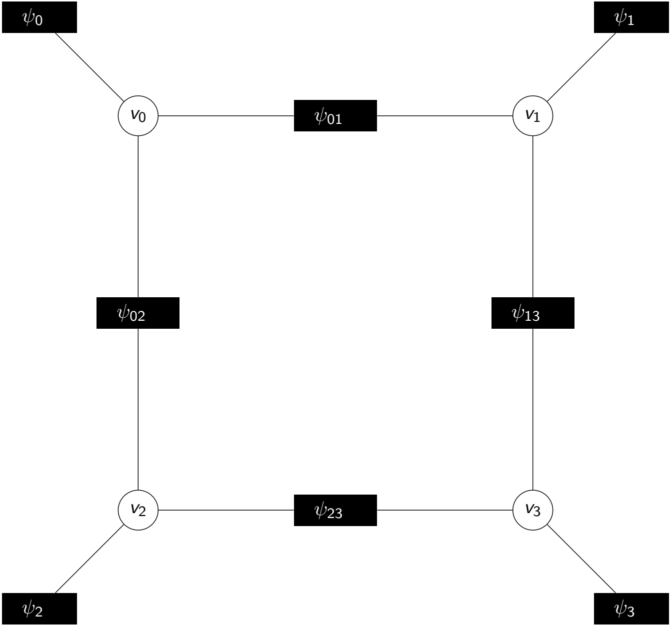
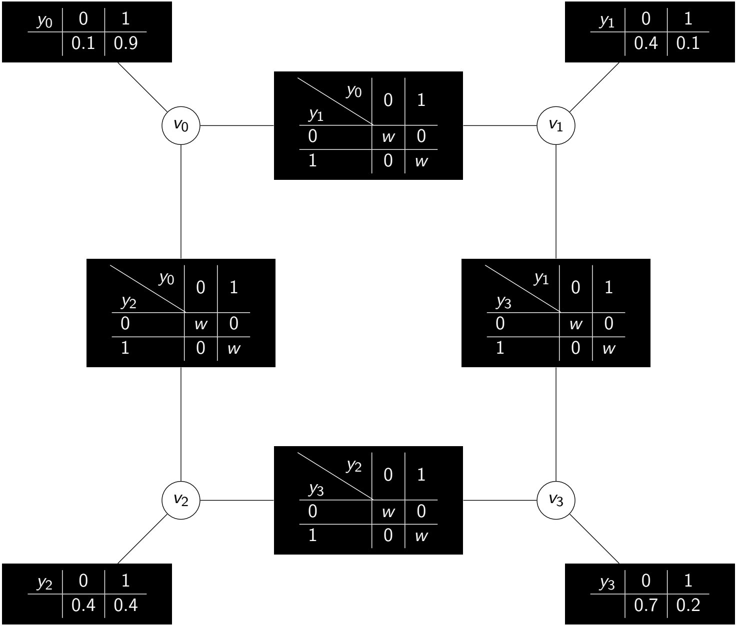
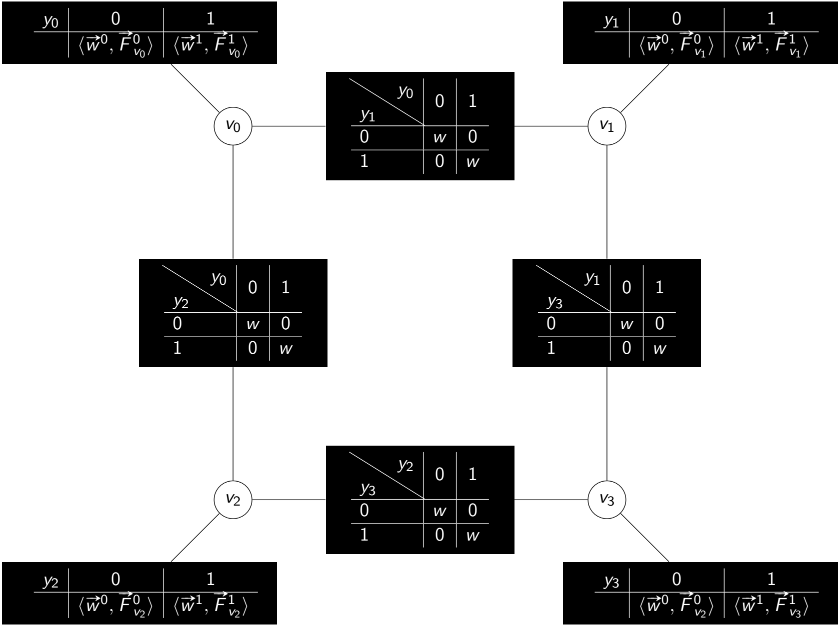
MLCV STRUCT - SVM

