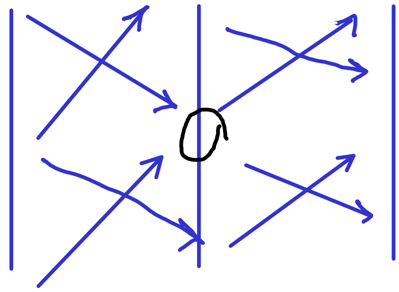
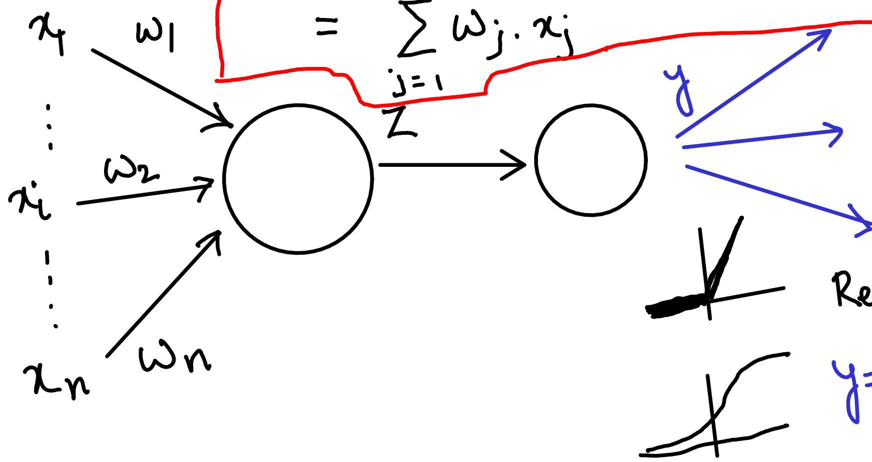
CS 620: FM in ML
Week 5: Modeling Neural Networks

Supratik Chakraborty

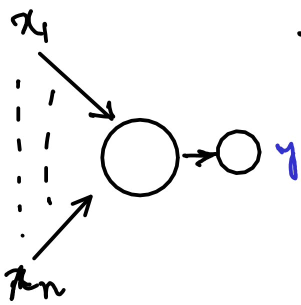


$$\boxed{\begin{aligned} Z &= w_1 \cdot x_1 + \dots + w_i \cdot x_i + \dots w_n \cdot x_n \\ &= \sum_{j=1}^{n} w_j \cdot x_j \end{aligned}}$$



$y = f(z)$

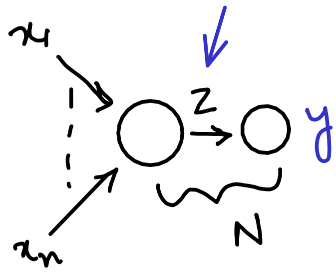
$y = \max(0, z)$

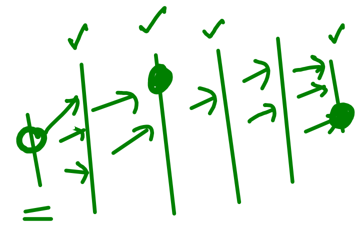


$$z = \sum_{j=1}^{n} w_j \cdot x_j$$

$$y = \max(0, z)$$

$$\exists z . \left(z = \sum_{j=1}^{n} w_j \cdot x_j \right) \wedge \left(y = \max(0, z) \right) \equiv \varphi_N(x_1, \ldots x_n, y)$$





y^{in}

Model of a single node

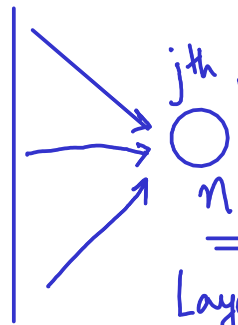
In

in_1

$\vdots$

in_K

Hidden layers



j^{th} node in layer i

$\underline{\underline{n_{i,j}}}$

Layer i

Output

out_1

$\vdots$

out_r

$$\text{pred}(n_{i,j}) = \left\{ n_{i-1,j'} \;\middle|\; \begin{array}{l} (n_{i-1,j'}, n_{i,j}) \\ \text{is an edge in the NN} \\ \text{(underlying graph)} \end{array} \right\}$$

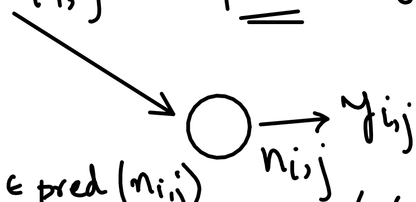
$$\varphi_{n_{i,j}}(x_1, \ldots, x_s, y_{i,j})$$

$$\begin{array}{ccc} x_1 & \cdots & x_s \\ \downarrow & & \downarrow \\ x_{1,i,j} & \cdots & x_{\underline{\underline{s}},\underline{\underline{i}},\underline{\underline{j}}} \end{array}$$

$$\bigwedge_{n_{i,j} \in \text{Hidden layers}} \varphi_{n_{i,j}} \;\wedge\; \bigwedge_{in_j \in In} \varphi_{in_j} \;\wedge$$

$$\varphi_N \equiv \bigwedge_{n_{i,j} \in \text{Hidden layers}} \varphi_{n_{i,j}} \quad \wedge \quad \bigwedge_{out_j \in \text{Out Layer}} \varphi_{out_j}$$

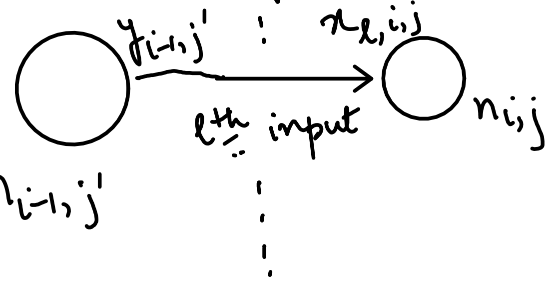
$n_{i-1,j'} \in \underline{\underline{pred}}(n_{i,j})$



$$\begin{pmatrix} x_{1,i,j} = y_{i-1,j'} & \cdots & n_{i-1,j'} \in pred(n_{i,j}) \\ x_{2,i,j} = y_{i-1,j''} & \cdots & n_{i-1,j''} \in pred(n_{i,j}) \end{pmatrix}$$

