Learning Structural SVMs with Latent Variables

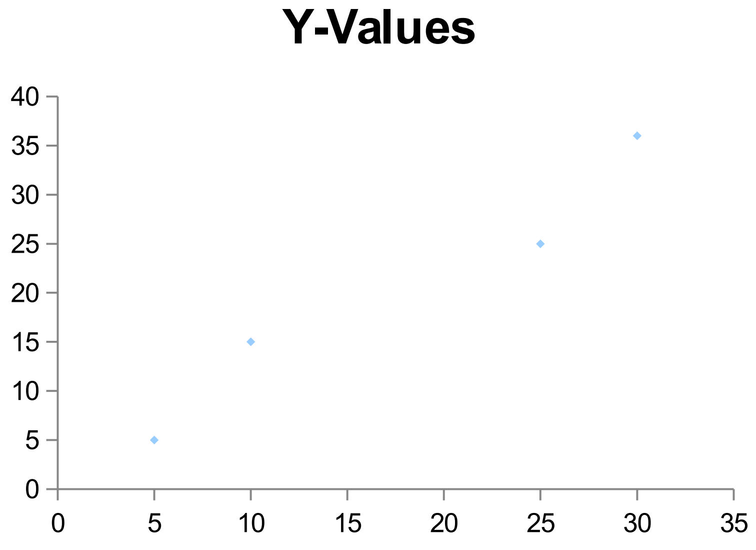
Presented By-

- Subhabrata Debnath(Roll-13111063)
- Anjan Banerjee(Roll-13111008)

Basics

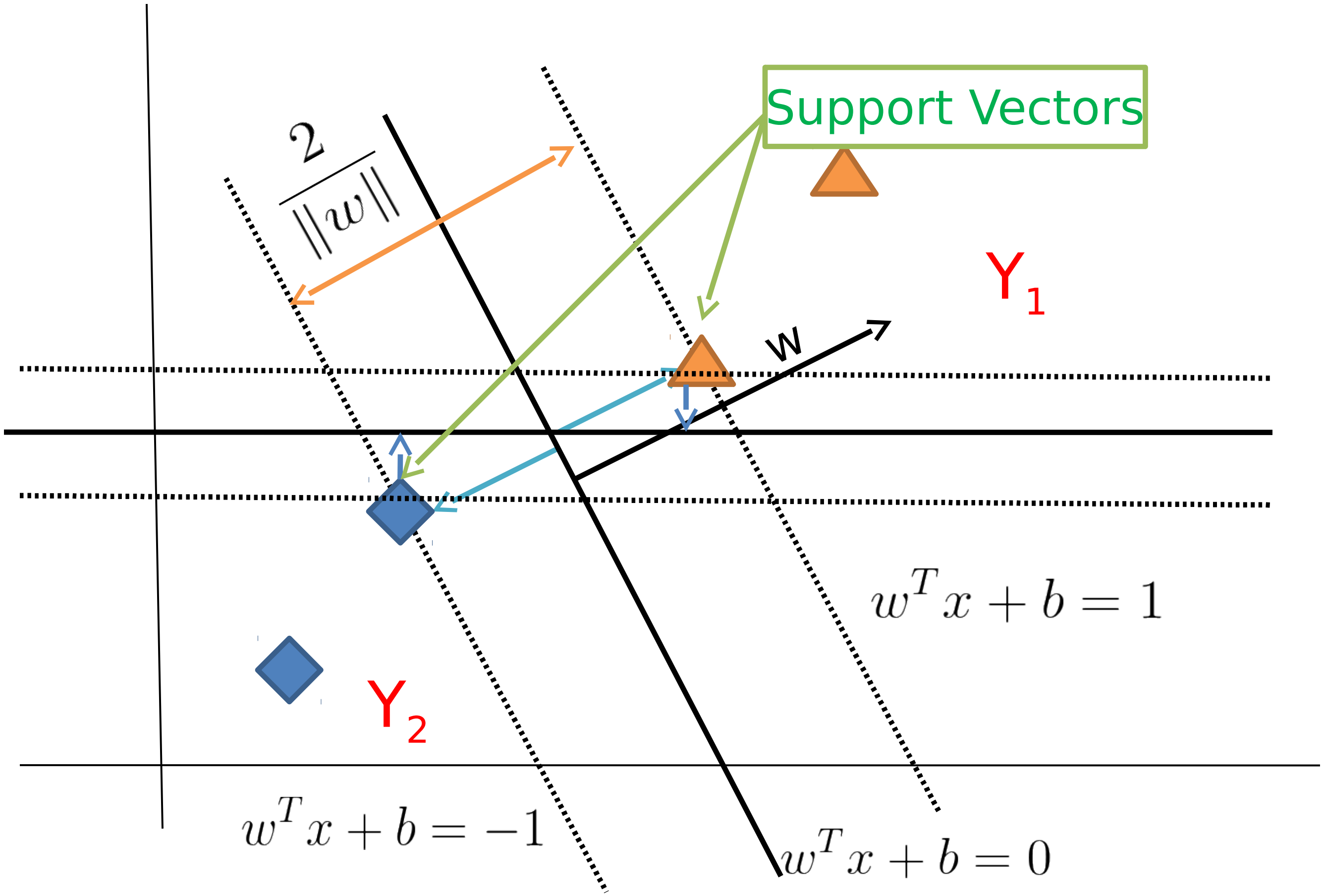
Machine Learning

Blood Pressure	Sugar	Hyper Tension
10	15	No
5	5	No
25	25	Yes
30	36	Yes



Basics

Machine Learning



Objective Function of SVM

$$\max_{w} \frac{2}{\|w\|}$$

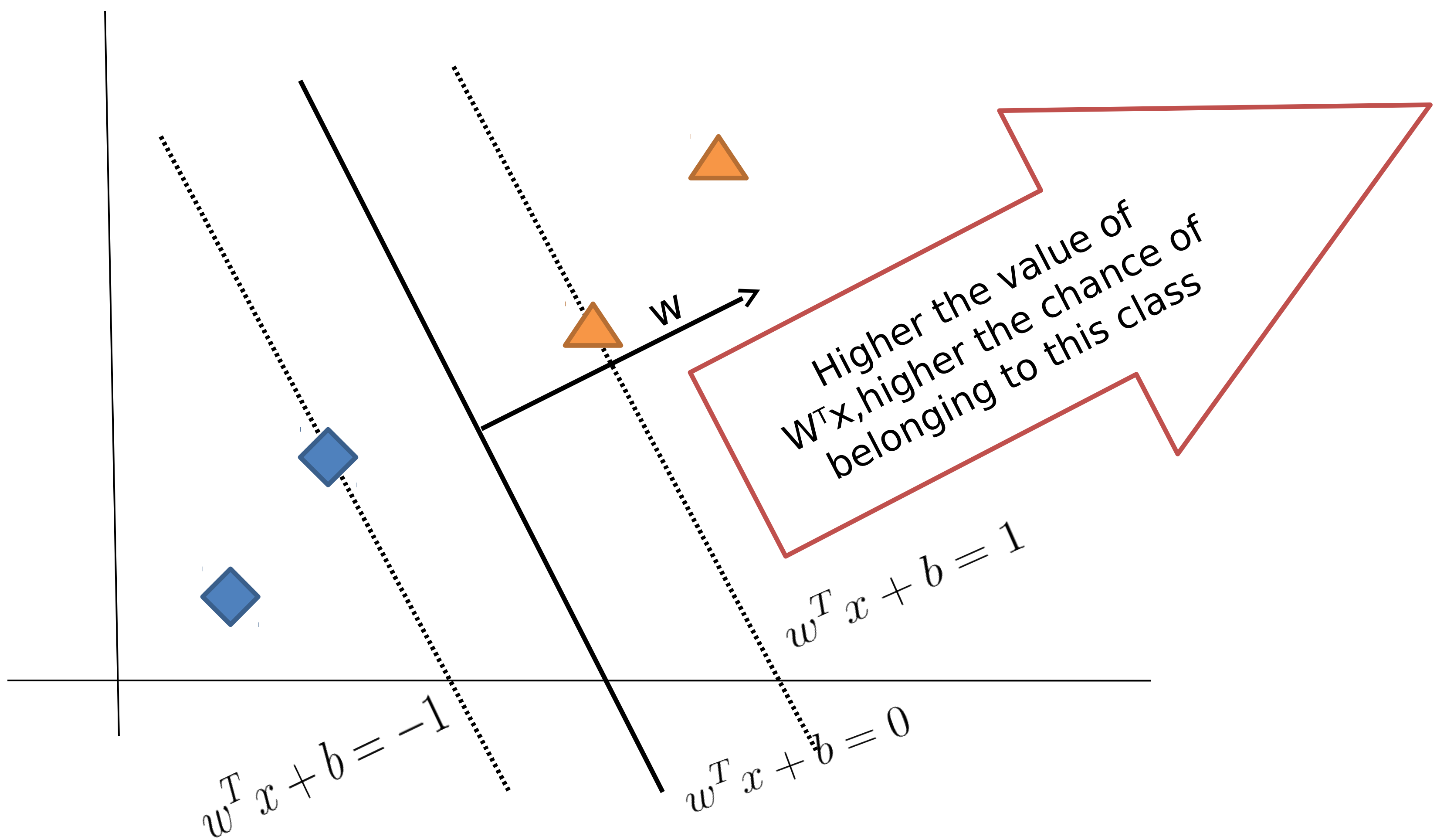
s.t.
$$\begin{aligned} &if \;\; y_i = Y_1, \quad w^T x_i + b \geq 1 \\ &if \;\; y_i = Y_2, \quad w^T x_i + b \leq -1 \end{aligned}$$

