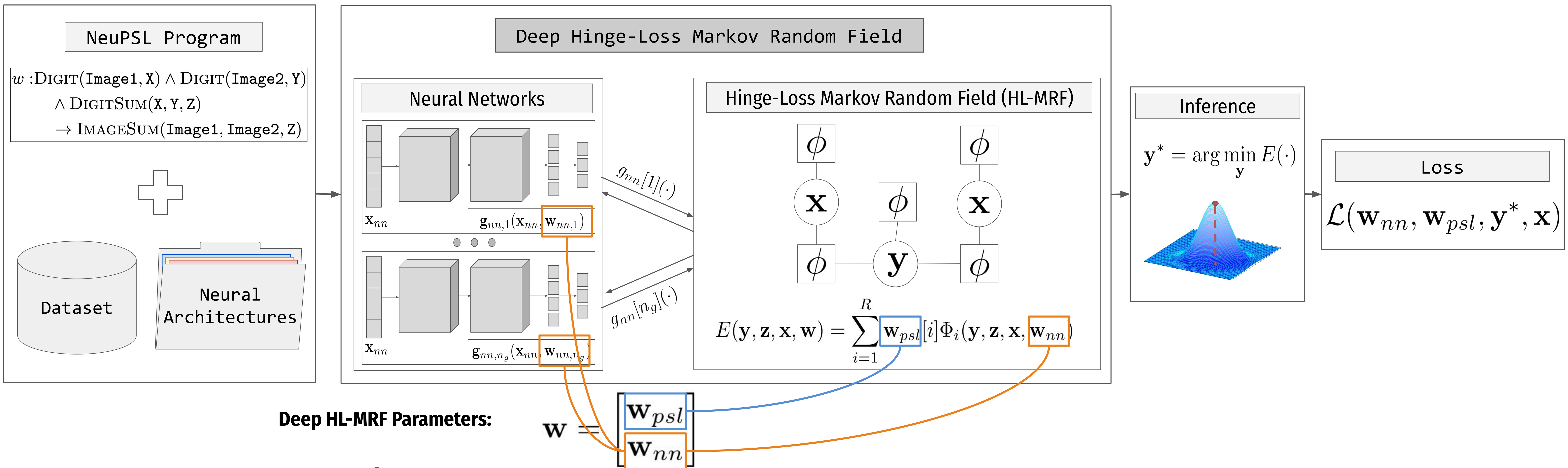
Efficient Learning Losses for Deep Hinge-Loss Markov Random Fields

Charles Dickens, Connor Pryor, Eriq Augustine, Alon Albalak, & Lise Getoor
University of California, Santa Cruz

Deep Hinge-Loss Markov Random Fields

Deep Hinge-Loss Markov Random Fields (Deep-HLMRFs)[1] are tractable probabilistic graphical models that integrate low-level neural perception with symbolic reasoning.



Learning

Training Examples:

$$\mathcal{S} = \{(\mathbf{y}_i, \mathbf{x}_i, \mathbf{x}_{i,nn}) : i = 1, \cdots, P\}$$

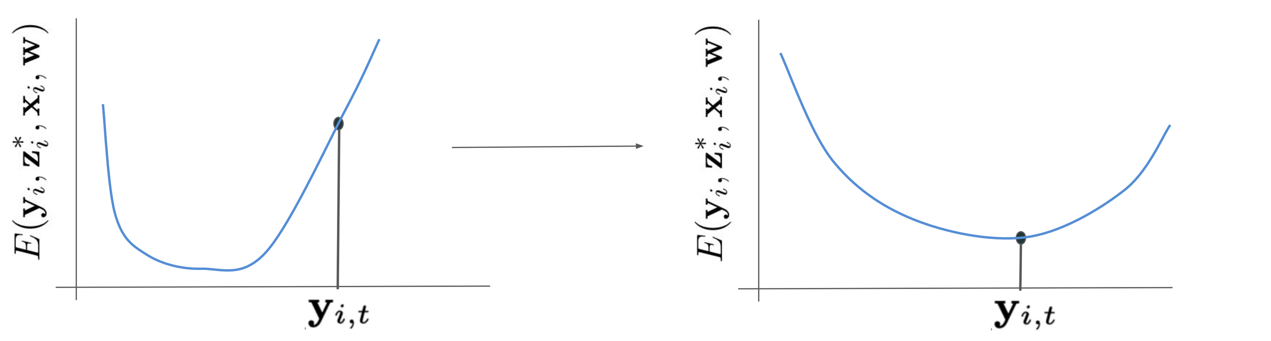
Latent Variables:

$$\mathbf{z}_i^* = \arg\min_{\mathbf{z} | ((\mathbf{y}_{i,t}, \mathbf{z}), \mathbf{x}) \in \Omega} E((\mathbf{y}_{i,t}, \mathbf{z}), \mathbf{x}_i, \mathbf{x}_{i,nn}, \mathbf{w}_{nn}, \mathbf{w}_{psl})$$