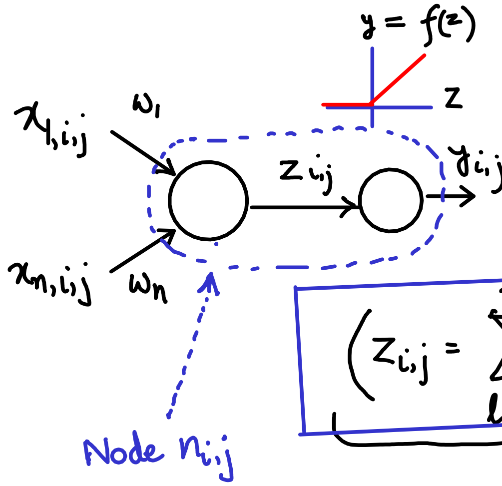
$$\varphi_{n_{i,j}} = (\underline{\underline{x_{1,i,j}}}, \ldots, \underline{\underline{x_{s,i,j}}}, \underline{\underline{y_{i,j}}})$$

$$\varphi_E \equiv \bigwedge_{n_{i,j} \in \text{Hidden layers} \,\cup\, \text{Output Layer}} \;\; \bigwedge_{\substack{n_{i-1,j'} \in \text{pred}(n_{i,j}) \\ (\ell^{th}\ \text{pred})}} \left(y_{i-1,j'} = x_{\ell,i,j} \right)$$



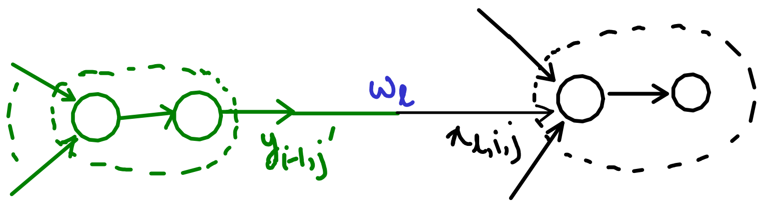
Recap:

Modeling nodes



$$\underbrace{\boxed{\left(z_{i,j} = \sum_{l=1}^{n} \omega_l \cdot x_{l,i,j}\right) \wedge \left(y_{i,j} = f(z_{i,j})\right)}}_{\varphi_{n_{i,j}}\left(x_{1,i,j},\, \text{----}\, x_{n,i,j},\, z_{i,j},\, y_{i,j}\right)}$$

Recap: Modeling edges



Node $n_{i-1,j'} \in \text{preds}(\text{Node } n_{i,j})$

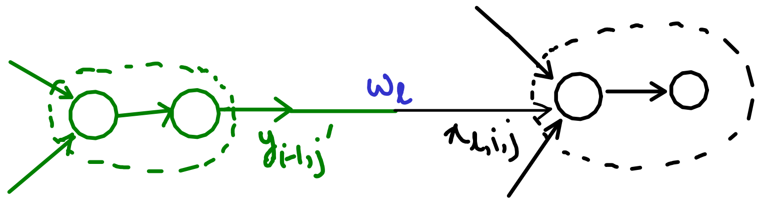
Edge $e = (n_{i-1,j'}, n_{i,j})$: $\boxed{y_{i-1,j'} = x_{\ell,i,j}} \equiv \Psi_e$

Putting Things Together

$$\left(\bigwedge_{n_{i,j} \in \text{Hidden Layers} \,\cup\, \text{Output Layers}} \varphi_{n_{i,j}} \;\wedge\; \bigwedge_{e \in \text{Edges}} \psi_e \right) \equiv \chi_{NN}$$

How many constraints?

How many free vars?



Putting Things Together

$$\left(\bigwedge_{n_{i,j} \in \text{Hidden Layers} \cup \text{Output layers}} \varphi_{n_{i,j}} \quad \wedge \quad \bigwedge_{e \in \text{Edges}} \psi_e \right) \equiv \chi_{NN}$$

\# nodes + \# edges

Constraints

How many constraints?

How many free vars?

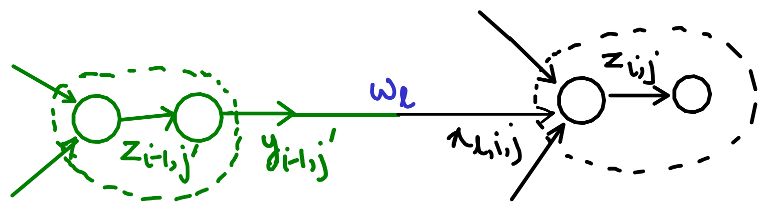
Putting Things Together

$$\left(\bigwedge_{\substack{n_{i,j} \in \text{Hidden Layers} \\ \cup \\ \text{Output Layers}}} \varphi_{n_{i,j}} \;\wedge\; \bigwedge_{e \in \text{Edges}} \psi_e \right) \equiv \chi_{NN}$$

How many constraints?

How many free vars?

Things blow-up quickly



2 x # edges + # nodes in hidden + output layers = # free vars in χ_{NN}

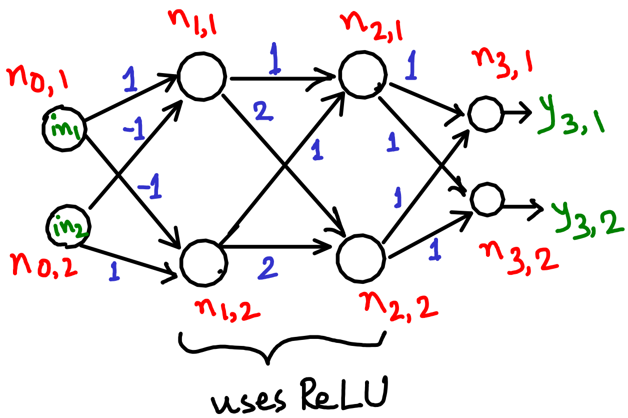
Some Perspective

Fully connected DNN

100 hidden layers + 1 output layer

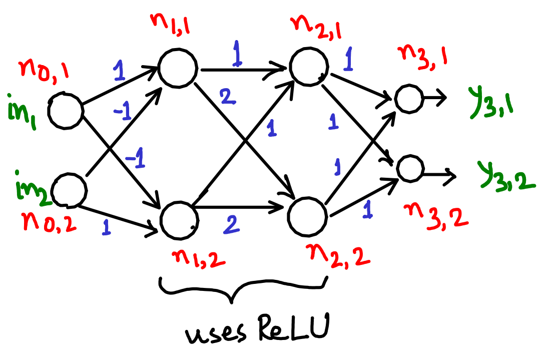
100 nodes / hidden layer + 10 nodes in output layer.