Objective Function of SVM

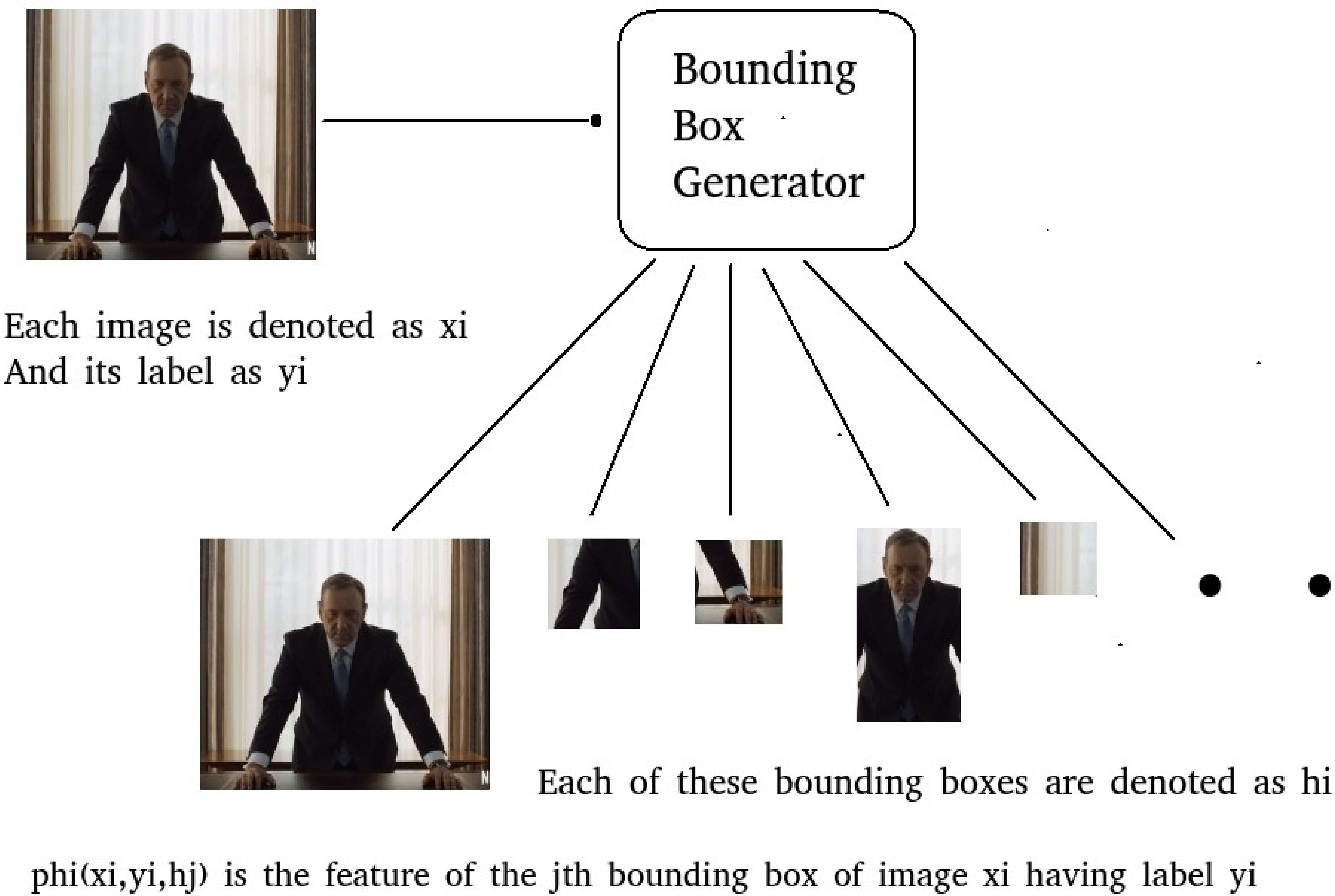
$$\min_{w} \frac{\|w\|^2}{2}$$

s.t. $$\begin{aligned} &if \;\; y_i = Y_1, \quad w^T x_i + b \geq 1 \\ &if \;\; y_i = Y_2, \quad w^T x_i + b \leq -1 \end{aligned}$$

Score



w
Higher the value of $W^T x$, higher the chance of belonging to this class
$w^T x + b = 1$
$w^T x + b = -1$
$w^T x + b = 0$


Bounding Box Generator
Each image is denoted as xi
And its label as yi
Each of these bounding boxes are denoted as hi
phi(xi,yi,hj) is the feature of the jth bounding box of image xi having label yi

Latent Structured SVM

$$\min_{\boldsymbol{w},\xi} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$\forall i, \xi_i \geq \max_{(\hat{y},\hat{h})\in\mathcal{Y}\times\mathcal{H}} [\boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,\hat{y},\hat{h}) + \Delta(y_i,\hat{y},\hat{h})] - \max_{h\in\mathcal{H}} \boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,y_i,h)$$

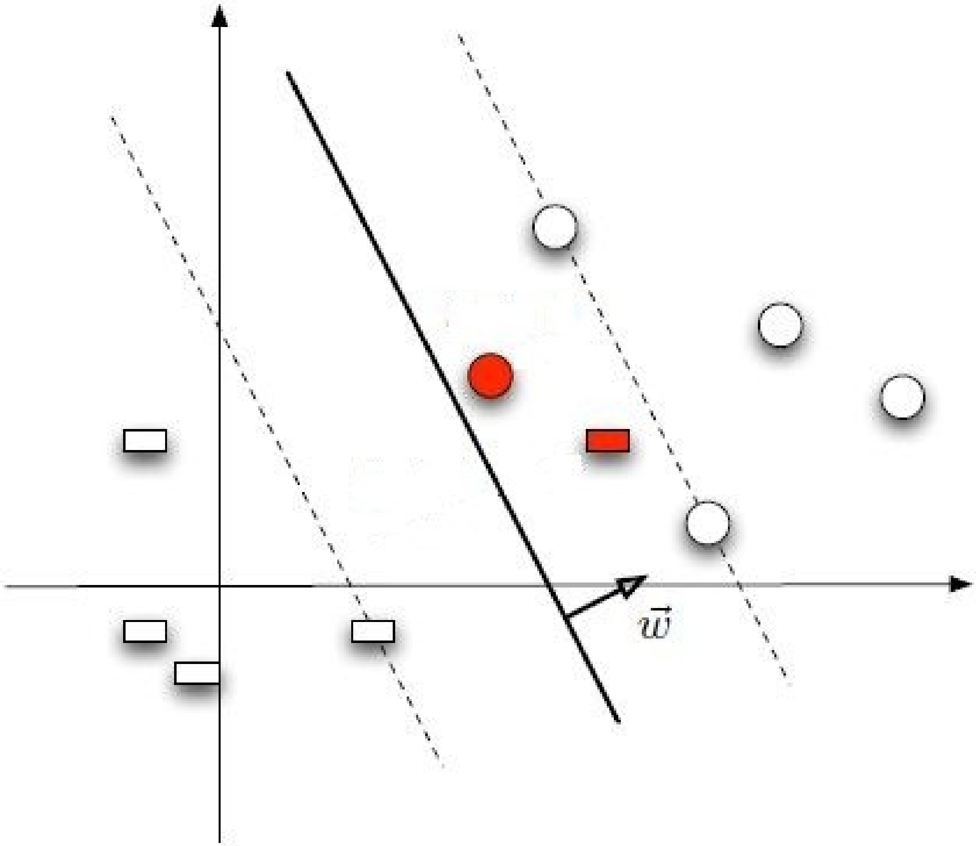
- **<u>Final Objective Function:</u>**

$$\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\left(\max_{(\hat{y},\hat{h})\in\mathcal{Y}\times\mathcal{H}} [\boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,\hat{y},\hat{h}) + \Delta(y_i,\hat{y},\hat{h})] - \max_{h\in\mathcal{H}} \boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,y_i,h)\right)$$

$$= \underbrace{\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\max_{(\hat{y},\hat{h})\in\mathcal{Y}\times\mathcal{H}} [\boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,\hat{y},\hat{h}) + \Delta(y_i,\hat{y},\hat{h})]}_{\text{convex}} \underbrace{-C\sum_{i=1}^{n}\max_{h\in\mathcal{H}} \boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,y_i,h)}_{\text{concave}}$$

