

Deep Learning With Constraints

Yatin Nandwani

Work done in collaboration with

Abhishek Pathak

Under the guidance of

Prof. Mausam and Prof. Parag Singla

Learning with Constraints:

Motivation

→ Modern day AI == Deep Learning (DL) [Learn from Data]

Learning with Constraints:

Motivation

- Modern day AI == Deep Learning (DL) [Learn from Data]
- Can we inject symbolic knowledge in Deep Learning? E.g. Person
 - => Noun [Learn from ~~Data~~ Knowledge](credit: Vivek S Kumar)

Learning with Constraints:

Motivation

- Modern day AI == Deep Learning (DL) [Learn from Data]
- Can we inject symbolic knowledge in Deep Learning? E.g. Person
=> Noun [Learn from ~~Data~~ Knowledge]
- **Constraints:** One of the ways of representing symbolic knowledge.

$$\mathbb{1}\{y_{PER.} = 1\} \implies \mathbb{1}\{y_{Noun.} = 1\}$$

Learning with Constraints:

Motivation

- **Task:** Fine Grained Entity Typing

Learning with Constraints:

Motivation

Input:

Bag of Mentions

Sample Mention:
the United States”

“Barack Obama is the President of

Output:
*president, leader,
politician...*

Learning with Constraints:

Motivation

Input:

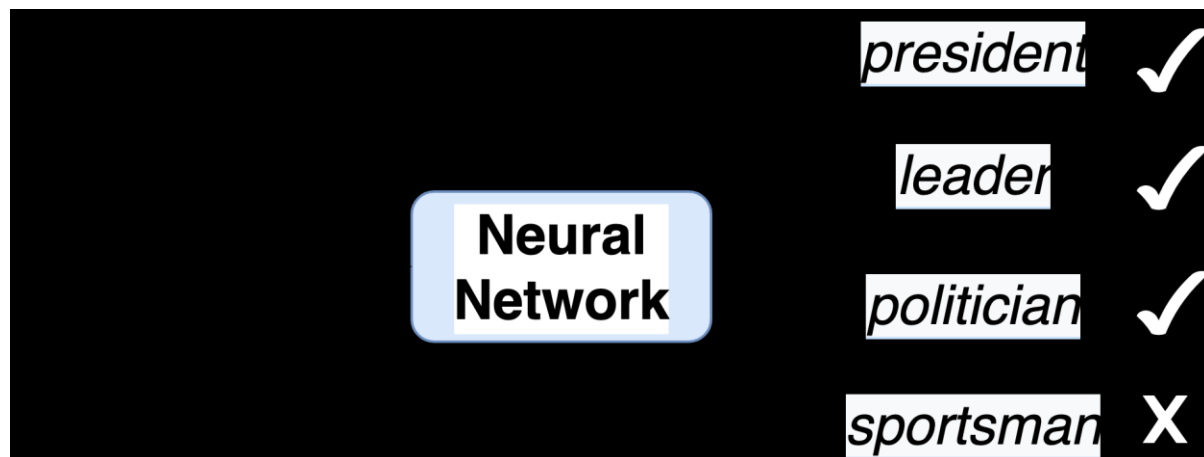
Bag of Mentions

Sample Mention:
the United States"

"Barack Obama is the President of

Output:

*president, leader,
politician...*



Learning with Constraints:

