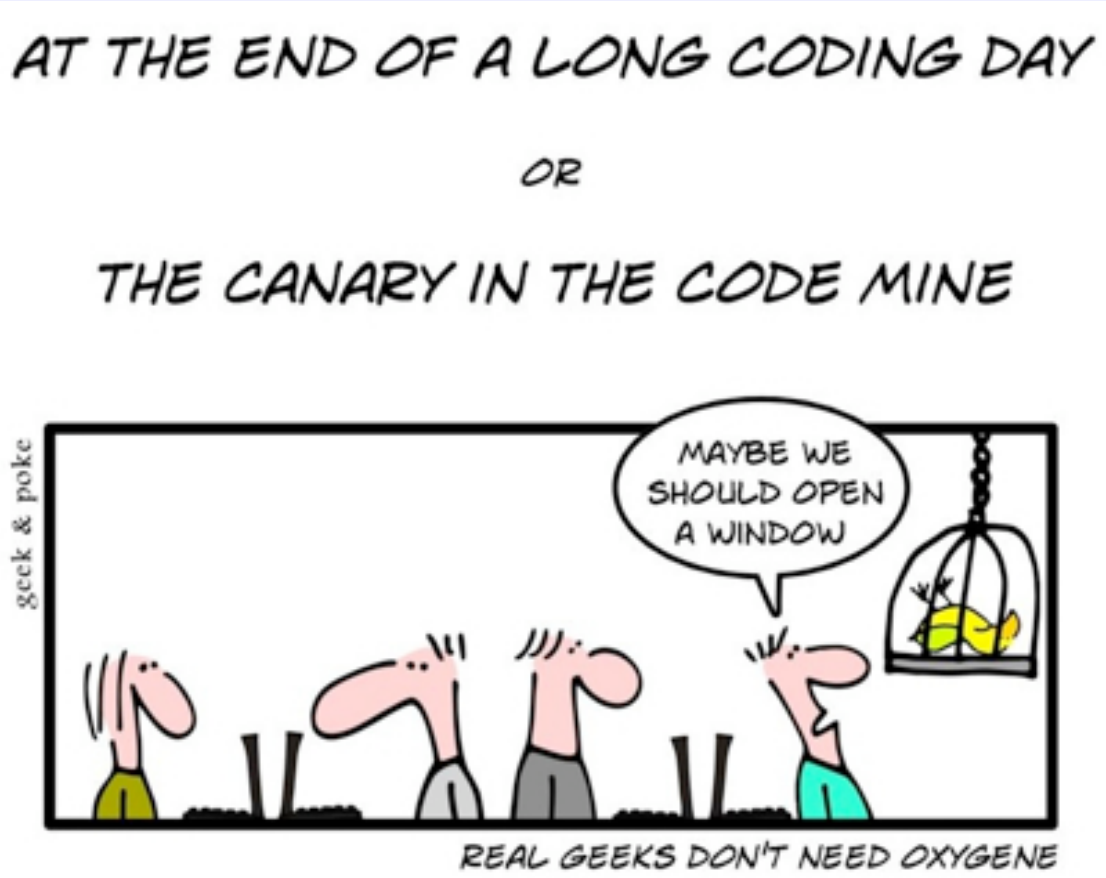


Ranking Loss, 100 Viewers



Ranking Loss, 200 Viewers





Demonstration

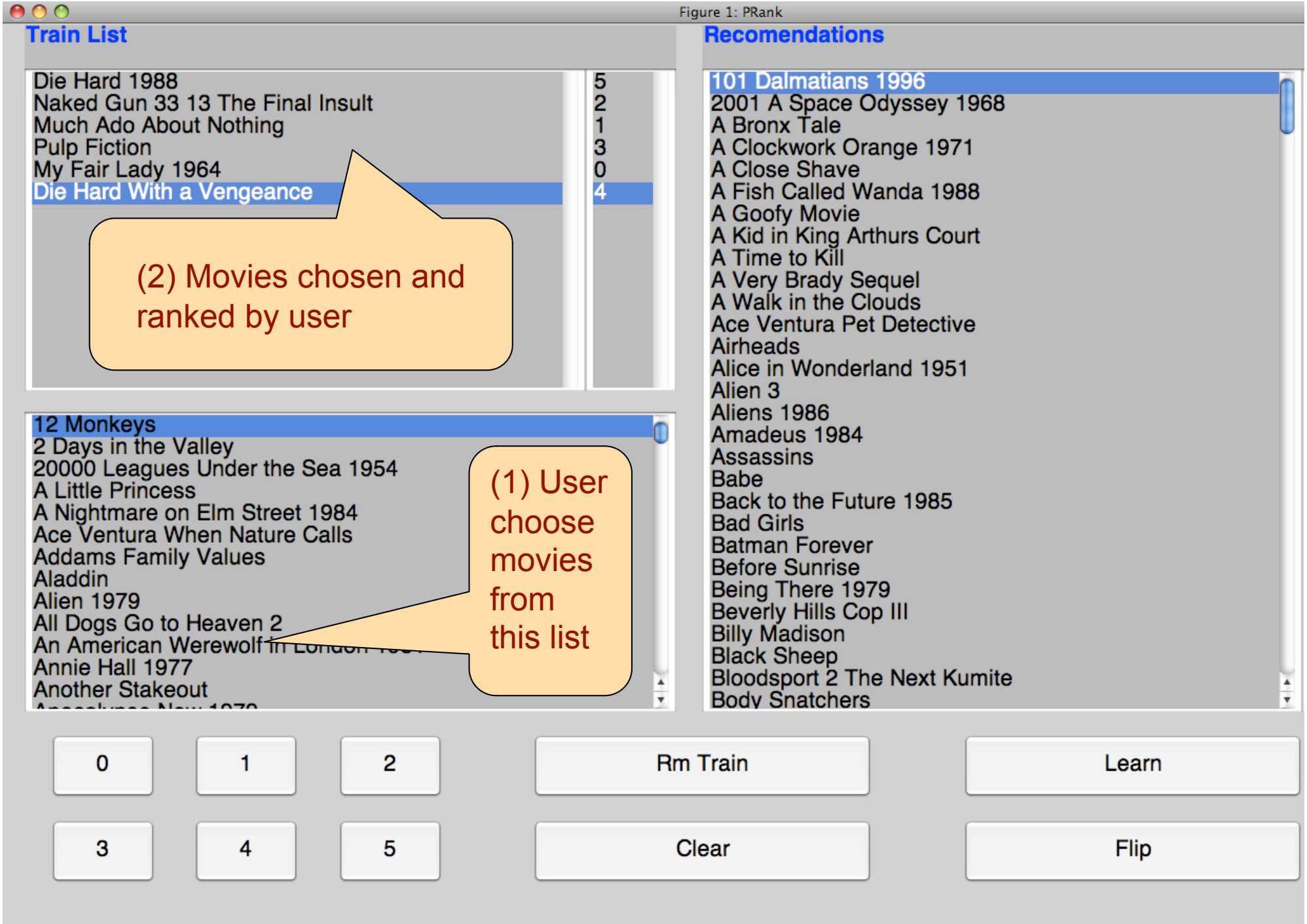


Figure 1: PRank
Train List
Die Hard 1988
Naked Gun 33 13 The Final Insult
Much Ado About Nothing
Pulp Fiction
My Fair Lady 1964
Die Hard With a Vengeance
5
2
1
3
0
4
(2) Movies chosen and ranked by user
12 Monkeys
2 Days in the Valley
20000 Leagues Under the Sea 1954
A Little Princess
A Nightmare on Elm Street 1984
Ace Ventura When Nature Calls
Addams Family Values
Aladdin
Alien 1979
All Dogs Go to Heaven 2
Annie Hall 1977
Another Stakeout
(1) User choose movies from this list
Recomendations
101 Dalmatians 1996
2001 A Space Odyssey 1968
A Bronx Tale
A Clockwork Orange 1971
A Close Shave
A Fish Called Wanda 1988
A Goofy Movie
A Kid in King Arthurs Court
A Time to Kill
A Very Brady Sequel
A Walk in the Clouds
Ace Ventura Pet Detective
Airheads
Alice in Wonderland 1951
Alien 3
Aliens 1986
Amadeus 1984
Assassins
Babe
Back to the Future 1985
Bad Girls
Batman Forever
Before Sunrise
Being There 1979
Beverly Hills Cop III
Billy Madison
Black Sheep
Bloodsport 2 The Next Kumite
Body Snatchers
0
1
2
Rm Train
Learn
3
4
5
Clear
Flip