Non-Convex Objective Function $\Rightarrow$ Can be solved by CCCP

Soft-Margin SVM



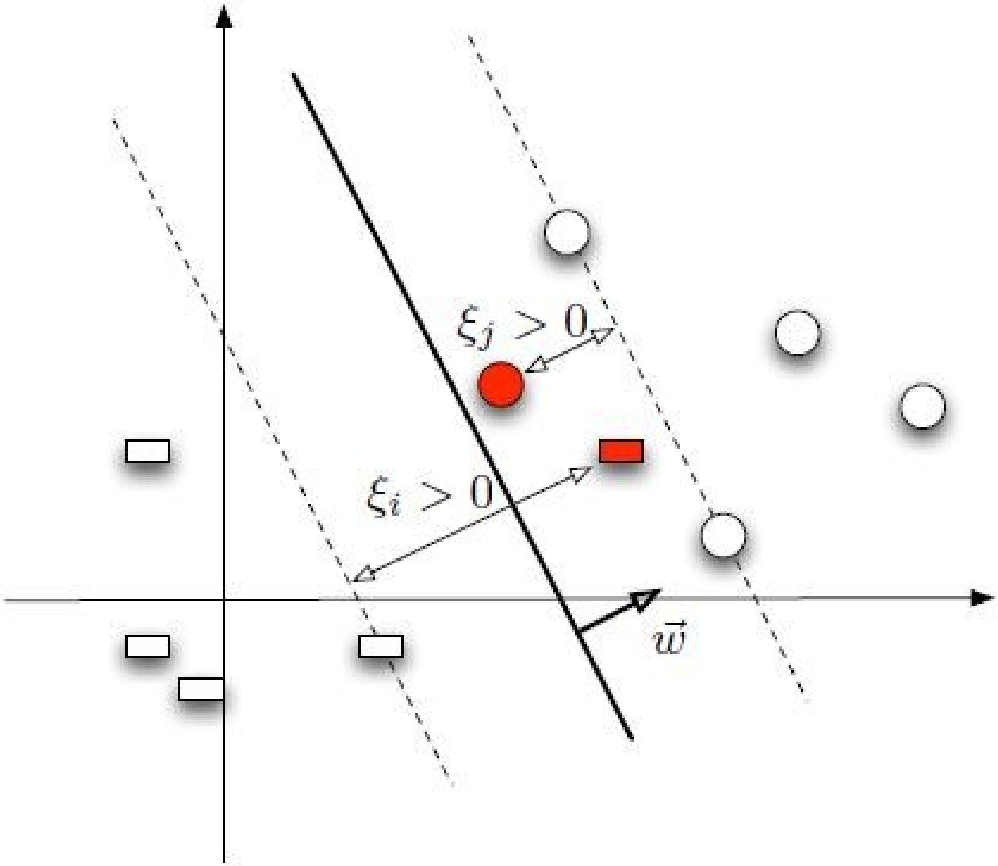
Soft-Margin SVM

$$\min_{w,\xi} \frac{\|w\|^2}{2} + C \sum_i \xi_i$$

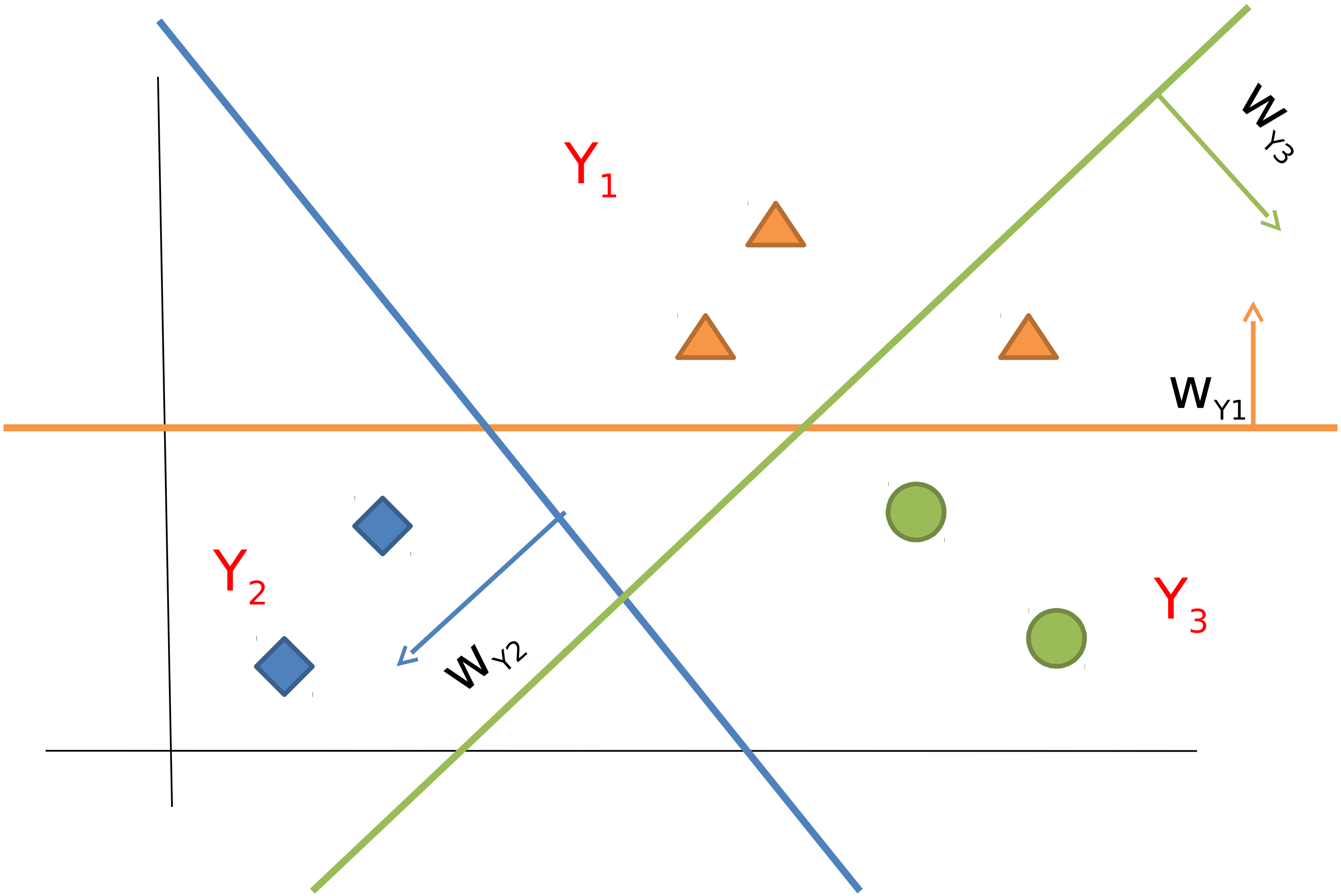
s.
t. $$\begin{aligned} &if \quad y_i = Y_1, \quad w^T x_i + b \geq 1 - \xi_i \\ &if \quad y_i = Y_2, \quad w^T x_i + b \leq -(1 - \xi_i) \end{aligned}$$

$$\forall i,\ \xi_i \geq 0$$

Soft-Margin SVM



Multiclass SVM



Predicted Class: $\max_{Y_i} \; w_{Y_i} \cdot x$

Multi-Class SVM

$$\min_{w,\xi} \frac{\|w\|^2}{2} + C \sum_i \xi_i$$

s.t.

$$w_{Y_i} \cdot x_i - w_{\hat{Y} \neq Y_i} \cdot x_i \geqslant 1 - \xi_i, \forall \ (x_i, Y_i)$$

$$\forall i, \hat{y} \neq y_i, \xi_i \geqslant 0$$

here, $w = \begin{pmatrix} w_{Y_1} \\ w_{Y_2} \\ w_{Y_3} \end{pmatrix}$

Multi-class SVM

- What if we don’t want the same amount of margin for all the classes?