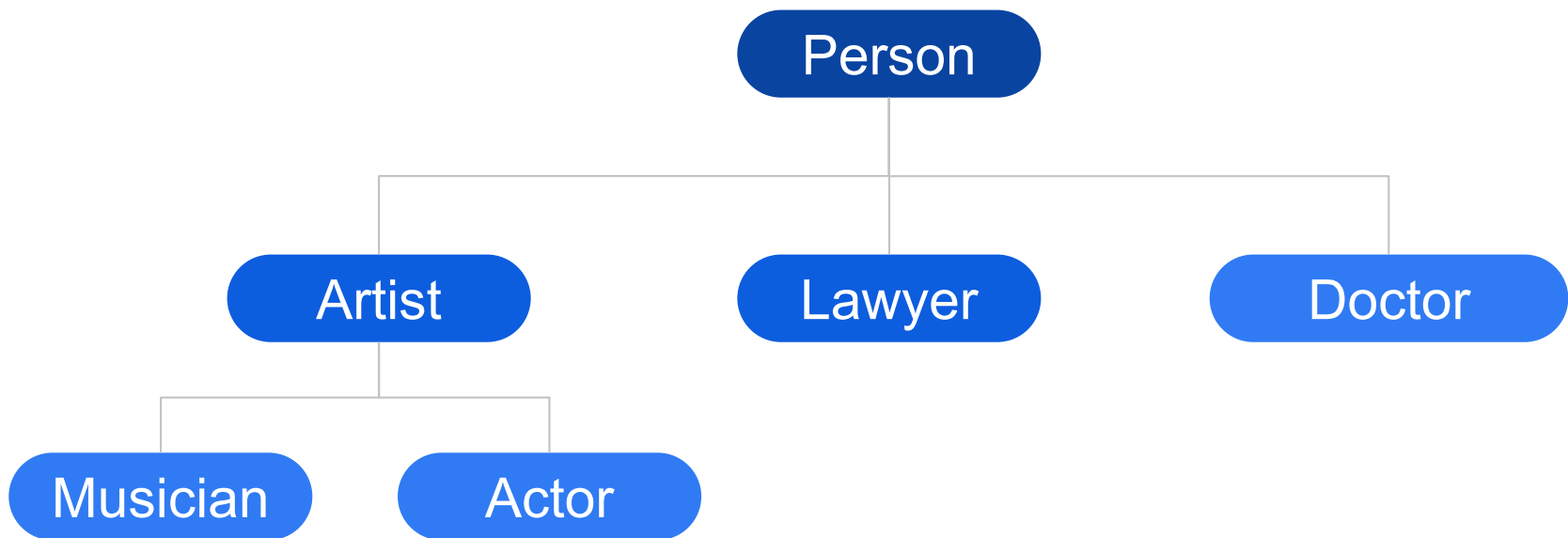
Motivation

- **Constraints:** Hierarchy on Output label space

Learning with Constraints:

Motivation

- **Constraints:** Hierarchy on Output label space



Learning with Constraints:

Motivation

→ Using Soft Logic

$$\mathbb{1}\{y_{ARTIST} = 1\} \implies \mathbb{1}\{y_{PERSON} = 1\}$$

Learning with Constraints:

Motivation

→ Using Soft Logic

$$\mathbb{1} \{y_{ARTIST} = 1\} \implies \mathbb{1} \{y_{PERSON} = 1\}$$

$$(\neg \mathbb{1} \{y_{ARTIST} = 1\}) \vee (\mathbb{1} \{y_{PERSON} = 1\})$$

Learning with Constraints:

Motivation

→ Using Soft Logic

$$\mathbb{1} \{y_{ARTIST} = 1\} \implies \mathbb{1} \{y_{PERSON} = 1\}$$

$$(\neg \mathbb{1} \{y_{ARTIST} = 1\}) \vee (\mathbb{1} \{y_{PERSON} = 1\})$$

$$(1 - p(y_{ARTIST})) + p(y_{PERSON})$$

Learning with Constraints:

Re

→

Boolean Expression	T-norm: Choice 1	T-norm: Choice 2
v	$p(v = 1)$	
$\neg v$	$1 - p(v = 1)$	
$v_1 \vee v_2$	$\min(p(v_1 = 1) + p(v_2 = 1), 1)$	$\max(p(v_1 = 1), p(v_2 = 1))$
$v_1 \wedge v_2$	$\max(p(v_1 = 1) + p(v_2 = 1) - 1, 0)$	$\min(p(v_1 = 1), p(v_2 = 1))$