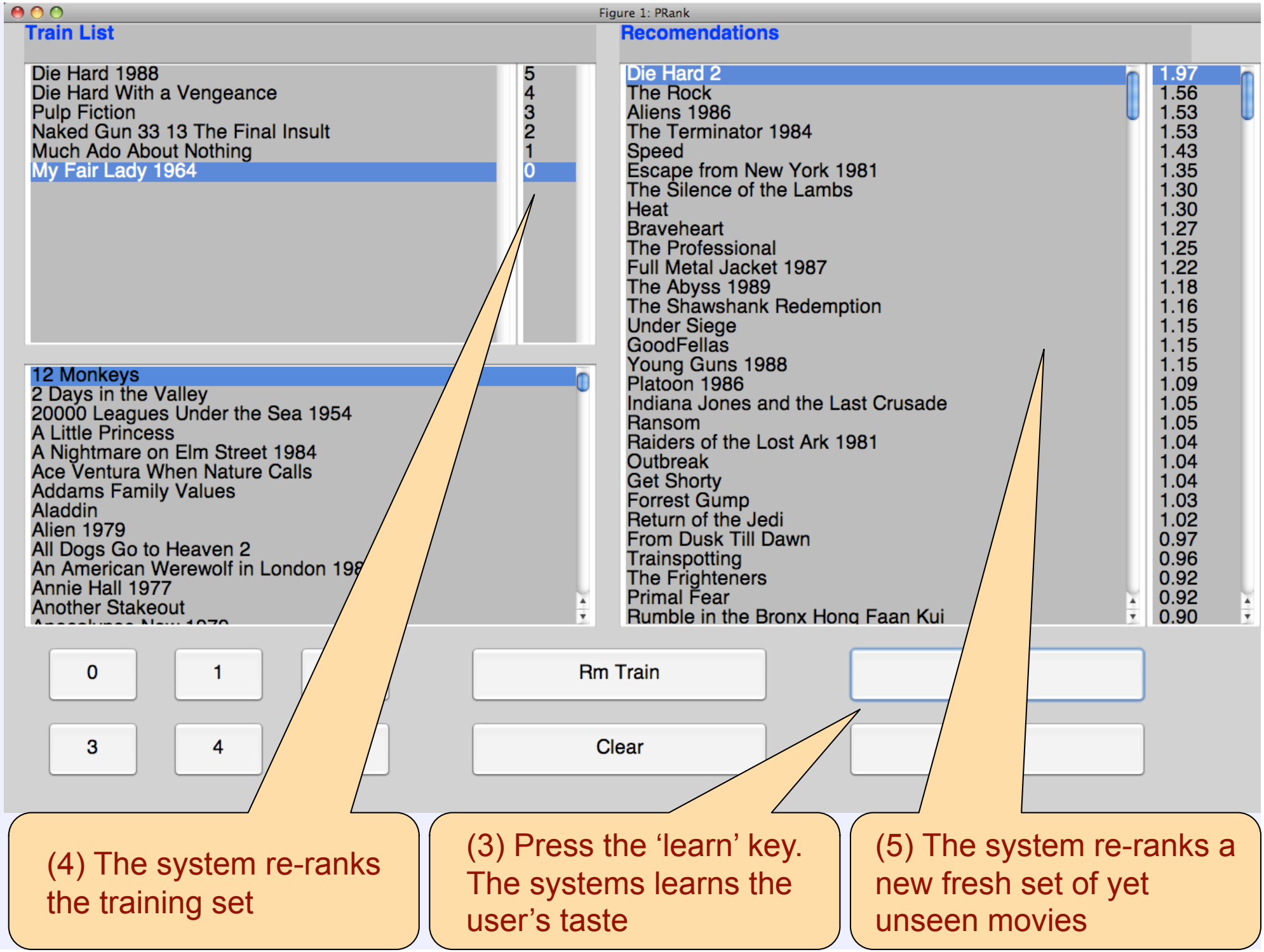


Figure 1: PRank
Train List
Die Hard 1988 5
Die Hard With a Vengeance 4
Pulp Fiction 3
Naked Gun 33 13 The Final Insult 2
Much Ado About Nothing 1
My Fair Lady 1964 0
12 Monkeys
2 Days in the Valley
20000 Leagues Under the Sea 1954
A Little Princess
A Nightmare on Elm Street 1984
Ace Ventura When Nature Calls
Addams Family Values
Aladdin
Alien 1979
All Dogs Go to Heaven 2
An American Werewolf in London 198
Annie Hall 1977
Another Stakeout
Recomendations
Die Hard 2 1.97
The Rock 1.56
Aliens 1986 1.53
The Terminator 1984 1.53
Speed 1.43
Escape from New York 1981 1.35
The Silence of the Lambs 1.30
Heat 1.30
Braveheart 1.27
The Professional 1.25
Full Metal Jacket 1987 1.22
The Abyss 1989 1.18
The Shawshank Redemption 1.16
Under Siege 1.15
GoodFellas 1.15
Young Guns 1988 1.15
Platoon 1986 1.09
Indiana Jones and the Last Crusade 1.05
Ransom 1.05
Raiders of the Lost Ark 1981 1.04
Outbreak 1.04
Get Shorty 1.04
Forrest Gump 1.03
Return of the Jedi 1.02
From Dusk Till Dawn 0.97
Trainspotting 0.96
The Frighteners 0.92
Primal Fear 0.92
Rumble in the Bronx Hong Faan Kui 0.90
0
1
3
4
Rm Train
Clear
(4) The system re-ranks the training set
(3) Press the 'learn' key. The systems learns the user's taste
(5) The system re-ranks a new fresh set of yet unseen movies