- E.g.: Given age, sex of an user and the movie genre, predict the rating(1-5) that the user will give.

- Highly Incorrect Class and Lesser Incorrect Class

Actual Rating	Predicted Rating	Loss
5	4	Less
5	1	High

Multi-Class SVM

$$\min_{w,\xi} \frac{\|w\|^2}{2} + C \sum_i \xi_i$$

s.t.

$$w_{Y_i} \cdot x_i - w_{\hat{Y} \neq Y_i} \cdot x_i \geqslant \Delta(Y_i, \hat{Y}) - \xi_i, \forall\ (x_i, Y_i)$$

$$\forall i, \hat{y} \neq y_i, \xi_i \geqslant 0$$

Multi-Class SVM

$$\min_{w,\xi} \frac{\|w\|^2}{2} + C\sum_i \xi_i$$

s.t.

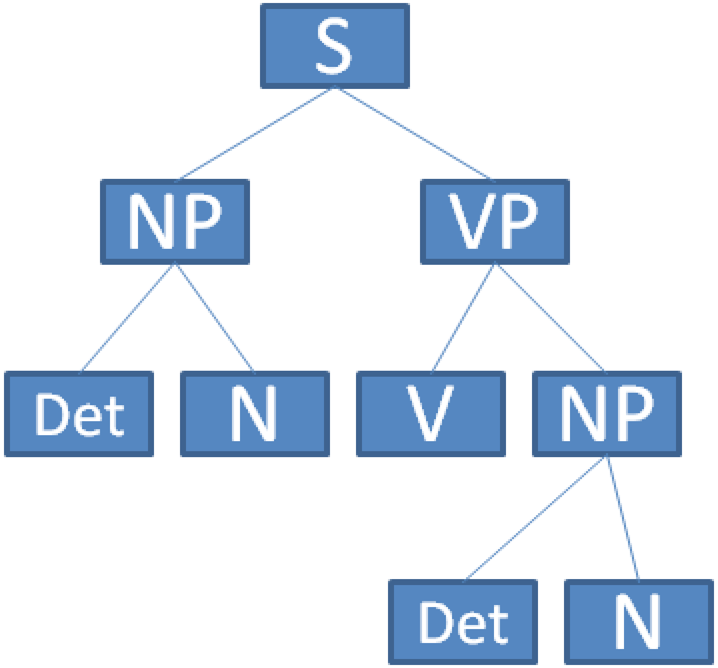
$$\boldsymbol{w} \cdot \boldsymbol{\Phi}(x_i, y_i) - \boldsymbol{w} \cdot \boldsymbol{\Phi}(x_i, \bar{y}) \geq \Delta(y_i, \bar{y}) - \xi_i$$

$$\forall i, \hat{y} \neq y_i, \xi_i \geqslant 0$$

$$w = \begin{pmatrix} w_{Y_1} \\ w_{Y_2} \\ \cdot \\ \cdot \\ \cdot \\ w_{Y_{k-1}} \\ w_{Y_k} \end{pmatrix} \quad \Phi(x, y) = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ x \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

Structured SVM

The dog chased the cat.



Input X

Output Y

Structured SVM

$$\min_{\vec{w},\vec{\xi}} \frac{1}{2}||\vec{w}||^2 + C\sum_{i=1}^{n} \xi_i$$

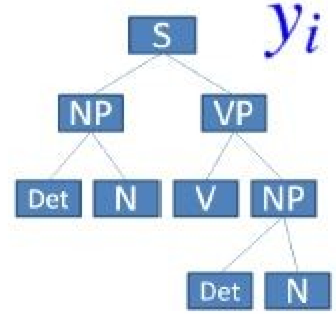
$s.t.$ for $1 \leq i \leq n$, for all output structures $\hat{y} \in \mathcal{Y}$,

$$\vec{w} \cdot \Phi(x_i, y_i) - \vec{w} \cdot \Phi(x_i, \hat{y}) \geq \Delta(y_i, \hat{y}) - \xi_i$$

correct parse tree

x_i y_i

$\vec{w} \cdot \Phi($ **The dog chased the cat.** ,



$)$

score of correct parse tree

Structured SVM

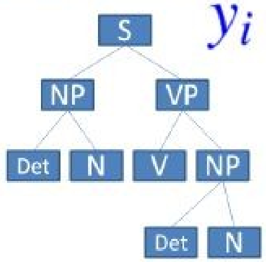
$$\min_{\vec{w},\vec{\xi}} \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$

$s.t.$ for $1 \le i \le n$, for all output structures $\hat{y} \in \mathcal{Y}$,

$$\vec{w}\cdot\Phi(x_i, y_i) - \vec{w}\cdot\Phi(x_i, \hat{y}) \ge \Delta(y_i, \hat{y}) - \xi_i$$

correct parse tree

$\vec{w}\cdot\Phi($ x_i The dog chased the cat. , y_i

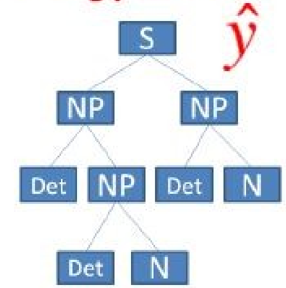


$)$

score of correct parse tree

wrong parse tree

$\vec{w}\cdot\Phi($ x_i The dog chased the cat. , $\hat{y}$



$)$

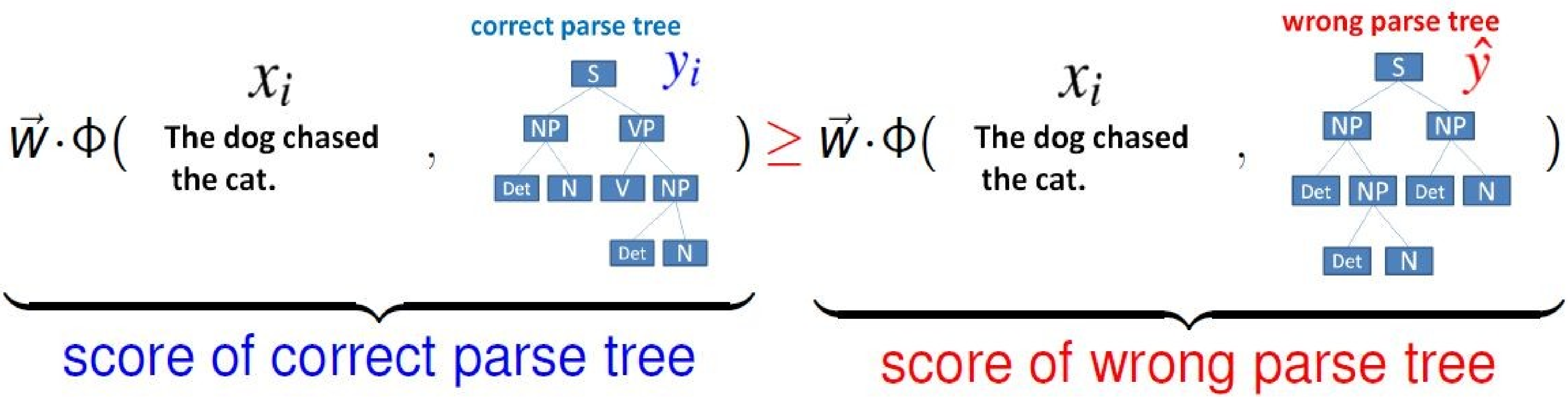
score of wrong parse tree

Structured SVM

$$\min_{\vec{w},\vec{\xi}} \frac{1}{2}||\vec{w}||^2 + C\sum_{i=1}^{n} \xi_i$$

$s.t.$ for $1 \leq i \leq n$, for all output structures $\hat{y} \in \mathcal{Y}$,

$$\vec{w} \cdot \Phi(x_i, y_i) - \vec{w} \cdot \Phi(x_i, \hat{y}) \geq \Delta(y_i, \hat{y}) - \xi_i$$