$$\mathbb{1} \{y_{ARTIST} = 1\} \implies \mathbb{1} \{y_{PERSON} = 1\}$$

$$(\neg \mathbb{1} \{y_{ARTIST} = 1\}) \vee (\mathbb{1} \{y_{PERSON} = 1\})$$

$$(1 - p(y_{ARTIST})) + p(y_{PERSON})$$

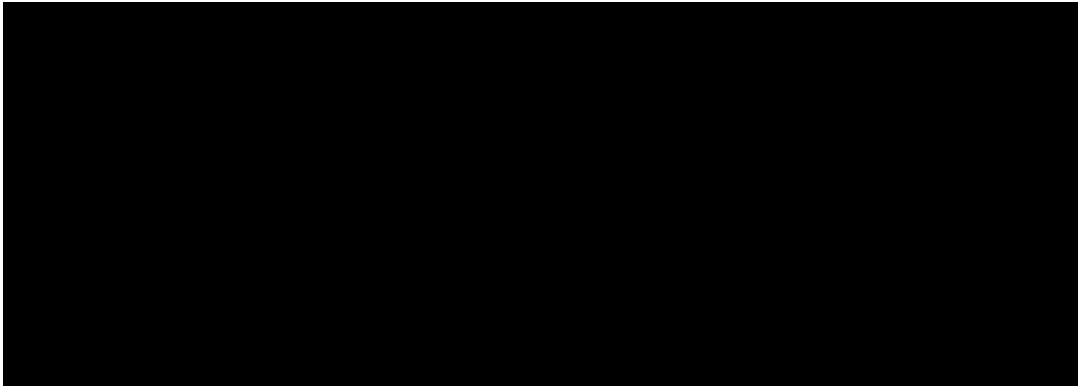
Learning with Constraints:

Representation of Constraints



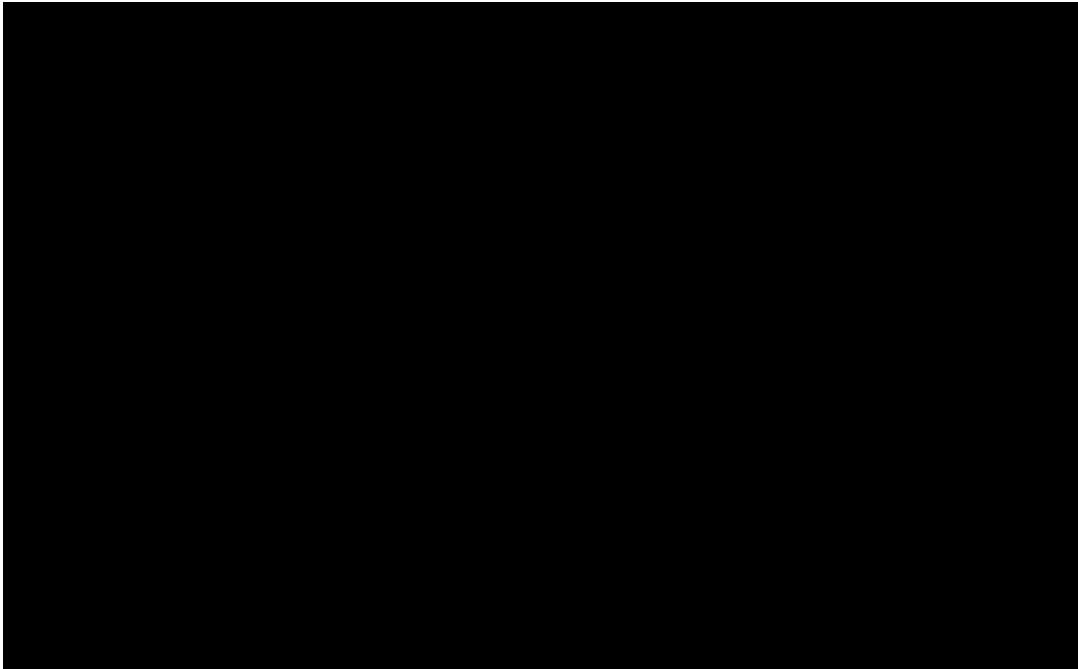
Learning with Constraints:

Representation of Constraints



Learning with Constraints:

Representation of Constraints



Learning with Constraints:

Representation of Constraints

Define:

$$f_k^i = p(y_{ARTIST}) - p(y_{PERSON})$$

k^{th} Constraint

i^{th} Data point

Inequality Constraint:

$$f_k^i \leq 0$$

Learning with Constraints:

Formulation

Unconstrained Problem

$$\min_w L(w)$$

$L(w)$: Any standard loss function,
say Cross Entropy

Learning with Constraints:

Formulation

Unconstrained Problem

$$\min_w L(w) \quad L(w) : \text{Any standard loss function, say Cross Entropy}$$