An Example



$$
\begin{aligned}
n_{1,1} &: \left(z_{1,1} = x_{1,1,1} - x_{2,1,1}\right) \wedge \\
&\quad \left(y_{1,1} = \max\left(0, z_{1,1}\right)\right) \wedge \\
&\vdots \\
n_{3,2} &: \text{------} \wedge \\
e_{0,1\to 1,1} &: \left(in_1 = x_{1,1,1}\right) \wedge \\
e_{0,1\to 1,2} &: \left(in_1 = x_{1,1,2}\right) \wedge \\
&\vdots \\
e_{2,2\to 3,2} &: \left(y_{2,2} = x_{2,3,2}\right)
\end{aligned}
$$

An Example



$$
X_{NN} \equiv
\begin{array}{ll}
n_{1,1}: & (z_{1,1} = x_{1,1,1} - x_{2,1,1}) \wedge \\
 & (y_{1,1} = \max(0, z_{1,1})) \wedge \\
\vdots & \qquad \vdots \\
n_{3,2}: & \text{------} \wedge \\
e_{0,1\to 1,1}: & (in_1 = x_{1,1,1}) \wedge \\
e_{0,1\to 1,2}: & (in_1 = x_{1,1,2}) \wedge \\
\vdots & \qquad \vdots \\
e_{2,2\to 3,2}: & (y_{2,2} = x_{2,3,2})
\end{array}
$$

What if I wanted a formula with free vars $in_1, in_2, y_{3,1}, y_{3,2}$?

$\hat{X}_{NN}(in_1, in_2, y_{3,1}, y_{3,2}) = ?$

Do we really need to existentially quantify?

$$\{\, Pre(in_1, in_2) \,\}$$

$$\boxed{\begin{array}{l} x_{NN} \\ \quad \text{or} \\ \hat{x}_{NN} = \exists \ldots\ x_{NN} \end{array}}$$

$$\{\, Post(y_{3,1}, y_{3,2}) \,\}$$

$$Pre(in_1, in_2) \wedge \boxed{}$$

$$\Rightarrow\ ?$$

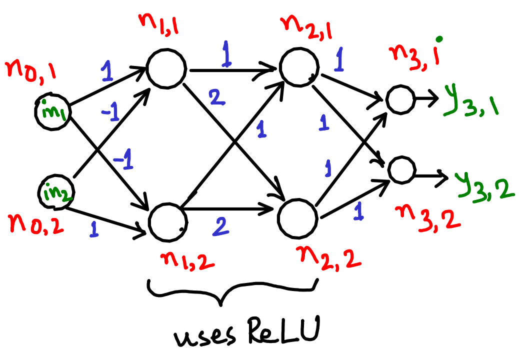
$$Post(y_{3,1}, y_{3,2})$$

Relational vs Functional Model

Functional: $(y_{3,1}, y_{3,2}) = \mathcal{V}(in_1, in_2)$

Relational:

$$\left[\begin{array}{l} X_{NN}(in_1, in_2, x_{1,1,1}, \ldots z_{3,2}, y_{3,1}, y_{3,2}) \\ \widehat{X}_{NN}(in_1, in_2, y_{3,1}, y_{3,2}) = \exists \ldots X_{NN} \end{array}\right.$$

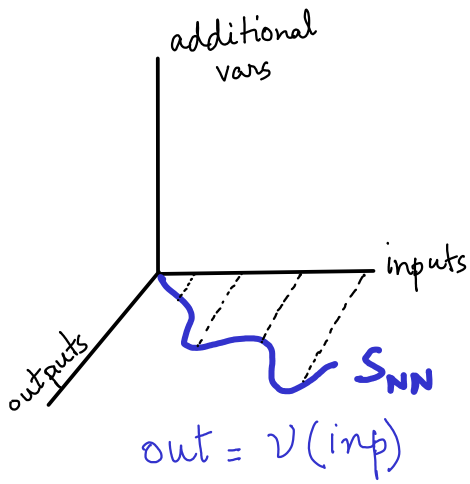


$$S_{\hat{X}_{NN}} = \{(a,b,c,d) \mid (a,b,c,d) \models \hat{X}_{NN}\}$$

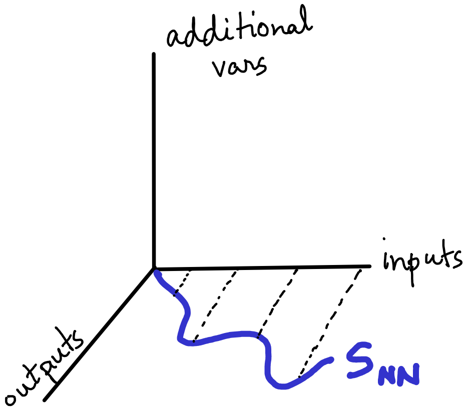
$$S_{X_{NN}} = \{(a,b,\ldots,c,d) \mid (a,b,\ldots,c,d) \models X_{NN}\}$$