Structured SVM

$$\min_{w,\xi} \frac{\|w\|^2}{2} + C\sum_i \xi_i$$

$$\xi_i \geqslant \Delta(Y_i, \hat{Y}) - [w \cdot \Phi(x_i, Y_i) - w \cdot \Phi(x_i, \hat{Y})]$$

$$\forall i, \hat{y} \neq y_i, \xi_i \geqslant 0$$

$$\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n} \max_{\bar{y}\in\mathcal{Y}}[\Delta(y_i, \bar{y}) - \boldsymbol{w} \cdot \boldsymbol{\Phi}(x_i, y_i) + \boldsymbol{w} \cdot \boldsymbol{\Phi}(x_i, \bar{y})]$$

Structured SVM

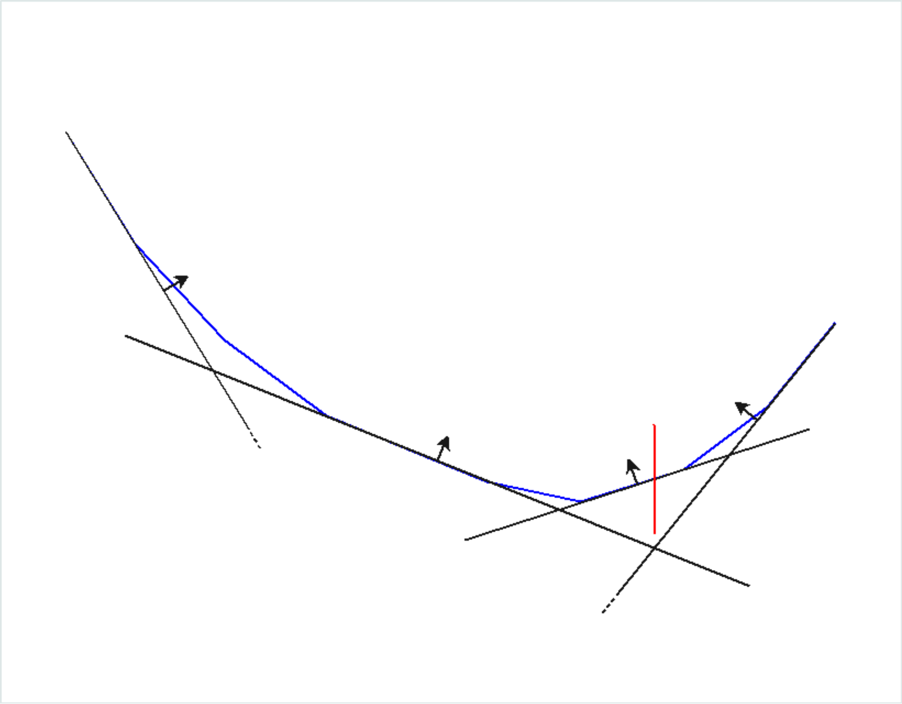
$$\underbrace{\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\max_{\bar{y}\in\mathcal{Y}}[\Delta(y_i,\bar{y}) - w\cdot\Phi(x_i,y_i) + w\cdot\Phi(x_i,\bar{y})]}_{convex}$$

- uld have been solved using any convex solver

- e only problem is the number of classes, hence the number of constraints are exponentially large.

- g. Number of possible parse trees for a given sentence is exponential in the number of words.

Cutting Plane Method



- lowever, this method gives a solution of the given convex optimization problem with precision ε.

$$\text{Our Convex Objective Function} \xrightarrow{\text{Cutting Plane Method}} W^*, \xi_i$$

Latent Information

- Hidden Information present in the training set that can improve our learning
- Let us denote these hidden/latent information as $\mathbf{h_i}$.

x_i

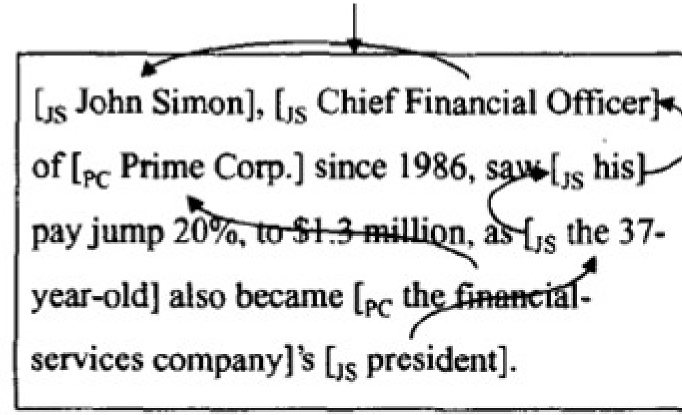
y_i h_i

(given/observed)

(hidden/unobserved)