Thorsten Beier

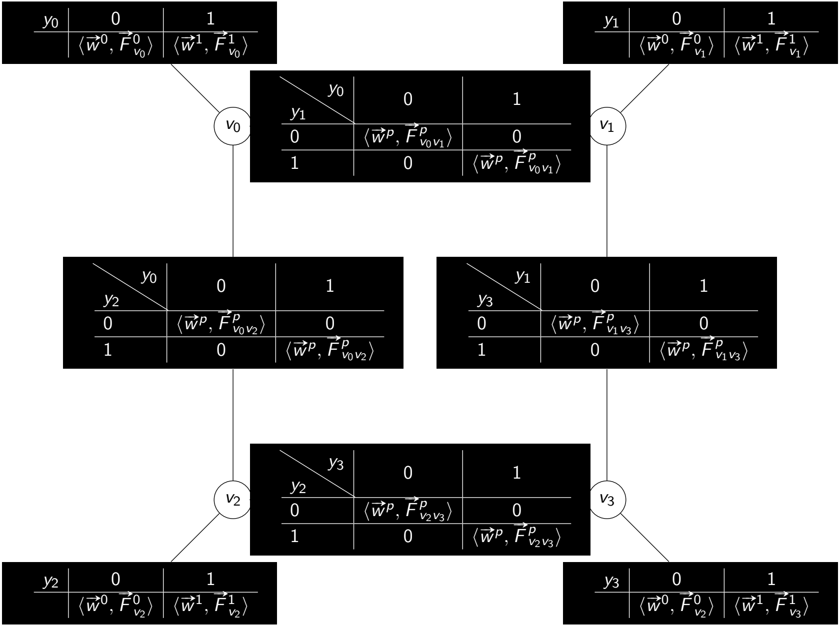
May 16, 2017


ψ_0
ψ_1
v_0
ψ_{01}
v_1
ψ_{02}
ψ_{13}
v_2
ψ_{23}
v_3
ψ_2
ψ_3


y0 | 0 | 1
0.1 | 0.9
y1 | 0 | 1
0.4 | 0.1
v0
v1
y0 y1 | 0 | 1
0 | w | 0
1 | 0 | w
y0 y2 | 0 | 1
0 | w | 0
1 | 0 | w
y1 y3 | 0 | 1
0 | w | 0
1 | 0 | w
y2 y3 | 0 | 1
0 | w | 0
1 | 0 | w
v2
v3
y2 | 0 | 1
0.4 | 0.4
y3 | 0 | 1
0.7 | 0.2



y_0 | 0 | 1
$\langle \vec{w}^0, \vec{F}^0_{v_0} \rangle$ | $\langle \vec{w}^1, \vec{F}^1_{v_0} \rangle$
y_1 | 0 | 1
$\langle \vec{w}^0, \vec{F}^0_{v_1} \rangle$ | $\langle \vec{w}^1, \vec{F}^1_{v_1} \rangle$
v_0
v_1
v_2
v_3
y_0 \ y_1 | 0 | 1
0 | w | 0
1 | 0 | w
y_0 \ y_2 | 0 | 1
0 | w | 0
1 | 0 | w
y_1 \ y_3 | 0 | 1
0 | w | 0
1 | 0 | w
y_2 \ y_3 | 0 | 1
0 | w | 0
1 | 0 | w
y_2 | 0 | 1
$\langle \vec{w}^0, \vec{F}^0_{v_2} \rangle$ | $\langle \vec{w}^1, \vec{F}^1_{v_2} \rangle$
y_3 | 0 | 1
$\langle \vec{w}^0, \vec{F}^0_{v_2} \rangle$ | $\langle \vec{w}^1, \vec{F}^1_{v_3} \rangle$