$$S_{NN} = \{(a,b,c,d) \mid (a,b) = \mathcal{V}(c,d)\}$$

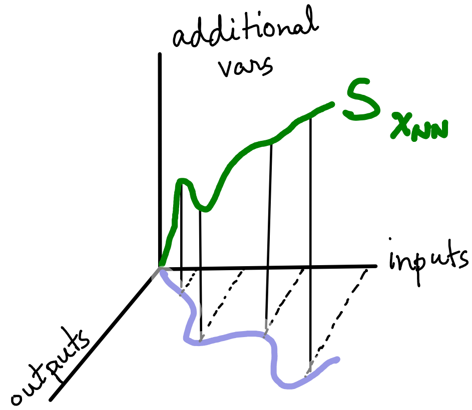
Relational vs Functional Model



Relational vs Functional Model

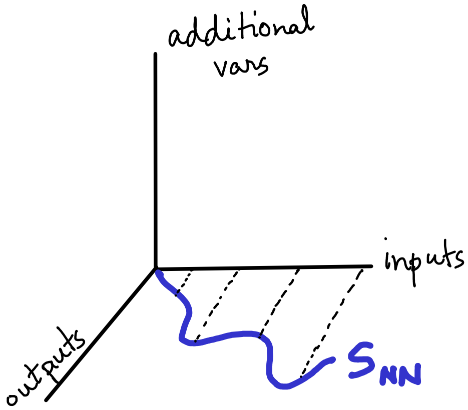


$out = \nu(inp)$

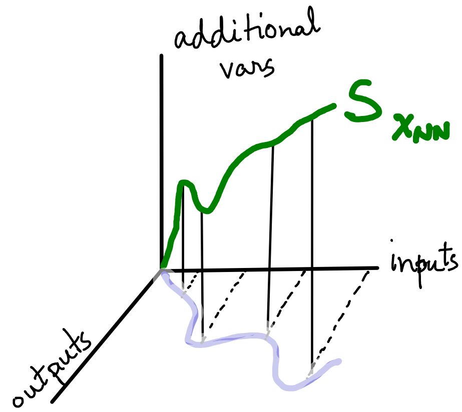


$X_{NN}(inp, \text{addnl vars}, out)$

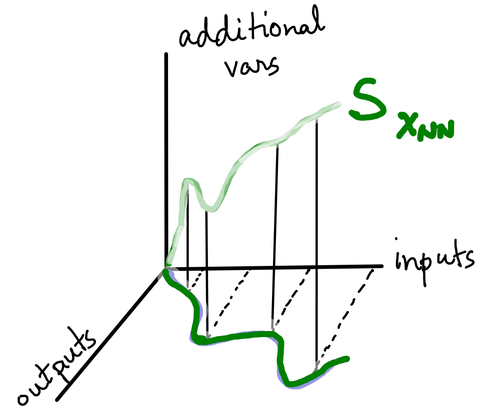
Relational vs Functional Model



$out = \nu(inp)$



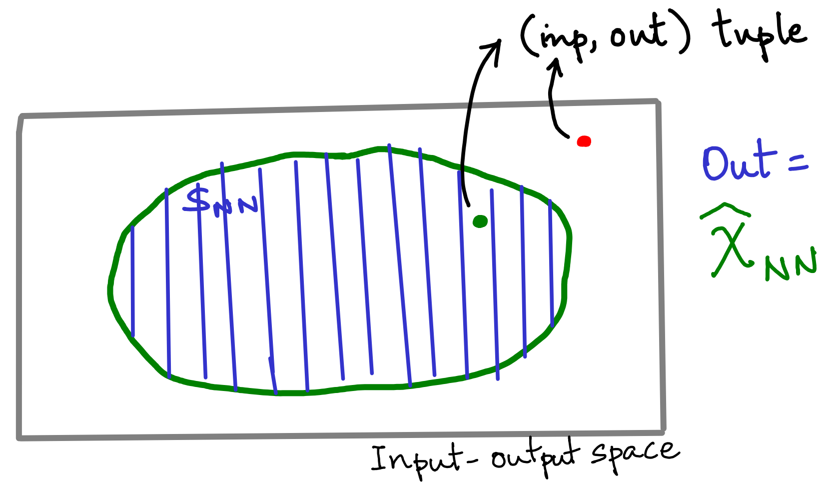
$\chi_{NN}(inp,\ addnl\ vars,\ out)$



$\hat{\chi}_{NN}(inp,\ outp) = \exists\, addnl\ vars\ \chi(in,\ a.v,\ out)$

Relational vs Functional Model

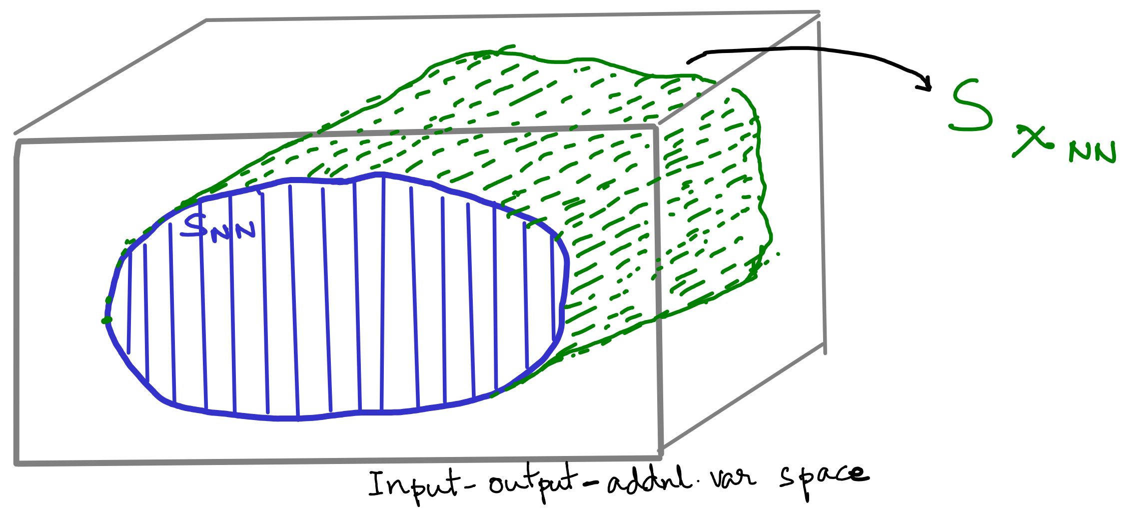
Another view



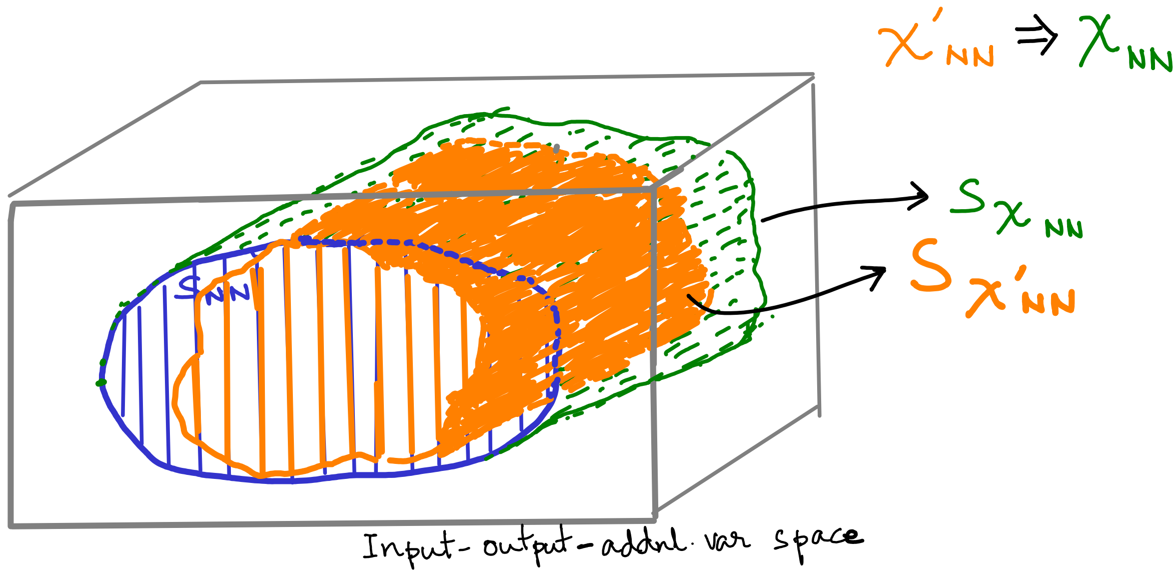