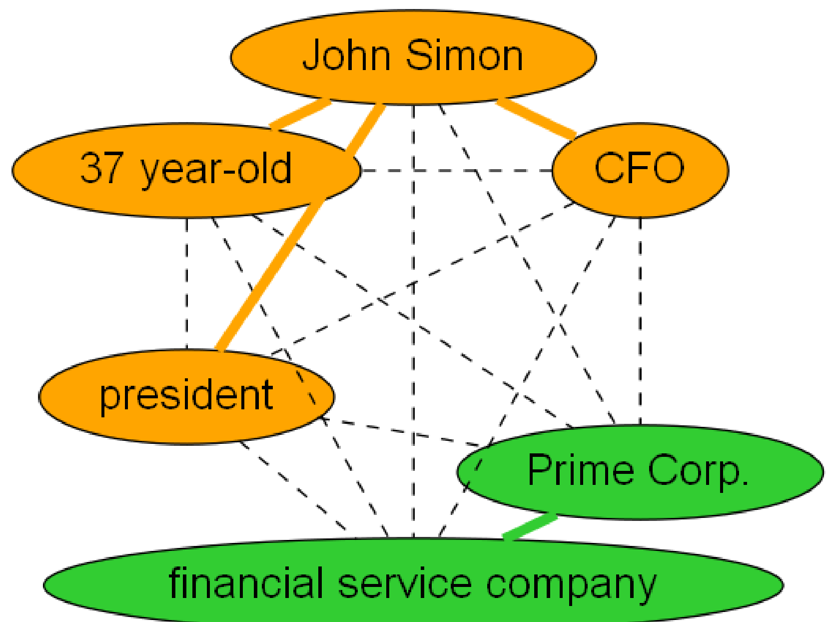
Latent Information

- **oun Phrase Coreference Proble**



- **nput x:**
Noun Phrases with edge features

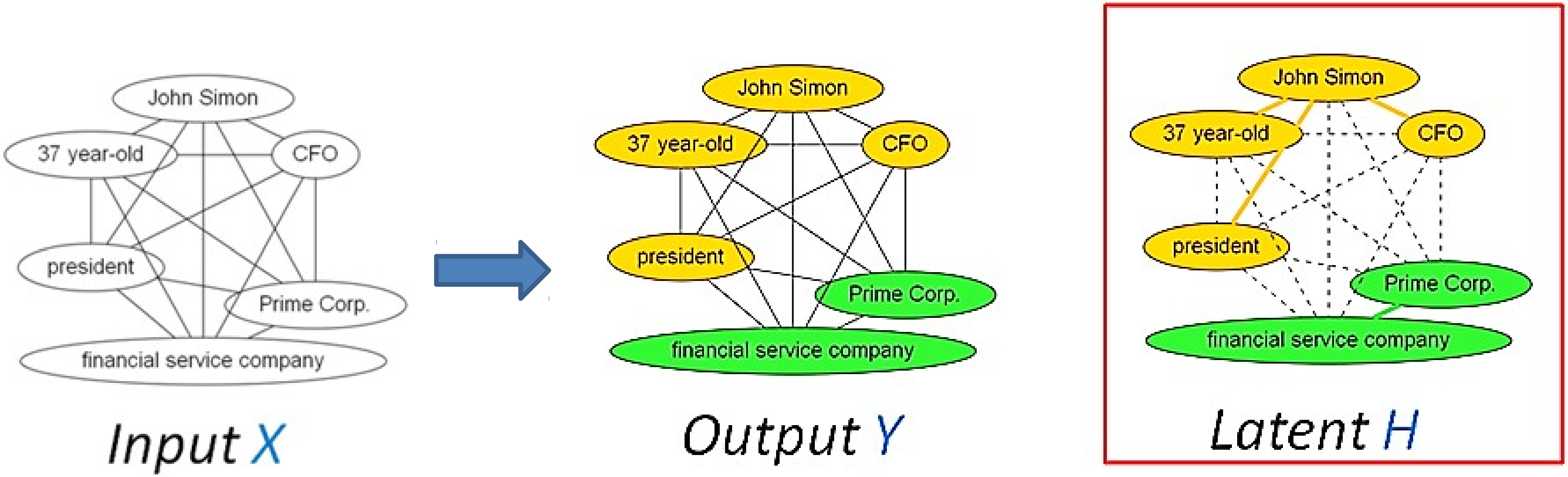
- **abels y:**
Clusters Of Noun Phrases



- **Latent Variable h:**
'Strong' links as trees

Latent Information

- **oun Phrase Coreference Problem:**

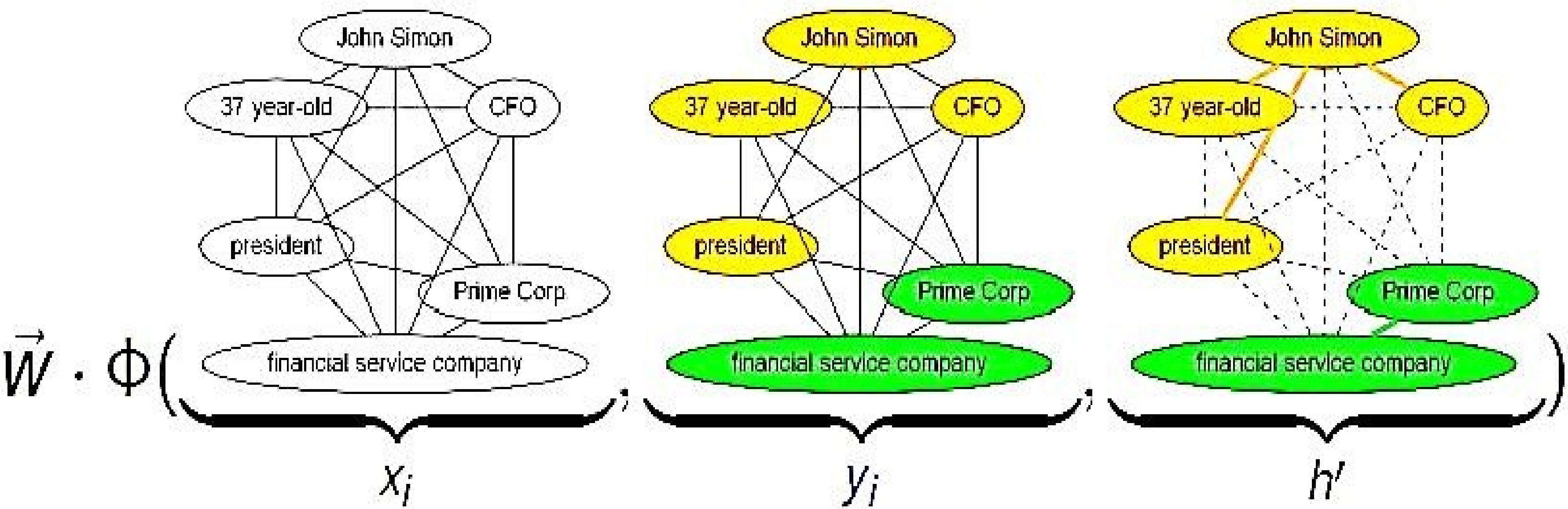


Latent Structured SVM

- **Objective function:**

$$\min_{\vec{w},\vec{\xi}} \frac{1}{2}||\vec{w}||^2 + C\sum_{i=1}^{n}\xi_i \;\; s.t. \text{ for } 1 \le i \le n, \text{ for all outputs } \hat{y} \in \mathcal{Y},$$

$$\max_{h\in\mathcal{H}} \vec{w}\cdot\Phi(x_i, y_i, h) - \max_{\hat{h}\in\mathcal{H}} \vec{w}\cdot\Phi(x_i,\hat{y},\hat{h}) \ge \Delta(y_i,\hat{y},\hat{h}) - \xi_i$$

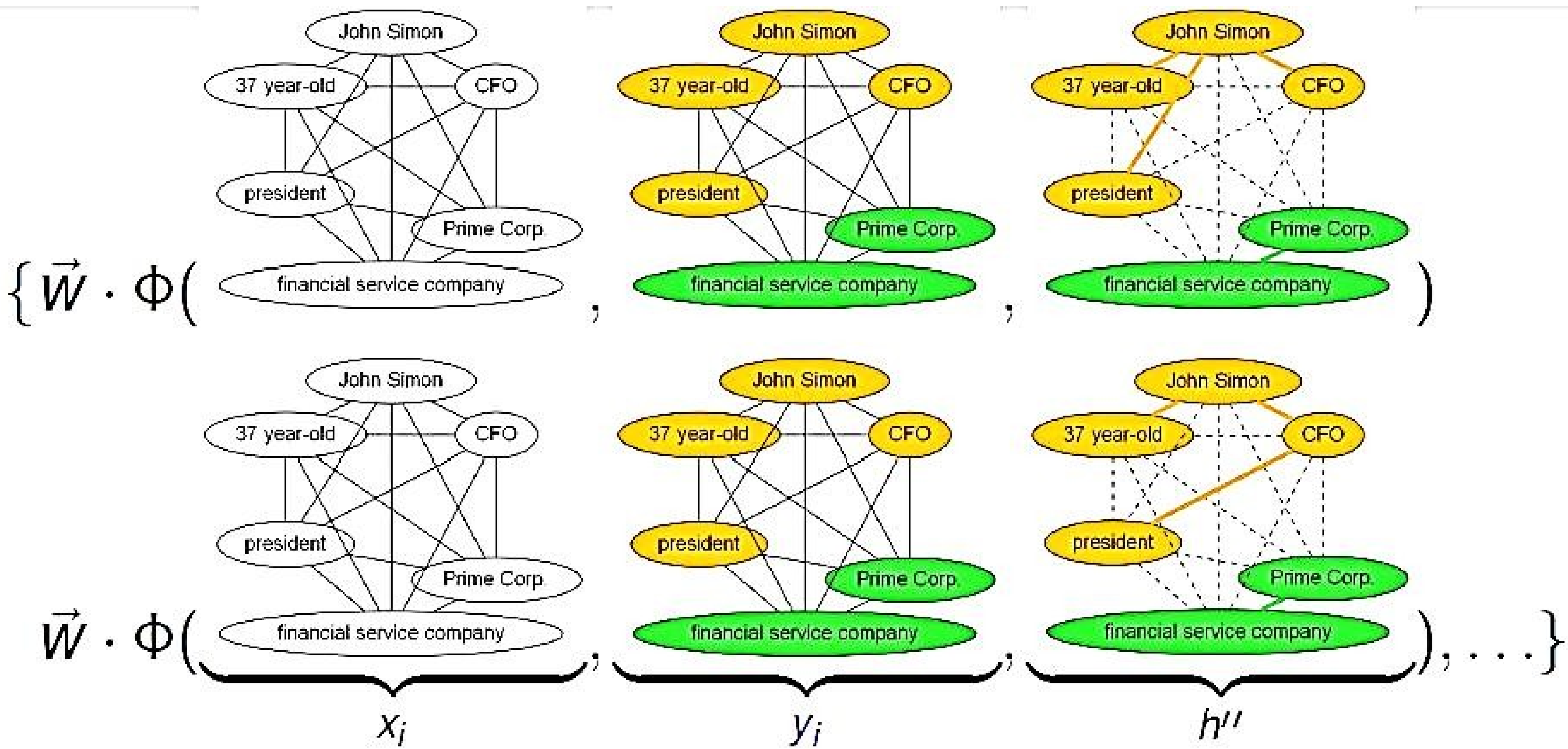


Latent Structured SVM

- **bjective function:**

$$\min_{\vec{w},\vec{\xi}} \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{n}\xi_i \quad s.t. \text{ for } 1 \le i \le n, \text{ for all outputs } \hat{y} \in \mathcal{Y},$$

$$\max_{h\in\mathcal{H}} \vec{w}\cdot\Phi(x_i, y_i, h) - \max_{\hat{h}\in\mathcal{H}} \vec{w}\cdot\Phi(x_i, \hat{y}, \hat{h}) \ge \Delta(y_i, \hat{y}, \hat{h}) - \xi_i$$