Constrained Problem

$$\min_w L(w) \quad \text{subject to} \quad f_k^i(w) \leq 0; \quad \forall 1 \leq i \leq m; \quad \forall 1 \leq k \leq K$$

Learning with Constraints:

Formulation

Constrained Problem

$$\min_w L(w) \quad \text{subject to} \quad f_k^i(w) \leq 0; \quad \forall 1 \leq i \leq m; \quad \forall 1 \leq k \leq K$$

Where:

m: Size of training data

K: Number of Constraints

Learning with Constraints:

Formulation

Constrained Problem

$$\min_w L(w) \quad \text{subject to} \quad f_i(w) \leq 0, \quad \forall 1 \leq i \leq m, \quad \forall 1 \leq k \leq K$$

$$\mathcal{L}(w, \Lambda) = L(w) + \sum_{i=1}^m \sum_{k=1}^K \lambda_k^i f_k^i(w)$$

Learning with Constraints:

Formulation

Constrained Problem

$$\min_w L(w) \quad \text{subject to} \quad f_k^i(w) \leq 0; \quad \forall 1 \leq i \leq m; \quad \forall 1 \leq k \leq K$$

Where:

m: Size of training data

K: Number of Constraints

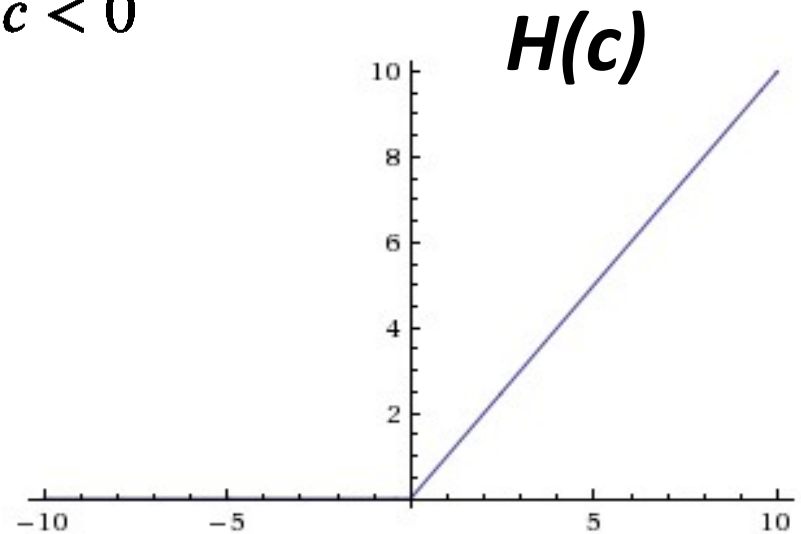
Issue:

O(mK) #constraints

i.e. ***mK*** Lagrange Multipliers!