$$Out = \nu(inp)$$

$$\widehat{X}_{NN}(inp, out)$$

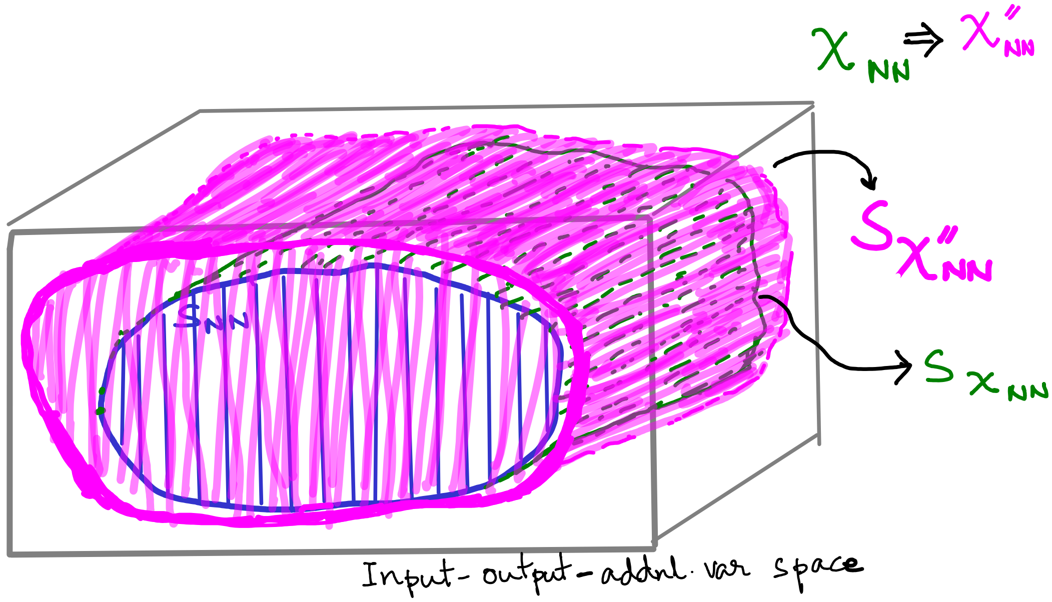
Relational vs Functional Model



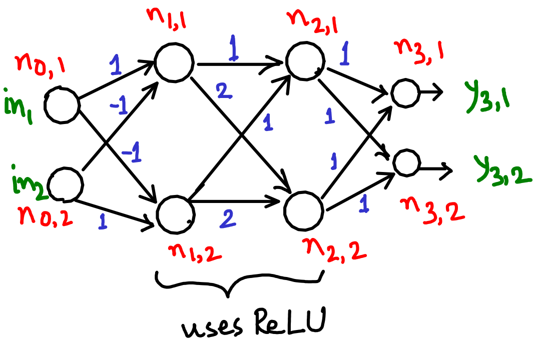
Approximate Relational Models



Approximate Relational Models

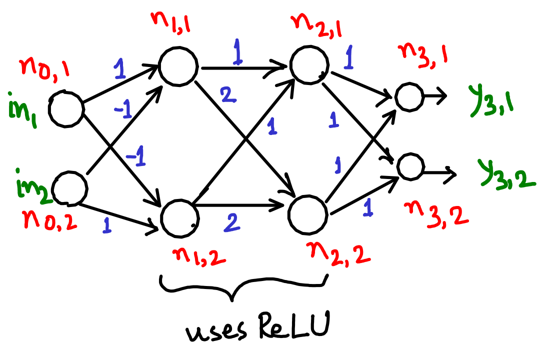


Quick Recap:



$$
\mathcal{X}_{NN} \left[\begin{array}{l}
(z_{1,1} = x_{1,1,1} - x_{2,1,1}) \wedge \\
(y_{1,1} = \max(0, z_{1,1})) \wedge \\
\quad \vdots \qquad\qquad\qquad\qquad \vdots \\
\quad \cdots \qquad\qquad\qquad\qquad \wedge \\
\quad (in_1 = x_{1,1,1}) \wedge \\
\quad (in_1 = x_{1,1,2}) \wedge \\
\quad \vdots \qquad\qquad\qquad\qquad \vdots \\
\quad (y_{2,2} = x_{2,3,2})
\end{array} \right.
$$