y_0 | 0 | 1
$\langle \vec{w}^0, \vec{F}^0_{v_0} \rangle$ | $\langle \vec{w}^1, \vec{F}^1_{v_0} \rangle$
y_1 | 0 | 1
$\langle \vec{w}^0, \vec{F}^0_{v_1} \rangle$ | $\langle \vec{w}^1, \vec{F}^1_{v_1} \rangle$
v_0
v_1
y_0 / y_1 | 0 | 1
0 | $\langle \vec{w}^p, \vec{F}^p_{v_0 v_1} \rangle$ | 0
1 | 0 | $\langle \vec{w}^p, \vec{F}^p_{v_0 v_1} \rangle$
y_0 / y_2 | 0 | 1
0 | $\langle \vec{w}^p, \vec{F}^p_{v_0 v_2} \rangle$ | 0
1 | 0 | $\langle \vec{w}^p, \vec{F}^p_{v_0 v_2} \rangle$
y_1 / y_3 | 0 | 1
0 | $\langle \vec{w}^p, \vec{F}^p_{v_1 v_3} \rangle$ | 0
1 | 0 | $\langle \vec{w}^p, \vec{F}^p_{v_1 v_3} \rangle$
v_2
v_3
y_3 / y_2 | 0 | 1
0 | $\langle \vec{w}^p, \vec{F}^p_{v_2 v_3} \rangle$ | 0
1 | 0 | $\langle \vec{w}^p, \vec{F}^p_{v_2 v_3} \rangle$
y_2 | 0 | 1
$\langle \vec{w}^0, \vec{F}^0_{v_2} \rangle$ | $\langle \vec{w}^1, \vec{F}^1_{v_2} \rangle$
y_3 | 0 | 1
$\langle \vec{w}^0, \vec{F}^0_{v_2} \rangle$ | $\langle \vec{w}^1, \vec{F}^1_{v_3} \rangle$

Let x be an image:
So far we had "arbitrary" potentials ψ and regularizers w

$$\hat{y} = \underset{y \in \{0,1\}^{\mathcal{V}}}{\operatorname{argmax}} \sum_{v_i \in \mathcal{V}} \psi_i(y_i) + \sum_{v_i v_j \in \mathcal{E}} w \cdot [y_i \neq y_j] \tag{1}$$

Let x be an image:
So far we had "arbitrary" potentials ψ and regularizers w

$$\hat{y} = \underset{y \in \{0,1\}^{\mathcal{V}}}{\operatorname{argmax}} \sum_{v_i \in \mathcal{V}} \psi_i(y_i) + \sum_{v_i v_j \in \mathcal{E}} w \cdot [y_i \neq y_j] \tag{1}$$

Lets generalize this a bit:
Let $\phi_i(y_i)$ and $\phi_{ij}(y_i, y_j)$ be functions returning feature vectors. w_a and w_b are the weight vectors we want to learn.

$$\hat{Y} = \underset{Y \in \{0,1\}^{\mathcal{V}}}{\operatorname{argmax}} \sum_{v_i \in \mathcal{V}} \langle \vec{w}_a, \phi_i(y_i) \rangle + \sum_{v_i v_j \in \mathcal{E}} \langle \vec{w}_b, \phi_{ij}(y_i, y_j) \rangle \tag{2}$$

Let x be an image:
So far we had "arbitrary" potentials ψ and regularizers w

$$\hat{y} = \underset{y \in \{0,1\}^{\mathcal{V}}}{\operatorname{argmax}} \sum_{v_i \in \mathcal{V}} \psi_i(y_i) + \sum_{v_i v_j \in \mathcal{E}} w \cdot [y_i \neq y_j] \tag{1}$$

Lets generalize this a bit:
Let $\phi_i(y_i)$ and $\phi_{ij}(y_i, y_j)$ be functions returning feature vectors. w_a and w_b are the weight vectors we want to learn.