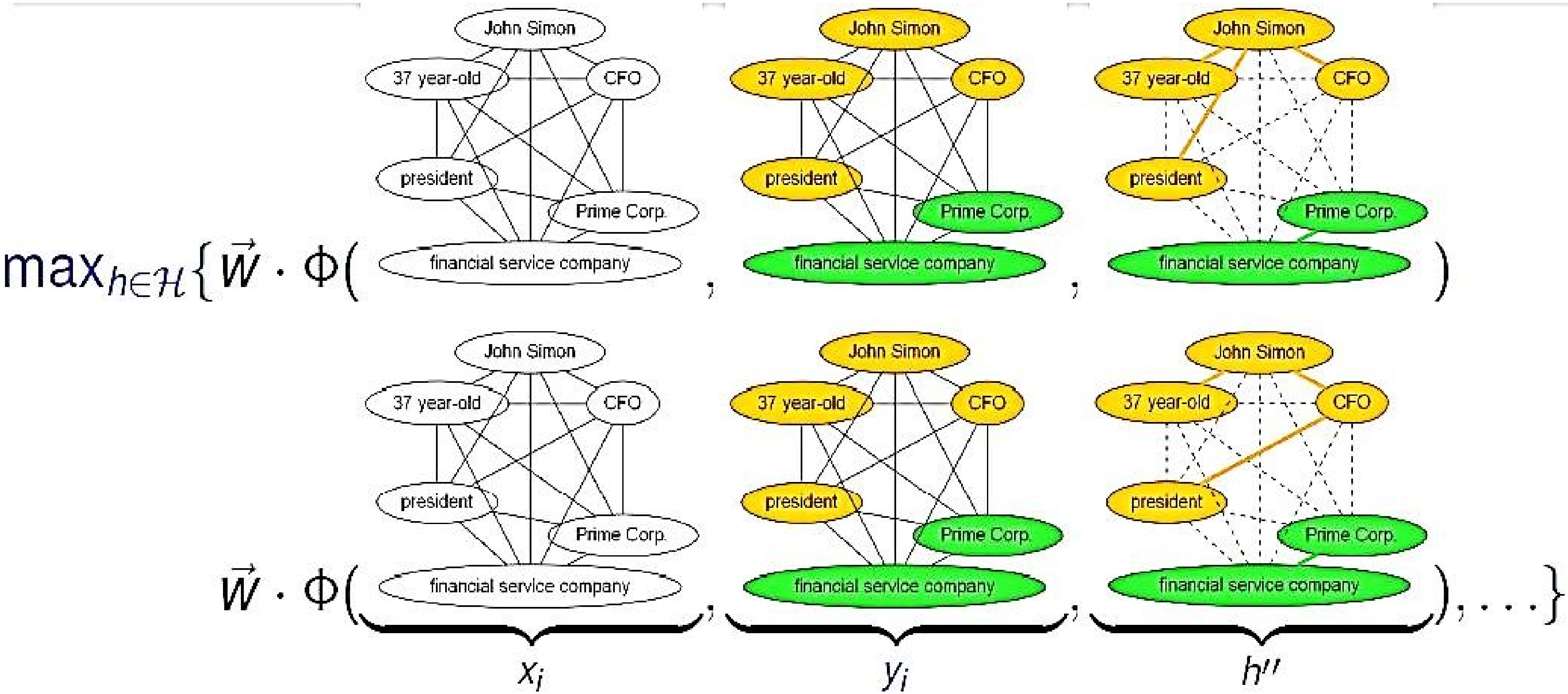


Latent Structured SVM

- **Objective function:**

$$\min_{\vec{w},\vec{\xi}} \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{n}\xi_i \quad s.t.\ \text{for } 1 \le i \le n, \text{ for all outputs } \hat{y} \in \mathcal{Y},$$

$$\max_{h\in\mathcal{H}} \vec{w}\cdot\Phi(x_i, y_i, h) - \max_{\hat{h}\in\mathcal{H}} \vec{w}\cdot\Phi(x_i, \hat{y}, \hat{h}) \ge \Delta(y_i, \hat{y}, \hat{h}) - \xi_i$$

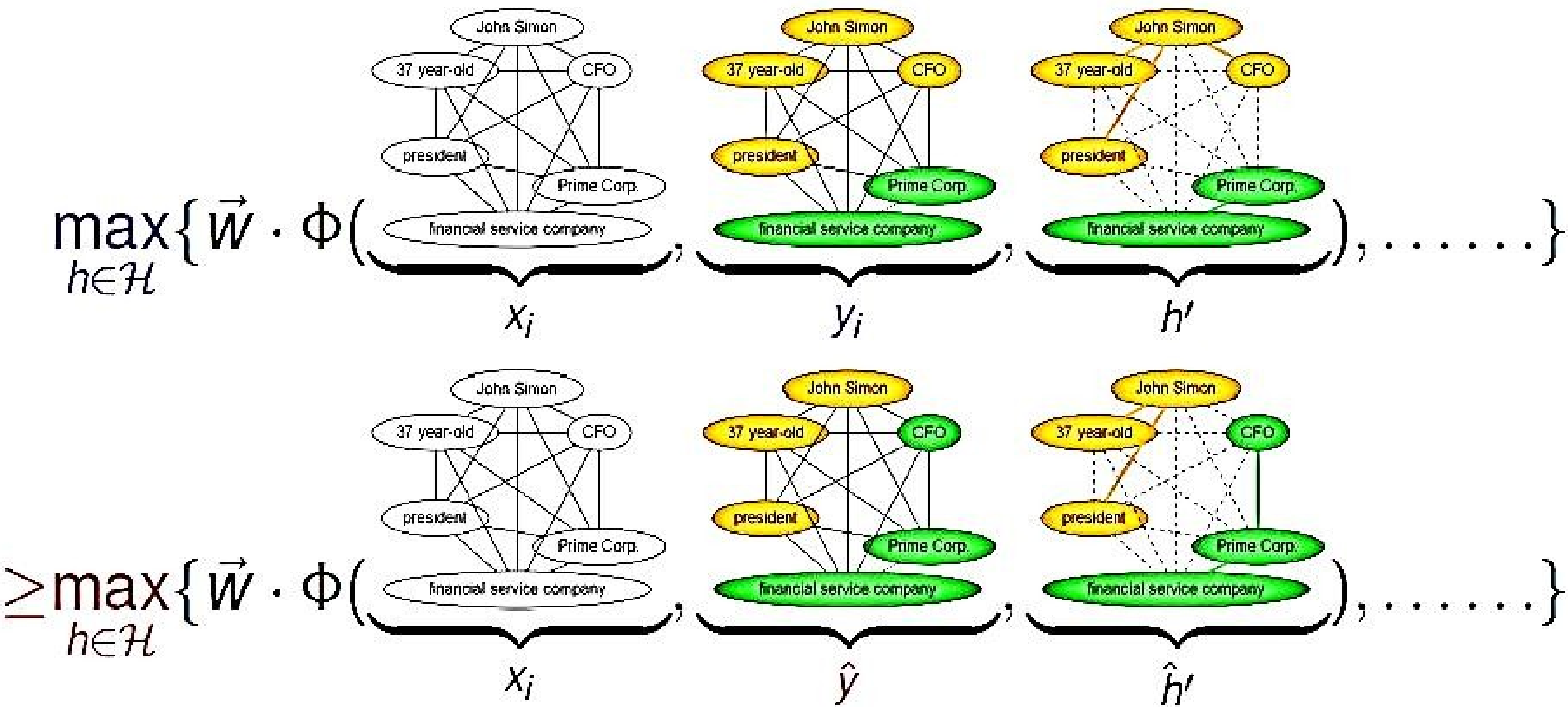


Latent Structured SVM

- **Objective function:**

$$\min_{\vec{w},\vec{\xi}} \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{n}\xi_i \quad s.t.\ \text{for } 1 \le i \le n, \text{ for all outputs } \hat{y} \in \mathcal{Y},$$

$$\max_{h\in\mathcal{H}} \vec{w}\cdot\Phi(x_i, y_i, h) - \max_{\hat{h}\in\mathcal{H}} \vec{w}\cdot\Phi(x_i, \hat{y}, \hat{h}) \ge \Delta(y_i, \hat{y}, \hat{h}) - \xi_i$$



Latent Structured SVM

$$\min_{\boldsymbol{w},\xi} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$\forall i, \xi_i \geq \max_{(\hat{y},\hat{h})\in\mathcal{Y}\times\mathcal{H}} [\boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,\hat{y},\hat{h}) + \Delta(y_i,\hat{y},\hat{h})] - \max_{h\in\mathcal{H}} \boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,y_i,h)$$