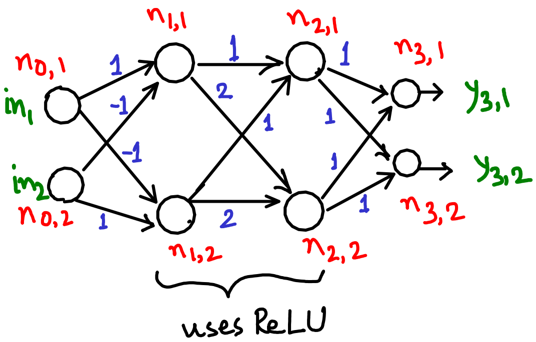
Quick Recap: Approximations



$$\chi'_{NN} \Downarrow \chi_{NN}$$

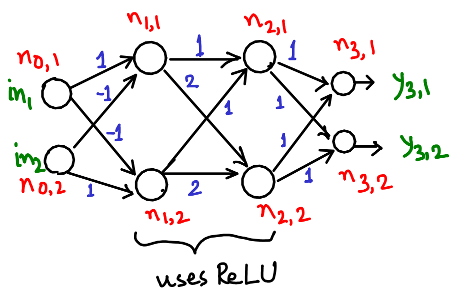
$$\left[\begin{array}{l}
(z_{1,1} = x_{1,1,1} - x_{2,1,1}) \wedge \\
\cancel{(y_{1,1} = \max(0, z_{1,1}))} \wedge \\
(z_{1,1} \geq 0) \wedge (y_{1,1} = z_{1,1}) \\
\vdots \\
\text{-------} \wedge \\
(in_1 = x_{1,1,1}) \wedge \\
(in_1 = x_{1,1,2}) \wedge \\
\vdots \\
(y_{2,2} = x_{2,3,2})
\end{array}\right.$$

Quick Recap: Approximations



$$
\begin{array}{c}
x''_{NN} \\
\Uparrow \\
X_{NN}
\end{array}
\left[
\begin{array}{l}
(Z_{1,1} = x_{1,1,1} - x_{2,1,1}) \wedge \\
\cancel{(y_{1,1} = \max(0, Z_{1,1}))} \wedge \\
(y_{1,1} \geq 0) \wedge (y_{1,1} \geq Z_{1,1}) \\
\quad \cdots \wedge \\
\quad (in_1 = x_{1,1,1}) \wedge \\
\quad (in_1 = x_{1,1,2}) \wedge \\
\quad \vdots \\
\quad (y_{2,2} = x_{2,3,2})
\end{array}
\right.
$$

Why Approximate ?



$$(z_{1,1} = x_{1,1,1} - x_{2,1,1}) \wedge$$

$$\boxed{(y_{1,1} = \max(0, z_{1,1}))} \wedge$$

$$\vdots$$

$$(in_1 = x_{1,1,1}) \wedge$$

$$(in_1 = x_{1,1,2}) \wedge$$

$$\vdots$$

$$(y_{2,2} = x_{2,3,2})$$

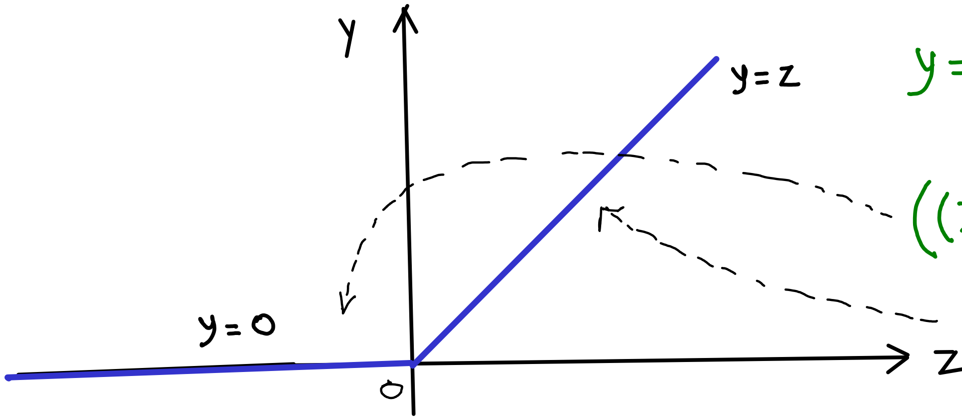
Non-linear (non-convex)

$\approx$

computationally harder to reason than linear constr.

How To Approximate ?

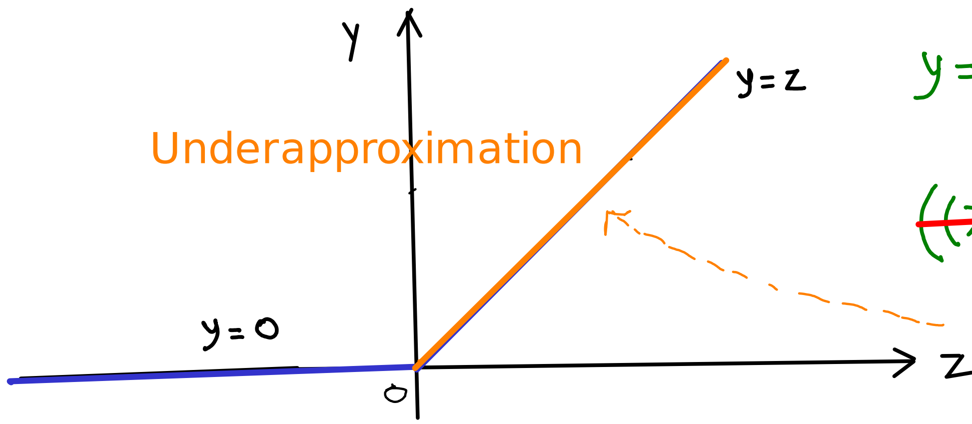
Can we use linear approximation of non-linear constraints?



$$y = \max(0, z)$$
$$\iff$$
$$((z \le 0) \wedge (y = 0))$$
$$\vee$$
$$((z > 0) \wedge (y = z))$$

How To Approximate ?