Learning with Constraints: *Reduce # Constraints*

$$H(c) = c \text{ for } c \geq 0, \text{ and } 0 \text{ for } c < 0$$

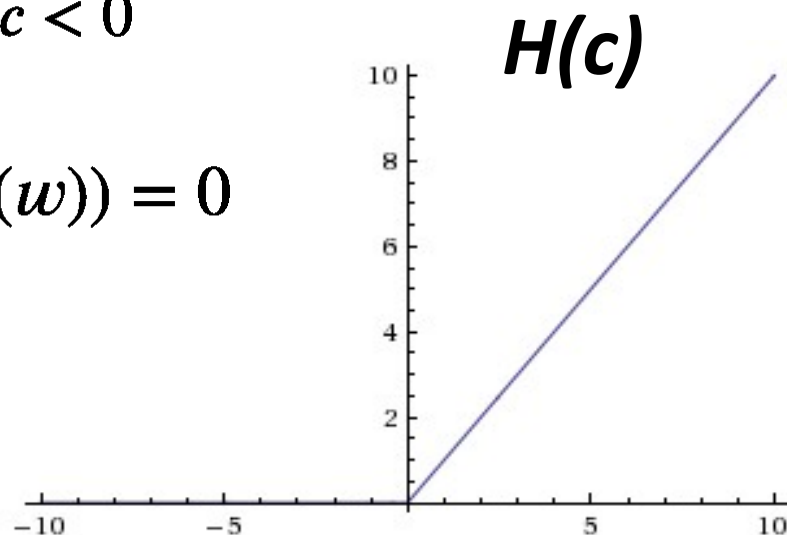


Learning with Constraints: *Reduce # Constraints*

$$H(c) = c \text{ for } c \geq 0, \text{ and } 0 \text{ for } c < 0$$

$$f_k^i(w) \leq 0 \quad \equiv \quad H(f_k^i(w)) = 0$$

Equivalent



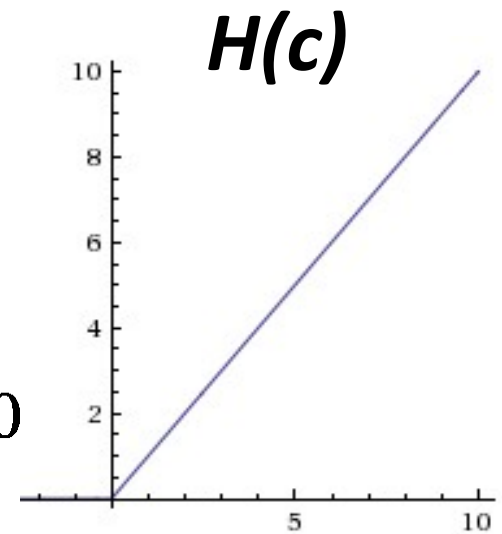
Learning with Constraints: *Reduce # Constraints*

$$H(c) = c \text{ for } c \geq 0, \text{ and } 0 \text{ for } c < 0$$

$$f_k^i(w) \leq 0 \quad \equiv \quad H(f_k^i(w)) = 0$$

Equivalent

$$\forall i : H(f_k^i(w)) = 0 \quad \equiv \quad \sum_i H(f_k^i(w)) = 0$$



Learning with Constraints: *Reduce # Constraints*

Originally:

$$\min_w L(w) \quad \text{subject to} \quad f_k^i(w) \leq 0; \quad \forall 1 \leq i \leq m; \quad \forall 1 \leq k \leq K$$

Learning with Constraints: *Reduce # Constraints*

Originally:

$$\min_w L(w) \quad \text{subject to} \quad f_k^i(w) \leq 0; \quad \forall 1 \leq i \leq m; \quad \forall 1 \leq k \leq K$$

Now:

$$\text{Define: } h_k(w) = \sum_i H(f_k^i(w))$$

$$\min_w L(w) \quad \text{subject to} \quad h_k(w) = 0; \quad \forall 1 \leq k \leq K$$

Learning with Constraints: *Reduce # Constraints*

Originally:

$$\min_w L(w) \quad \text{subject to} \quad f_k^i(w) \leq 0; \quad \forall 1 \leq i \leq m; \quad \forall 1 \leq k \leq K$$

Now:

Define: $h_k(w) = \sum_i H(f_k^i(w))$ $O(K)$ #constraints

$$\min_w L(w) \quad \text{subject to} \quad h_k(w) = 0; \quad \forall 1 \leq k \leq K$$

Learning with Constraints: *Reduce # Constraints*

$$\min_w L(w) \quad \text{subject to} \quad h_k(w) = 0; \quad \forall 1 \leq k \leq K$$

Lagrangian

$$\mathcal{L}(w; \Lambda) = L(w) + \sum_{k=1}^K \lambda_k h_k(w)$$

Learning with Constraints:

Experiments: Typenet

	MAP Scores			Constraint Violations		
Scenario	5% Data	10% Data	100% Data	5% Data	10% Data	100% Data
B	68.6			22,715		
B+H	68.71			22,928		
B+C						
B+S						

Learning with Constraints:

Experiments: Typenet

	MAP Scores			Constraint Violations		
Scenario	5% Data	10% Data	100% Data	5% Data	10% Data	100% Data
B	68.6			22,715		
B+H	68.71			22,928		
B+C	80.13			25		
B+S	82.22			41		

Learning with Constraints:

Experiments: Typenet

	MAP Scores			Constraint Violations		
Scenario	5% Data	10% Data	100% Data	5% Data	10% Data	100% Data
B	68.6	69.2	70.5	22,715	21,451	22,359
B+H	68.71	69.31	71.77	22,928	21,157	24,650
B+C	80.13	81.36	82.80	25	45	12
B+S	82.22	83.81		41	26	