$$\hat{Y} = \underset{Y \in \{0,1\}^{\mathcal{V}}}{\operatorname{argmax}} \sum_{v_i \in \mathcal{V}} \langle \vec{w}_a, \phi_i(y_i) \rangle + \sum_{v_i v_j \in \mathcal{E}} \langle \vec{w}_b, \phi_{ij}(y_i, y_j) \rangle \tag{2}$$

We can generalize this a bit more:
Let x be the input data. $\phi(x, y)$ is a function extracting feature vectors for a given labeling y.

$$\hat{y} = \underset{y \in \{0,1\}}{\operatorname{argmax}} \quad \langle \vec{w}, \phi(x, y) \rangle \tag{3}$$

Struct Svm Ingredients:

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\dots,n}$

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\ldots,n}$
  - x and y can be of any structure

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\dots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\ldots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\dots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$
  - Expresses relation between x and y

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\ldots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$
  - Expresses relation between x and y
  - $\phi(x, y)$ has a constant dimensionality

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\dots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$
  - Expresses relation between x and y
  - $\phi(x, y)$ has a constant dimensionality
- Set of weights $\mathrm{w} \in \mathcal{W}$

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\dots,n}$
    - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$
    - Expresses relation between x and y
    - $\phi(x, y)$ has a constant dimensionality
- Set of weights $\mathrm{w} \in \mathcal{W} \rightarrow$ we want to learn these weights

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\ldots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$
  - Expresses relation between x and y
  - $\phi(x, y)$ has a constant dimensionality
- Set of weights w $\in \mathcal{W} \rightarrow$ we want to learn these weights
- Score Function $f(x, y, w) = w^T \phi(x, y)$

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\ldots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$
  - Expresses relation between x and y
  - $\phi(x, y)$ has a constant dimensionality
- Set of weights w $\in \mathcal{W} \rightarrow$ we want to learn these weights
- Score Function $f(x, y, w) = w^T \phi(x, y)$
  - Score should we high for y^i and low for all other $y \in \mathcal{Y}(x_i)$

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\dots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$
  - Expresses relation between x and y
  - $\phi(x, y)$ has a constant dimensionality
- Set of weights $\mathrm{w} \in \mathcal{W} \rightarrow$ we want to learn these weights
- Score Function $f(x, y, w) = w^T \phi(x, y)$
  - Score should we high for y^i and low for all other $y \in \mathcal{Y}(x_i)$
- Loss Function $\Delta(y^i, y)$

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\ldots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$
  - Expresses relation between x and y
  - $\phi(x, y)$ has a constant dimensionality
- Set of weights w $\in \mathcal{W} \rightarrow$ we want to learn these weights
- Score Function $f(x, y, w) = w^T\phi(x, y)$
  - Score should we high for y^i and low for all other $y \in \mathcal{Y}(x_i)$
- Loss Function $\Delta(y^i, y)$

$$\langle w, \phi(x^i, y^i)\rangle - \langle w, \phi(x^i, y)\rangle \quad \geq 1 \qquad \forall i \in \{1, \ldots, n\}$$
$$\forall y \in \mathcal{Y}(x^i) \setminus y^i$$

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\dots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$
  - Expresses relation between x and y
  - $\phi(x, y)$ has a constant dimensionality
- Set of weights $\mathrm{w} \in \mathcal{W} \rightarrow$ we want to learn these weights
- Score Function $f(x, y, w) = w^T \phi(x, y)$
  - Score should we high for y^i and low for all other $y \in \mathcal{Y}(x_i)$
- Loss Function $\Delta(y^i, y)$

$$\langle w, \phi(x^i, y^i)\rangle - \langle w, \phi(x^i, y)\rangle \quad \geq 1 - \xi^i \qquad \begin{array}{r} \forall i \in \{1, \dots, n\} \\ \forall y \in \mathcal{Y}(x^i) \setminus y^i \end{array}$$