MAP State / Prediction:

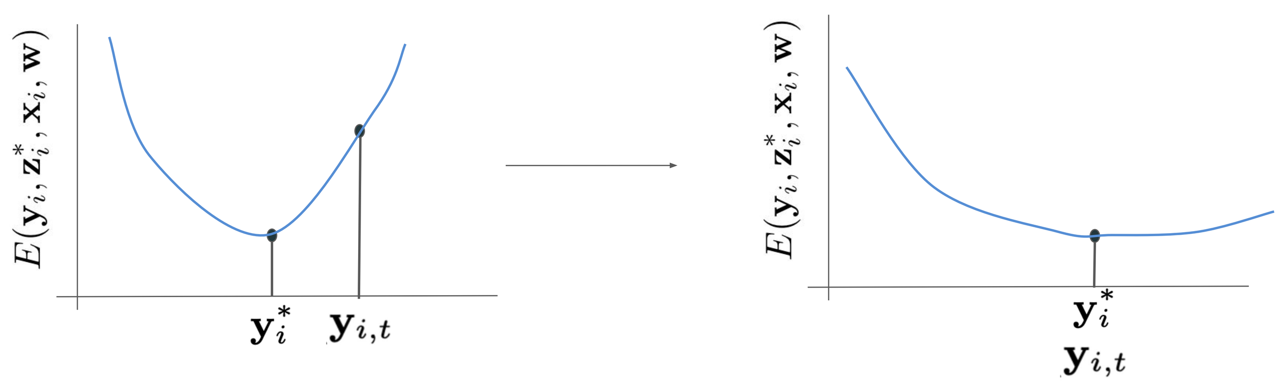
$$\mathbf{y}_i^* = \arg\min_{\mathbf{y} | (\mathbf{y}, \mathbf{x}_i) \in \Omega} E(\mathbf{y}, \mathbf{x}_i, \mathbf{x}_{i,nn}, \mathbf{w}_{nn}, \mathbf{w}_{psl})$$

Energy Loss



$$\arg\min_{\mathbf{w}_{nn}, \mathbf{w}_{psl}} \mathcal{L}_{Energy}(\mathbf{w}_{nn}, \mathbf{w}_{psl}, \mathcal{S}) = \arg\min_{\mathbf{w}_{nn}, \mathbf{w}_{psl}} \sum_{i=1}^{P} E((\mathbf{y}_{i,t}, \mathbf{z}_i^*), \mathbf{x}_i, \mathbf{x}_{i,nn}, \mathbf{w}_{nn}, \mathbf{w}_{psl})$$