Can we use linear approximation of non-linear constraints?



$$y = \max(0, z)$$

$$\Updownarrow$$

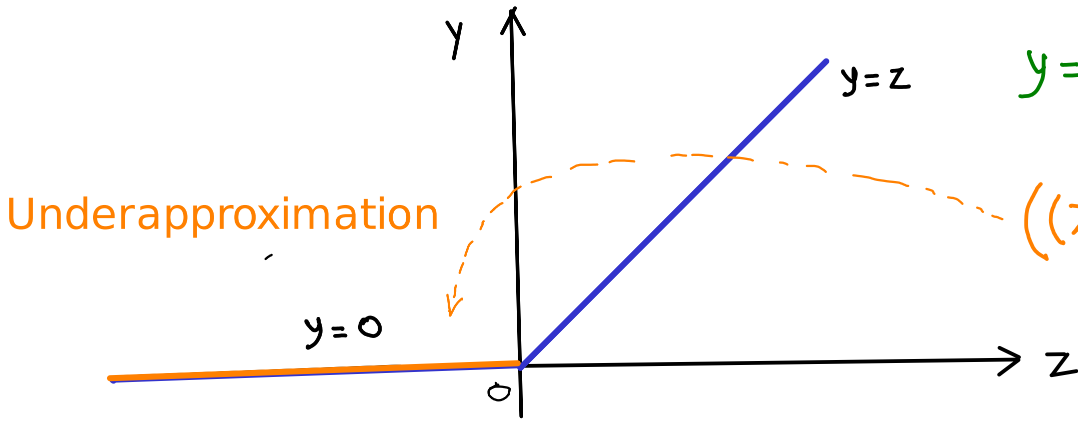
$$\cancel{((z \leq 0) \wedge (y = 0))}$$

$$\vee$$

$$((z > 0) \wedge (y = z))$$

How To Approximate ?

Can we use linear approximation of non-linear constraints?



$$y = \max(0, z)$$

$$\iff$$

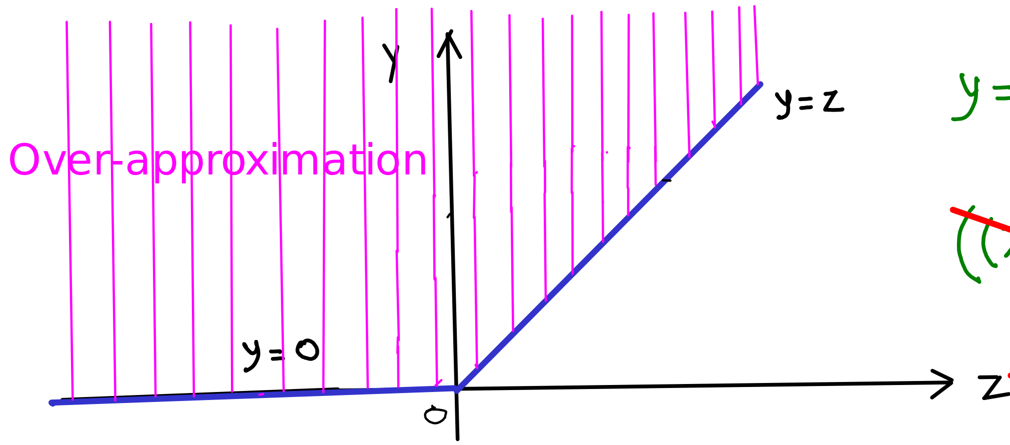
$$((z \leq 0) \wedge (y = 0))$$

$$\cancel{\vee}$$

$$\cancel{((z > 0) \wedge (y = z))}$$

How To Approximate ?

Can we use linear approximation of non-linear constraints?



$y = \max(0, z)$

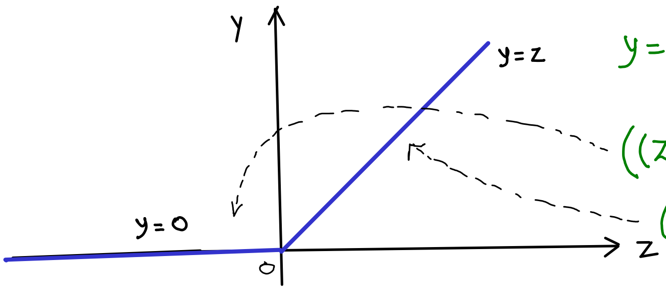
$\Longleftrightarrow$

$\cancel{((z \leq 0) \wedge (y = 0)) \vee ((z > 0) \wedge (y = z))}$

$(y \geq 0) \wedge (y \geq z)$

Piece-wise linear approximation

Recall:



$y = \max(0, z)$

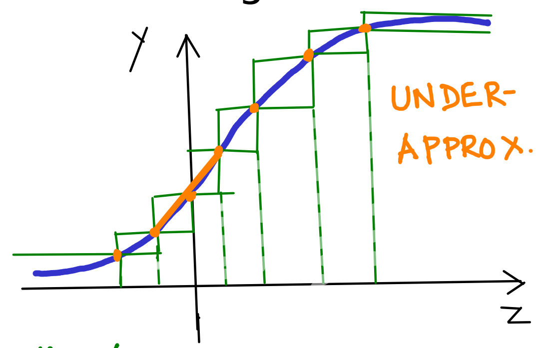
$\Leftrightarrow$

$((z \le 0) \wedge (y = 0))$

$\vee$

$((z > 0) \wedge (y = z))$

Generalizing:

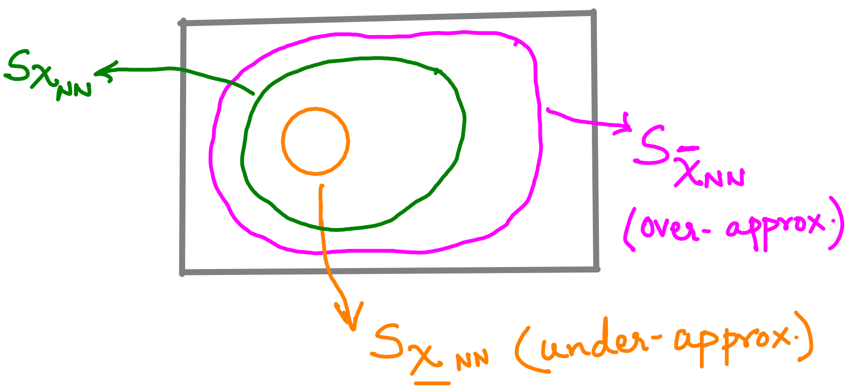


$$\bigvee_{i=1}^{n} \big((a_i \le z < b_i) \wedge (c_i \le y < d_i)\big)$$

OVER-APPROX

Approximations in Modeling

When to use which approximation?