- **Final Objective Function:**

$$\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\left(\max_{(\hat{y},\hat{h})\in\mathcal{Y}\times\mathcal{H}} [\boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,\hat{y},\hat{h}) + \Delta(y_i,\hat{y},\hat{h})] - \max_{h\in\mathcal{H}} \boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,y_i,h)\right)$$

$$= \underbrace{\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\max_{(\hat{y},\hat{h})\in\mathcal{Y}\times\mathcal{H}} [\boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,\hat{y},\hat{h}) + \Delta(y_i,\hat{y},\hat{h})]}_{\text{convex}} \underbrace{-C\sum_{i=1}^{n}\max_{h\in\mathcal{H}} \boldsymbol{w}\cdot\boldsymbol{\Phi}(x_i,y_i,h)}_{\text{concave}}$$

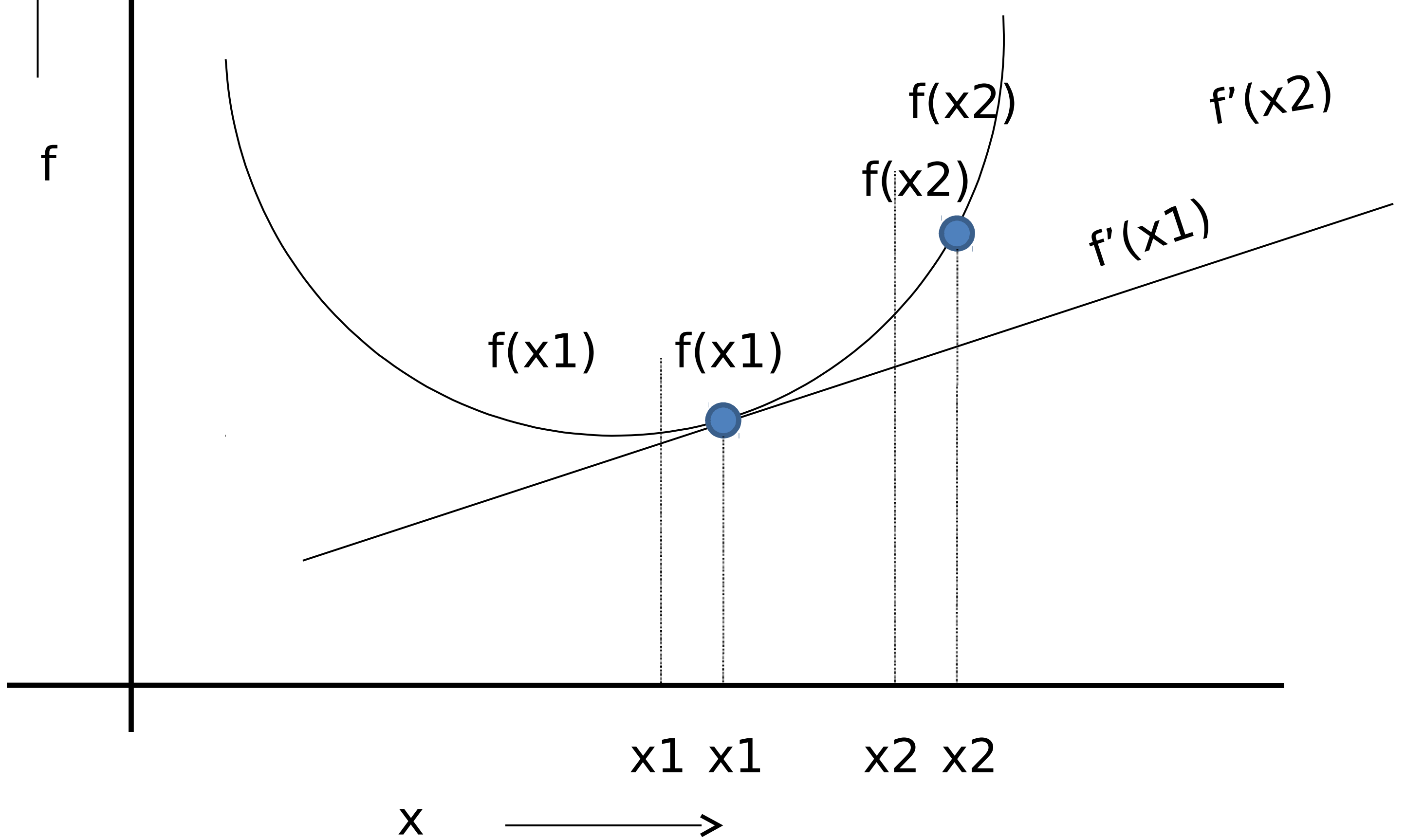
Non-Convex Objective Function ➡ Can't be solved using Cutting plane

Property Of Concave Function

Property Of Convex Function

$$f(x_2) \le f(x_1) + (x_2 - x_1) * f'(x_1)$$

$$f(x_1) \ge f(x_2) + (x_1 - x_2) * f'(x_2)$$

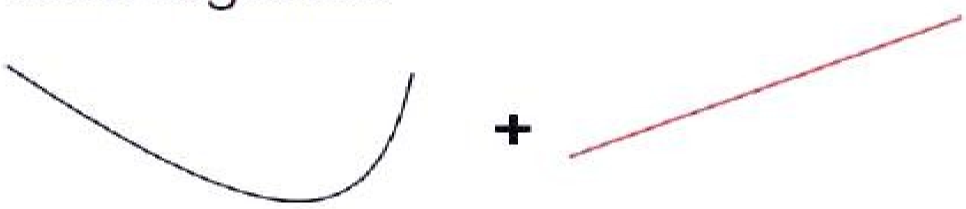


Concave-Convex Procedure

- Decompose the objective into convex and concave part

- Upper bound the concave part with a hyperplane

- Minimize the resulting convex sum. Iterate until convergence



Concave-Convex Procedure

- Decompose the objective into convex and concave part

$$\underbrace{\left[\frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{n} \max_{(\hat{y},\hat{h})\in\mathcal{Y}\times\mathcal{H}} [\vec{w}\cdot\Phi(x_i,\hat{y},\hat{h}) + \Delta(y_i,\hat{y},\hat{h})]\right]}_{\text{convex}}$$

$$\underbrace{-\left[C\sum_{i=1}^{n} \max_{h\in\mathcal{H}} \vec{w}\cdot\Phi(x_i,y_i,h)\right]}_{\text{concave}}$$

Concave-Convex Procedure

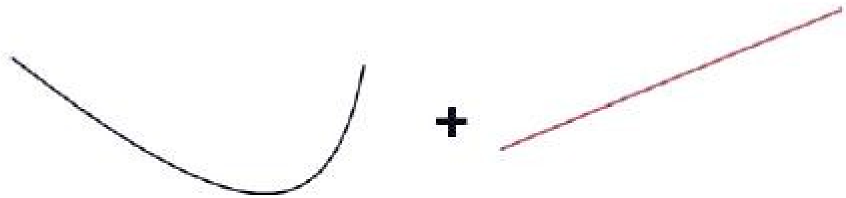
- Upper bound the concave part with a hyperplane

$$\forall \vec{w}, \underbrace{-\left[C\sum_{i=1}^{n} \max_{h\in\mathcal{H}} \vec{w}\cdot\Phi(x_i, y_i, h)\right]}_{\text{concave}} \leq \underbrace{-\left[C\sum_{i=1}^{n} \vec{w}\cdot\Phi(x_i, y_i, h_i^*)\right]}_{\text{linear}}$$

where $h_i^* = \operatorname*{argmax}_{h\in\mathcal{H}} \vec{w}_t\cdot\Phi(x_i, y_i, h)$

Concave-Convex Procedure

- Minimize the resulting sum



$$\min_{\vec{w}} \underbrace{\left[\frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{n} \max_{(\hat{y},\hat{h})\in\mathcal{Y}\times\mathcal{H}} [\vec{w}\cdot\Phi(x_i,\hat{y},\hat{h}) + \Delta(y_i,\hat{y},\hat{h})]\right]}_{\text{convex}} \underbrace{- \left[C\sum_{i=1}^{n} \vec{w}\cdot\Phi(x_i,y_i,h_i^*)\right]}_{\text{linear}}$$

↓

Cutting plane Algorithm → W_{t+1}

- Iterate till desired precision

Overview of the CCCP

- Initialize w_0

repeat
  - Find h* using the w_i
  - Obtain w_{i+1} by optimizing the convex function using cutting plane.
  - Set $w_i = w_{i+1}$

till objective function improves by at least ε

THANK YOU