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\ldots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$
  - Expresses relation between x and y
  - $\phi(x, y)$ has a constant dimensionality
- Set of weights w $\in \mathcal{W} \rightarrow$ we want to learn these weights
- Score Function $f(x, y, w) = w^T \phi(x, y)$
  - Score should we high for y^i and low for all other $y \in \mathcal{Y}(x_i)$
- Loss Function $\Delta(y^i, y)$

$$
\begin{aligned}
w_{opt} = \underset{w}{\operatorname{argmin}} \quad & \frac{c}{n} \sum \xi^i \\
s.t. \quad & \xi^i \geq 0 \\
& \langle w, \phi(x^i, y^i) \rangle - \langle w, \phi(x^i, y) \rangle \quad \geq 1 - \xi^i \qquad \forall i \in \{1, \ldots, n\} \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall y \in \mathcal{Y}(x^i) \setminus y^i
\end{aligned}
$$

Struct Svm Ingredients:

- Training Set: $\mathcal{T} = \{x^i, y^i\}_{i=1,\ldots,n}$
  - x and y can be of any structure
- Set of feasible outputs $\mathcal{Y}(x)$ for some input x
- feature function: $\phi(x, y)$
  - Expresses relation between x and y
  - $\phi(x, y)$ has a constant dimensionality
- Set of weights w $\in \mathcal{W} \rightarrow$ we want to learn these weights
- Score Function $f(x, y, w) = w^T \phi(x, y)$
  - Score should we high for y^i and low for all other $y \in \mathcal{Y}(x_i)$
- Loss Function $\Delta(y^i, y)$

$$
\begin{aligned}
w_{opt} = \underset{w}{\operatorname{argmin}} \quad & w^T w + \frac{c}{n} \sum \xi^i \\
s.t. \quad & \xi^i \geq 0 \\
& \langle w, \phi(x^i, y^i) \rangle - \langle w, \phi(x^i, y) \rangle \quad \geq 1 - \xi^i \qquad \forall i \in \{1, \ldots, n\} \\
& \qquad \forall y \in \mathcal{Y}(x^i) \setminus y^i
\end{aligned}
$$

Struct SVM Formulation:

$$w_{opt} = \underset{w}{\operatorname{argmin}} \quad w^T w + \frac{c}{n} \sum \xi^i$$

$$s.t. \quad \xi^i \geq 0$$

$$\langle w, \phi(x^i, y^i) \rangle - \langle w, \phi(x^i, y) \rangle \quad \geq 1 - \xi^i \qquad \forall i \in \{1, \ldots, n\}$$

$$\forall y \in \mathcal{Y}(x^i) \setminus y^i$$

Struct SVM Formulation:

$$w_{opt} = \underset{w}{\operatorname{argmin}} \quad w^T w + \frac{c}{n} \sum \xi^i$$

$$s.t. \quad \xi^i \geq 0$$

$$\langle w, \phi(x^i, y^i) \rangle - \langle w, \phi(x^i, y) \rangle \quad \geq 1 - \xi^i \qquad \forall i \in \{1, \ldots, n\}$$

$$\forall y \in \mathcal{Y}(x^i) \setminus y^i$$

- Quadratic Program (QP)

Struct SVM Formulation:

$$
\begin{aligned}
w_{opt} = \underset{w}{\operatorname{argmin}} \quad & w^T w + \frac{c}{n} \sum \xi^i \\
s.t. \quad & \xi^i \geq 0 \\
& \langle w, \phi(x^i, y^i) \rangle - \langle w, \phi(x^i, y) \rangle \quad \geq 1 - \xi^i \qquad \forall i \in \{1, \dots, n\} \\
& \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \forall y \in \mathcal{Y}(x^i) \setminus y^i
\end{aligned}
$$

- Quadratic Program (QP)

- $\mathcal{Y}(x)$ is exponentially large!

Struct SVM Formulation:

$$
\begin{aligned}
w_{opt} = \underset{w}{\operatorname{argmin}} \quad & w^T w + \frac{c}{n} \sum \xi^i \\
s.t. \quad & \xi^i \geq 0 \\
& \langle w, \phi(x^i, y^i) \rangle - \langle w, \phi(x^i, y) \rangle \quad \geq 1 - \xi^i \qquad \forall i \in \{1, \dots, n\} \\
& \qquad \forall y \in \mathcal{Y}(x^i) \setminus y^i
\end{aligned}
$$

- Quadratic Program (QP)

- $\mathcal{Y}(x)$ is exponentially large!