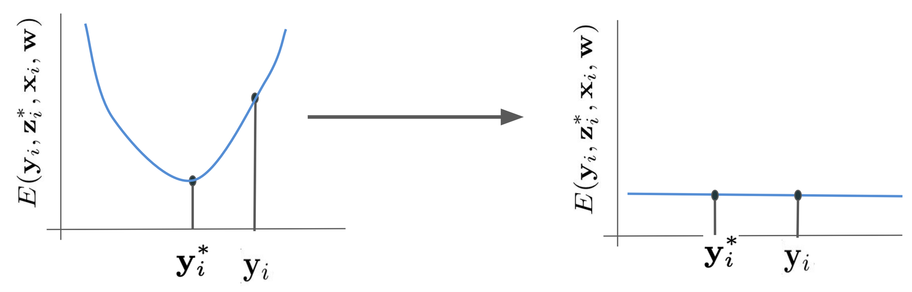
Structured Perceptron Loss



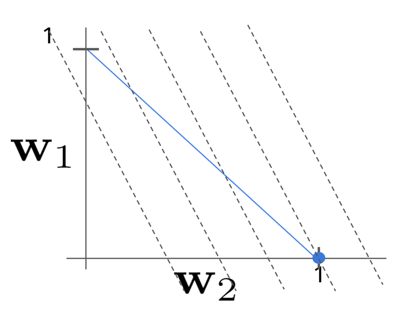
$$\arg\min_{\mathbf{w}_{nn}, \mathbf{w}_{psl}} \mathcal{L}_{SP}(\mathbf{w}_{nn}, \mathbf{w}_{psl}, \mathcal{S}) = \arg\min_{\mathbf{w}_{nn}, \mathbf{w}_{psl}} \sum_{i=1}^{P} E((\mathbf{y}_{i,t}, \mathbf{z}_i^*), \mathbf{x}_i, \mathbf{x}_{i,nn}, \mathbf{w}_{psl}, \mathbf{w}_{nn}) - E(\mathbf{y}_i^*, \mathbf{x}_i, \mathbf{x}_{i,nn}, \mathbf{w}_{psl}, \mathbf{w}_{nn})$$

Degenerate solutions

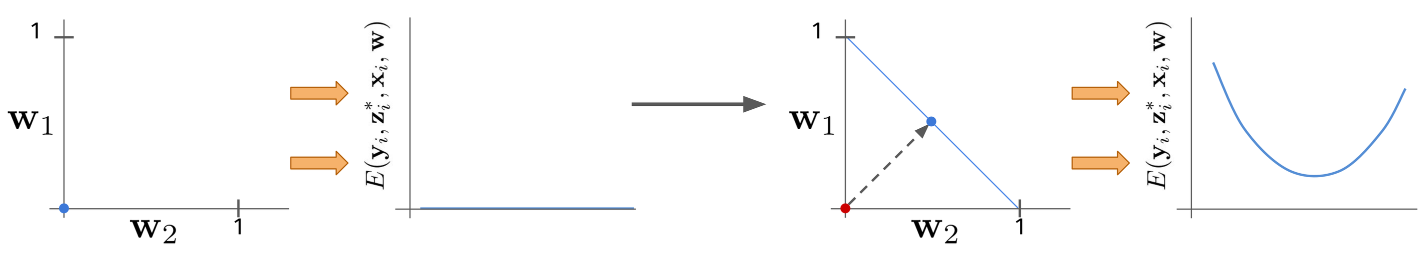
Direct optimization of losses leads collapsed energy functions



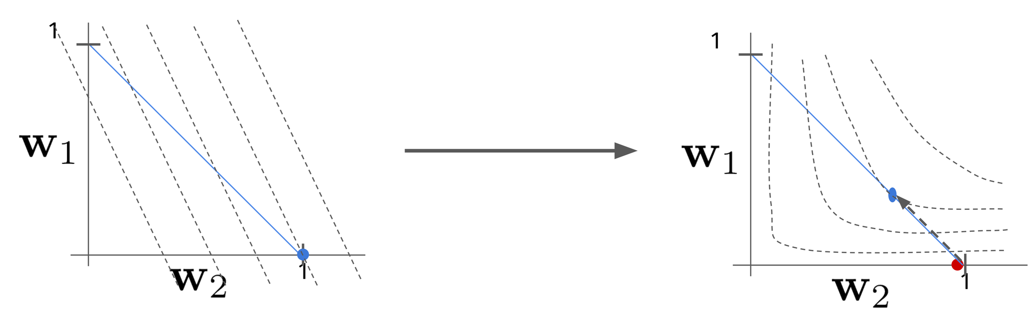
Simplex Constrained Parameters Leads to Corner Solutions



Restrict weights to be on the simplex.



Add regularization to the energy loss pushing weights away from 0



Regularized and Simplex Constrained Loss Function:

$$\min_{\mathbf{w} \in \mathcal{W}} \mathcal{L}(\mathbf{w}, \mathcal{S}) - \sum_{i=1}^{R} \log_b(\mathbf{w}[i])$$

$$\text{s.t.} \quad \mathbf{w}_{psl} \in \Delta^R$$

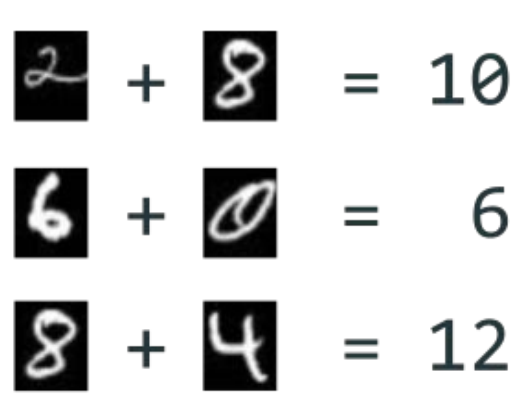
Mirror Descent Updates

$$\mathbf{w}_{nn}^{k+1} = \mathbf{w}_{nn}^{k} + \eta \nabla_{w_{nn}} \mathcal{L}(\mathbf{w}_{nn}^{k}, \mathbf{w}_{psl}^{k}, S)$$

