Proximal Deep Structured Models

Shenlong Wang, Sanja Fidler, & Raquel Urtasun

University of Toronto

NIPS 2016
Presenter: Jack Lanchantin

Outline

1. Intro

Continuous-Valued Deep Structured Models

Given input $x \in \mathcal{X}$, let $y = (y_1, ..., y_n)$ be the set of random variables we want to predict. The output spacie is a product space of all the elements $y \in \mathcal{Y} = \prod_{i=1}^{N} \mathcal{Y}_i$. Let $E(x, y; w) : \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^K \rightarrow \mathbb{R}$ be an energy function which encodes the problem.

$$E(x, y; w) = \sum_i f_i(y_i, x; w_u) + \sum_\alpha f_\alpha(y_\alpha, x, w_\alpha) \tag{1}$$

$$y^* = argmin_{y \in \mathcal{Y}} \sum_i f_\alpha(y_i; x, w_u) + \sum_\alpha f_\alpha(y_\alpha, x, w_\alpha) \tag{2}$$

where $f_i(y_i : x, w_u) : \mathcal{Y}_i \times \mathcal{X} \rightarrow \mathbb{R}$ is a function that depends on a single variable and $f_\alpha(y_i) : \mathcal{Y}_\alpha \times \mathcal{X} \rightarrow \mathbb{R}$ depends on a subset of variables $y_\alpha = (y_i)_{i \in \alpha}$ are nonlinear functions of the parameters.

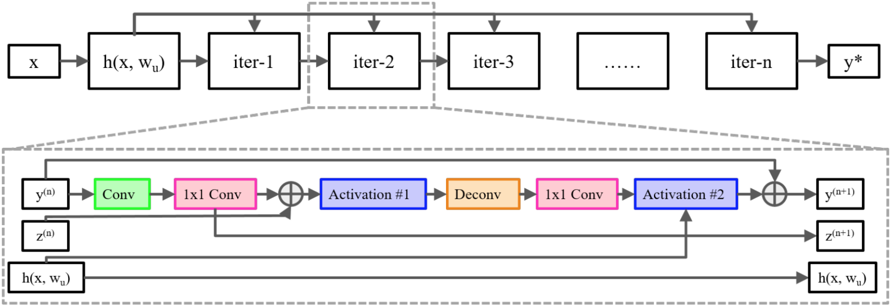
Proximal Models

$$prox_f(x_0) = argmin_y (y - x_0)^2 + f(y) \tag{3}$$

$$E(x, y; w) = \sum_i g_i(y_i, h_i(x; w)) + \sum_\alpha h_\alpha(x; w) g_\alpha(w_\alpha^T y_\alpha) \tag{4}$$

$$\begin{aligned} &min_{y\in\mathcal{Y}} max(z \in \mathcal{Z}) \sum_i g_i(y_i, h_i(x; w)) \\ &\quad - \sum_\alpha h_\alpha(x, w) g^*_\alpha(w_\alpha^T y_\alpha) + \sum_\alpha h_\alpha(x, w) \langle w_\alpha^T y_\alpha, z_\alpha \rangle \end{aligned} \tag{5}$$

$$w^* = argmin_w \sum_n \ell(y_n^*, y_n^{gt}) + \gamma r(w) \tag{6}$$



$$
\begin{cases}
z_\alpha^{(t+1)} &= \operatorname{prox}_{g_\alpha^*}\left(z_\alpha^{(t)} + \frac{\sigma_\rho}{h_\alpha(\mathbf{x};\mathbf{w})}\mathbf{w}_\alpha^T \bar{\mathbf{y}}_\alpha^{(t)}\right) \\
y_i^{(t+1)} &= \operatorname{prox}_{g_i,h_i(\mathbf{x},\mathbf{w})}\left(y_i^{(t)} - \frac{\sigma_\tau}{h_\alpha(\mathbf{x};\mathbf{w})}\mathbf{w}_{\cdot,i}^{*T}\mathbf{z}^{(t+1)}\right) \\
\bar{y}_i^{(t+1)} &= y_i^{(t+1)} + \sigma_{ex}\left(y_i^{(t+1)} - y_i^{(t)}\right)
\end{cases}
$$