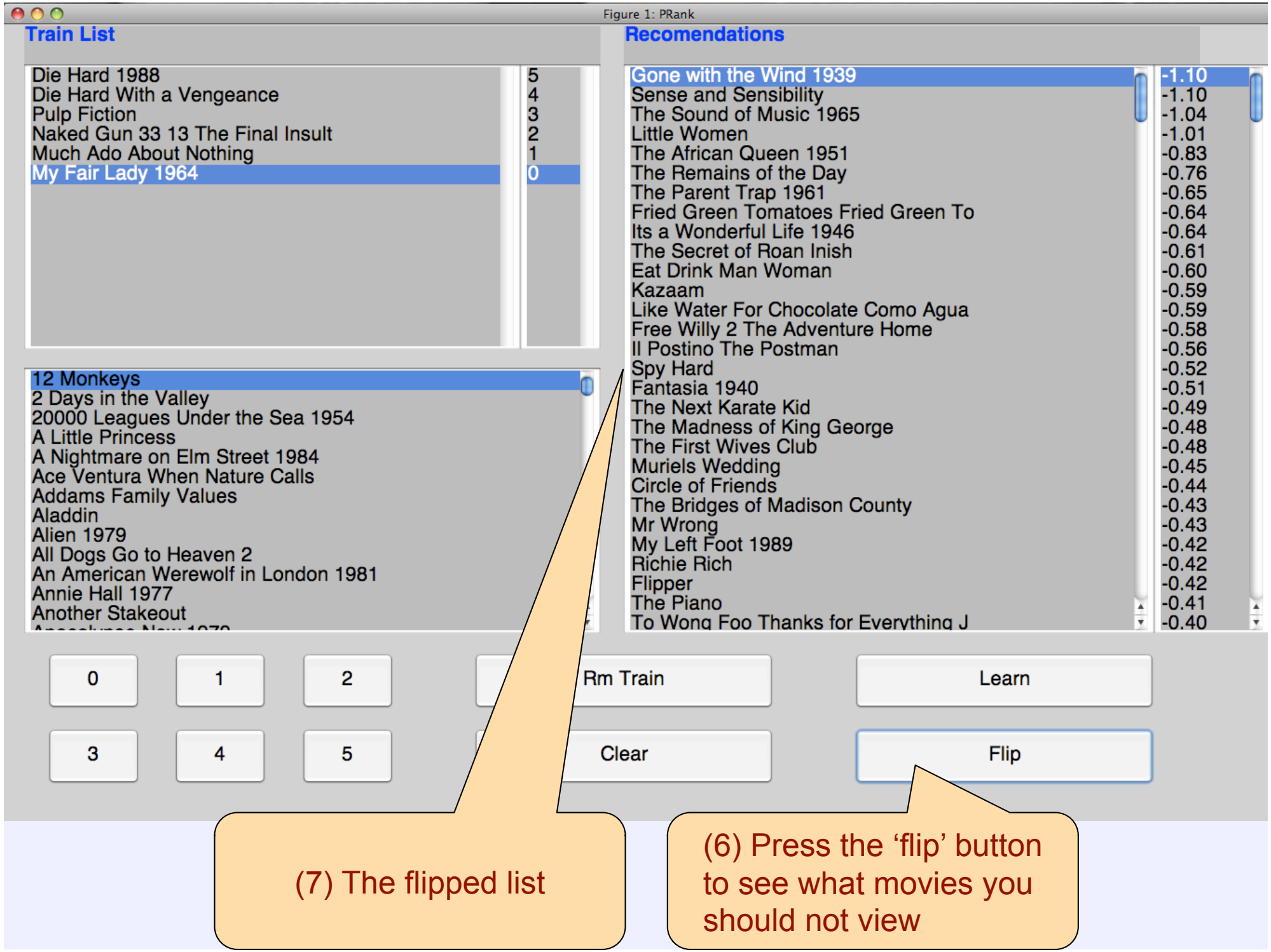


Figure 1: PRank
Train List
Die Hard 1988 5
Die Hard With a Vengeance 4
Pulp Fiction 3
Naked Gun 33 13 The Final Insult 2
Much Ado About Nothing 1
My Fair Lady 1964 0
12 Monkeys
2 Days in the Valley
20000 Leagues Under the Sea 1954
A Little Princess
A Nightmare on Elm Street 1984
Ace Ventura When Nature Calls
Addams Family Values
Aladdin
Alien 1979
All Dogs Go to Heaven 2
An American Werewolf in London 1981
Annie Hall 1977
Another Stakeout
Recomendations
Gone with the Wind 1939 -1.10
Sense and Sensibility -1.10
The Sound of Music 1965 -1.04
Little Women -1.01
The African Queen 1951 -0.83
The Remains of the Day -0.76
The Parent Trap 1961 -0.65
Fried Green Tomatoes Fried Green To -0.64
Its a Wonderful Life 1946 -0.64
The Secret of Roan Inish -0.61
Eat Drink Man Woman -0.60
Kazaam -0.59
Like Water For Chocolate Como Agua -0.59
Free Willy 2 The Adventure Home -0.58
Il Postino The Postman -0.56
Spy Hard -0.52
Fantasia 1940 -0.51
The Next Karate Kid -0.49
The Madness of King George -0.48
The First Wives Club -0.48
Muriels Wedding -0.45
Circle of Friends -0.44
The Bridges of Madison County -0.43
Mr Wrong -0.43
My Left Foot 1989 -0.42
Richie Rich -0.42
Flipper -0.42
The Piano -0.41
To Wong Foo Thanks for Everything J -0.40
0
1
2
Rm Train
Learn
3
4
5
Clear
Flip
(7) The flipped list
(6) Press the 'flip' button to see what movies you should not view

- Many alternatives to formulate ranking
- Choose one that models best your problem
- Exploit and Incorporate structure
- Specifically:
  - Online algorithm for ranking problems via projections and conservative update of the projection's direction and the threshold values
  - Experiments on a synthetic dataset and on Each-Movie data set indicate that the PRank algorithm performs better then algorithms for classification and regression