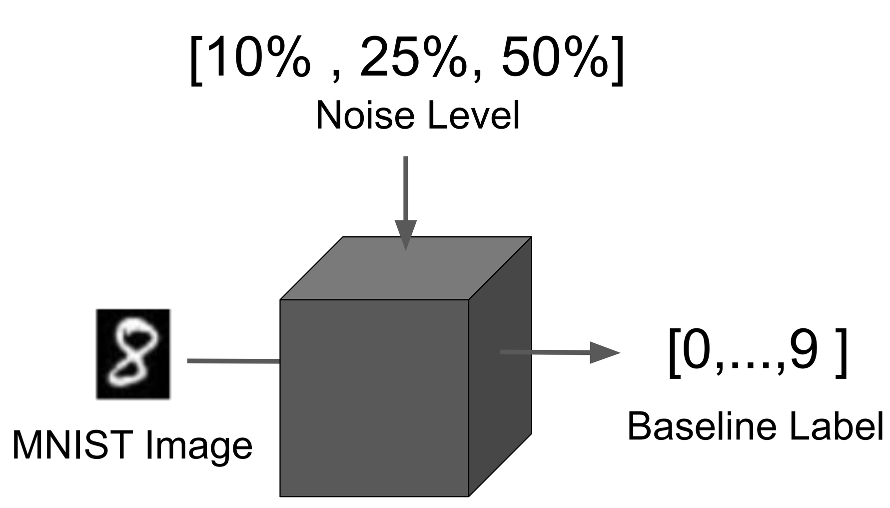
$$\mathbf{w}_{psl}^{k+1}[i] = \frac{\mathbf{w}_{psl}^{k}[i] \exp\{-\eta \frac{\partial \mathcal{L}(\mathbf{w}_{nn}^{k}, \mathbf{w}_{psl}^{k}, S)}{\partial \mathbf{w}_{psl}^{k}[i]}\}}{\sum_{j=1}^{r} \exp\{-\frac{\partial \mathcal{L}(\mathbf{w}_{nn}^{k}, \mathbf{w}_{psl}^{k}, S)}{\partial \mathbf{w}_{psl}^{k}[j]}\}}$$

Empirical Evaluation

MNIST Addition [2]



Independent Baseline and Noise



Model	Learning Method	Noise (%)	Accuracy	Runtime (sec)
NeuPSL	Energy	10	77.1 ± 2.5	120.3 ± 0.7
	Energy	25	75.3 ± 4.6	120.3 ± 0.5
	Energy	50	70.4 ± 4.2	121.2 ± 0.7
	Structured Perceptron	10	71.2 ± 3.9	266.5 ± 2.0
	Structured Perceptron	25	72.0 ± 4.6	281.3 ± 2.4
	Structured Perceptron	50	75.1 ± 3.8	289.9 ± 2.5
Independent Baseline	-	10	81.8 ± 2.5	-
	-	25	59.0 ± 2.7	-
	-	50	30.1 ± 2.5	-

Energy loss learning can achieve higher accuracy in roughly half the runtime of structured perceptron.

Conclusion

Energy and structured perceptron learning losses were presented for Deep-HLMRFs. Degenerate solutions of the losses were identified and we proposed constraints, regularizations, and a tractable optimization technique to overcome them. The performance of learning losses was tested on a canonical NeSY dataset and we found, surprisingly, training with the simpler energy loss can achieve higher accuracy in roughly half the runtime of structured perceptron.

This work was partially supported by the NSF grants CCF-1740850,CCF-2023495, and an unrestricted gift from Google

[1] Pryor, Dickens, Augustine, Albalak, Wang, & Getoor. *NeuPSL: Neural Probabilistic Soft Logic*. arXiv (2022).
[2] Manhaeve, Dumancic, Kimmig, Demeester, and De Raedt. *Deepproblog: Neural probabilistic logic programming*. NeurIPS(2018).