$\{Pre(inp)\}$

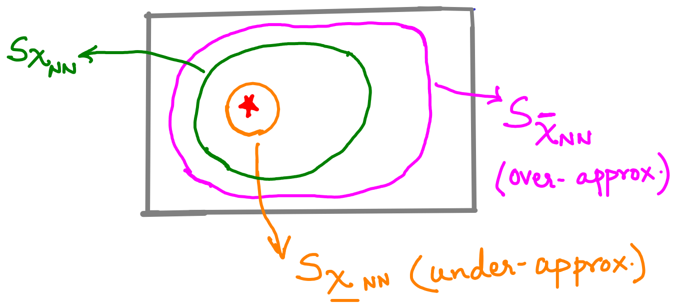
$$\boxed{x_{NN} \text{ or } \underline{x}_{NN} \text{ or } \bar{x}_{NN} \text{ ???}}$$

$\{Post(out)\}$

$$\underline{x}_{NN} \Rightarrow x_{NN} \Rightarrow \bar{x}_{NN}$$

Approximations in Modeling

When to use which approximation?



$\{ Pre(inp) \}$

$\boxed{X_{NN} \text{ or } \underline{X}_{NN} \text{ or } \overline{X}_{NN} \text{ ???}}$

$\{ Post(out) \}$

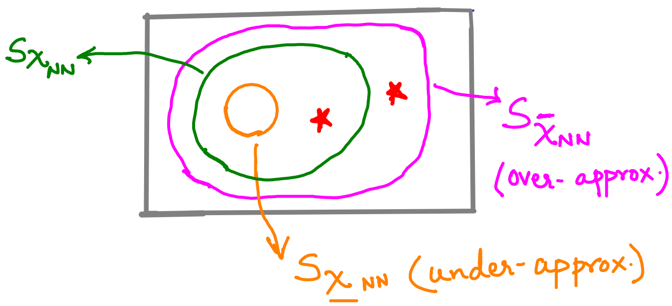
$Pre \wedge \underline{X}_{NN} \wedge \neg Post$ satisfiable

NECESSARILY IMPLIES

$Pre \wedge X_{NN} \wedge \neg Post$ satisfiable.

Approximations in Modeling

When to use which approximation?



$\{Pre(inp)\}$

$\boxed{x_{NN} \text{ or } \underline{x}_{NN} \text{ or } \bar{x}_{NN} \text{ ???}}$

$\{Post(out)\}$

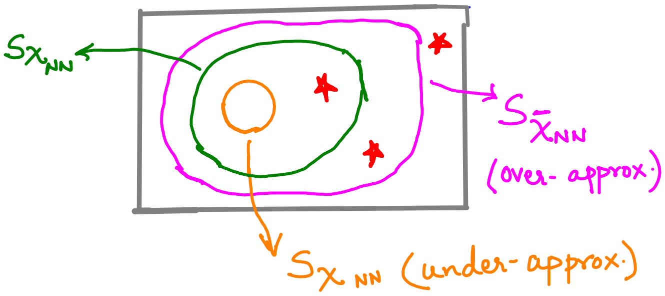
$Pre \wedge \bar{x}_{NN} \wedge \neg Post$ satisfiable

DOES NOT IMPLY

$Pre \wedge x_{NN} \wedge \neg Post$ satisfiable

Approximations in Modeling

When to use which approximation?



$\{ Pre(inp) \}$

$$\boxed{\begin{array}{c} \chi_{NN} \text{ or } \underline{\chi}_{NN} \\ \text{or} \\ \overline{\chi}_{NN} \text{ ???} \end{array}}$$

$\{ Post(out) \}$

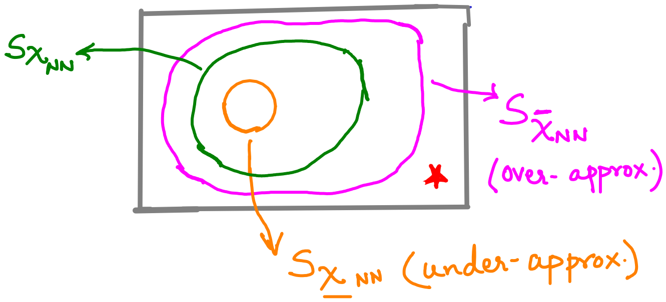
$Pe \wedge \underline{\chi}_{NN} \Rightarrow Post$

DOES NOT IMPLY

$Pre \wedge \chi_{NN} \Rightarrow Post$

Approximations in Modeling

When to use which approximation?



$\{ Pre(inp) \}$

$\boxed{x_{NN} \text{ or } \underline{x}_{NN} \text{ or } \bar{x}_{NN} \text{ ???}}$

$\{ Post(out) \}$

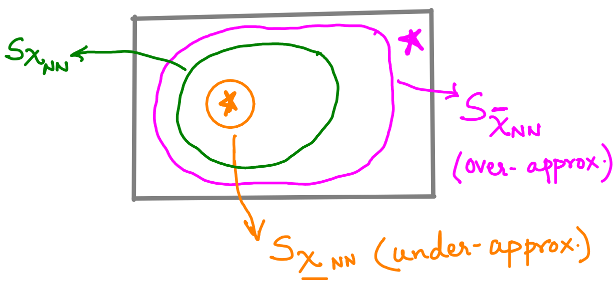
$Pe \wedge \bar{X}_{NN} \Rightarrow Post$

NECESSARILY IMPLIES

$Pre \wedge X_{NN} \Rightarrow Post$

Approximations in Modeling

When to use which approximation?



$\{Pre(inp)\}$

$\boxed{\mathcal{X}_{NN} \text{ or } \underline{\mathcal{X}}_{NN} \text{ or } \overline{\mathcal{X}}_{NN} \text{ ???}}$

$\{Post(out)\}$

$\underline{\mathcal{X}}_{NN} \Rightarrow \mathcal{X}_{NN} \Rightarrow \overline{\mathcal{X}}_{NN}$

$\underline{\mathcal{X}}_{NN}$ good for bug-hunting

$\overline{\mathcal{X}}_{NN}$ good for proving absence of bugs