- $\rightarrow$ Cutting planes approach

Struct SVM Formulation:

$$w_{opt} = \underset{w}{\operatorname{argmin}} \quad w^T w + \frac{c}{n} \sum \xi^i$$

$$s.t. \quad \xi^i \geq 0$$

$$\langle w, \phi(x^i, y^i) \rangle - \langle w, \phi(x^i, y) \rangle \quad \geq 1 - \xi^i \qquad \forall i \in \{1, \ldots, n\}$$

$$\forall y \in \mathcal{Y}(x^i) \setminus y^i$$

- Quadratic Program (QP)

- $\mathcal{Y}(x)$ is exponentially large!

- $\rightarrow$ Cutting planes approach
- $\rightarrow$ Stochastic sub-gradient decent

Cutting Plane Struct SVM

Instead of adding all constraints in to a QP

$$\begin{aligned} \langle w, \phi(x^i, y^i)\rangle - \langle w, \phi(x^i, y)\rangle \quad & \geq 1 - \xi^i \quad \forall i \in \{1, \dots, n\} \\ & \qquad\qquad \forall y \in \mathcal{Y}(x^i) \setminus y^i \end{aligned}$$

we add only the *most violated constraints*:

Cutting Plane Struct SVM:

$w = 0 \quad \xi = 0$ Constraint Set: $\mathcal{K} = \cup_i \{\xi^i \geq 0\}$

Cutting Plane Struct SVM:

$w = 0 \quad \xi = 0$ Constraint Set: $\mathcal{K} = \cup_i \{\xi^i \geq 0\}$

REPEAT:

Cutting Plane Struct SVM:

$w = 0 \quad \xi = 0$ Constraint Set: $\mathcal{K} = \cup_i \{\xi^i \geq 0\}$

REPEAT:

$\mathcal{K}_{old} = \mathcal{K}$

Cutting Plane Struct SVM:

$w = 0 \quad \xi = 0$ Constraint Set: $\mathcal{K} = \cup_i \{\xi^i \geq 0\}$

REPEAT:

$\mathcal{K}_{old} = \mathcal{K}$

- **FOR** $i = 1$ **TO** n :

Cutting Plane Struct SVM:

$w = 0 \quad \xi = 0$ Constraint Set: $\mathcal{K} = \cup_i \{\xi^i \geq 0\}$

REPEAT:

$\mathcal{K}_{old} = \mathcal{K}$

- **FOR** $i = 1$ **TO** n :
  - $$\hat{y} = \underset{y \in \mathcal{Y}(x^i) \setminus y^i}{\operatorname{argmax}} - \left(\overbrace{w^T \phi(x^i, y^i) - w^T \phi(x^i, y)}^{=:\text{margin}} \right)$$

Cutting Plane Struct SVM:

$w = 0 \quad \xi = 0$ Constraint Set: $\mathcal{K} = \cup_i \{\xi^i \geq 0\}$

REPEAT:

$\mathcal{K}_{old} = \mathcal{K}$

- **FOR** $i = 1$ **TO** n :
  - $\hat{y} = \underset{y \in \mathcal{Y}(x^i) \setminus y^i}{\operatorname{argmax}} - \left(\overbrace{w^T \phi(x^i, y^i) - w^T \phi(x^i, y)}^{=:\text{margin}} \right)$
  - **IF** $\left(\underbrace{w^T \phi(x^i, y^i) - w^T \phi(x^i, \hat{y})}_{=:\text{margin of } \hat{y}} \right) \leq 1 - \xi^i$:

Cutting Plane Struct SVM:

$w = 0 \quad \xi = 0$ Constraint Set: $\mathcal{K} = \cup_i \{\xi^i \geq 0\}$

REPEAT:

$\mathcal{K}_{old} = \mathcal{K}$

- **FOR** $i = 1$ **TO** n :
  - $$\hat{y} = \underset{y \in \mathcal{Y}(x^i) \setminus y^i}{\operatorname{argmax}} - \left(\overbrace{w^T \phi(x^i, y^i) - w^T \phi(x^i, y)}^{=:\text{margin}} \right)$$
  - **IF** $$\left(\underbrace{w^T \phi(x^i, y^i) - w^T \phi(x^i, \hat{y})}_{=:\text{margin of } \hat{y}} \right) \leq 1 - \xi^i :$$
    - $\mathcal{K} = \mathcal{K} \cup \{w^T \phi(x^i, y^i) - w^T \phi(x^i, \hat{y}) \geq 1 - \xi^i\}$

Cutting Plane Struct SVM:

$w = 0 \quad \xi = 0$ Constraint Set: $\mathcal{K} = \cup_i \{\xi^i \geq 0\}$

REPEAT:

$\mathcal{K}_{old} = \mathcal{K}$

- **FOR** $i = 1$ **TO** n :
  - $\hat{y} = \underset{y \in \mathcal{Y}(x^i) \setminus y^i}{\text{argmax}} - \left(\overbrace{w^T\phi(x^i, y^i) - w^T\phi(x^i, y)}^{=:\text{margin}} \right)$
  - **IF** $\left(\underbrace{w^T\phi(x^i, y^i) - w^T\phi(x^i, \hat{y})}_{=:\text{margin of } \hat{y}} \right) \leq 1 - \xi^i$:
    - $\mathcal{K} = \mathcal{K} \cup \left\{ w^T\phi(x^i, y^i) - w^T\phi(x^i, \hat{y}) \geq 1 - \xi^i \right\}$
    - $\{w, \xi\} = \underset{w,\xi}{\text{argmin}}\, w^T w + \frac{c}{n} \sum_i^n \xi^i \qquad s.t \quad \mathcal{K}$

Cutting Plane Struct SVM:

$w = 0 \quad \xi = 0$ Constraint Set: $\mathcal{K} = \cup_i \{\xi^i \geq 0\}$

REPEAT:

$\mathcal{K}_{old} = \mathcal{K}$

- **FOR** $i = 1$ **TO** n :
  - $\hat{y} = \underset{y \in \mathcal{Y}(x^i) \setminus y^i}{\text{argmax}} - \left(\overbrace{w^T \phi(x^i, y^i) - w^T \phi(x^i, y)}^{=:\text{margin}} \right)$
  - **IF** $\left(\underbrace{w^T \phi(x^i, y^i) - w^T \phi(x^i, \hat{y})}_{=:\text{margin of } \hat{y}} \right) \leq 1 - \xi^i$:
    - $\mathcal{K} = \mathcal{K} \cup \{ w^T \phi(x^i, y^i) - w^T \phi(x^i, \hat{y}) \geq 1 - \xi^i \}$
    - $\{w, \xi\} = \underset{w, \xi}{\text{argmin}}\, w^T w + \frac{c}{n} \sum_i^n \xi^i \qquad s.t \quad \mathcal{K}$

UNTIL: $\mathcal{K}_{old} = \mathcal{K}$

Margin Rescaled Cutting Plane Struct SVM:

$w = 0 \quad \xi = 0$ Constraint Set: $\mathcal{K} = \cup_i \{\xi^i \geq 0\}$

REPEAT:

$\mathcal{K}_{old} = \mathcal{K}$

- **FOR** $i = 1$ **TO** n :
  - $\hat{y} = \underset{y \in \mathcal{Y}(x^i) \setminus y^i}{\text{argmax}} - \left(\overbrace{w^T \phi(x^i, y^i) - w^T \phi(x^i, y)}^{=:\text{margin}} \right) + \Delta(y^i, y)$
  - **IF** $\left(\underbrace{w^T \phi(x^i, y^i) - w^T \phi(x^i, \hat{y})}_{=:\text{margin of } \hat{y}} \right) \leq \Delta(y^i, \hat{y}) + 1 - \xi^i$:
    - $\mathcal{K} = \mathcal{K} \cup \left\{ w^T \phi(x^i, y^i) - w^T \phi(x^i, \hat{y}) \geq \Delta(y^i, \hat{y}) + 1 - \xi^i \right\}$
    - $\{w, \xi\} = \underset{w,\xi}{\text{argmin}}\, w^T w + \frac{c}{n} \sum_i^n \xi^i \qquad s.t \quad \mathcal{K}$

UNTIL: $\mathcal{K}_{old} = \mathcal{K}$