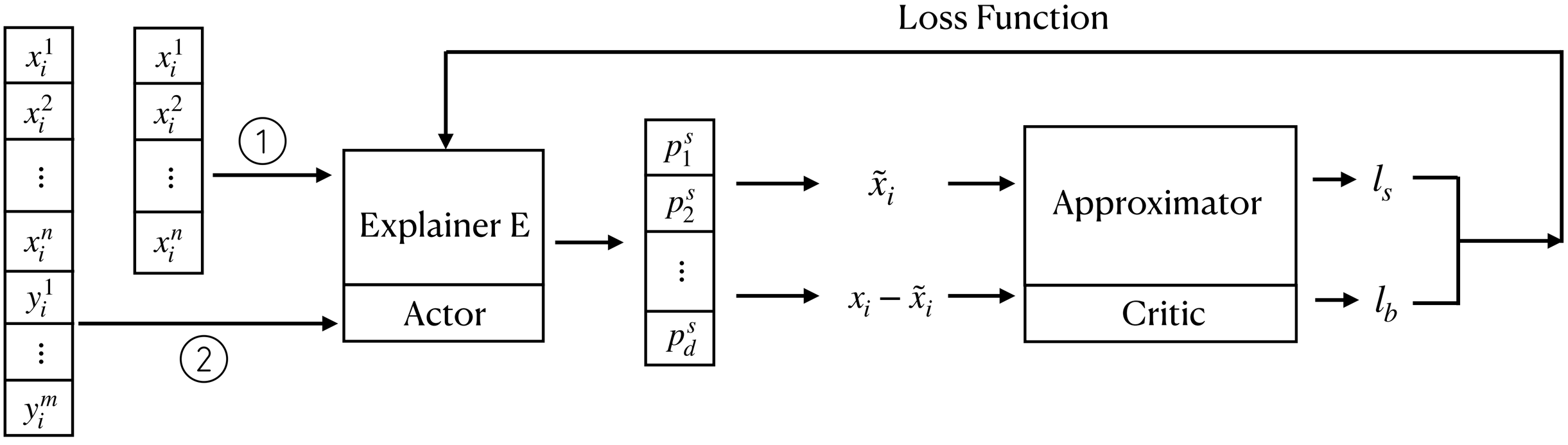

Loss Function
x_i^1
x_i^2
x_i^n
y_i^1
y_i^m
1
2
Explainer E
Actor
p_1^s
p_2^s
p_d^s
$\tilde{x}_i$
$x_i - \tilde{x}_i$
Approximator
Critic
l_s
l_b

Exploration of Potential Improvements to INVASE

Loss Function Revision

✤ INVASE's original loss function

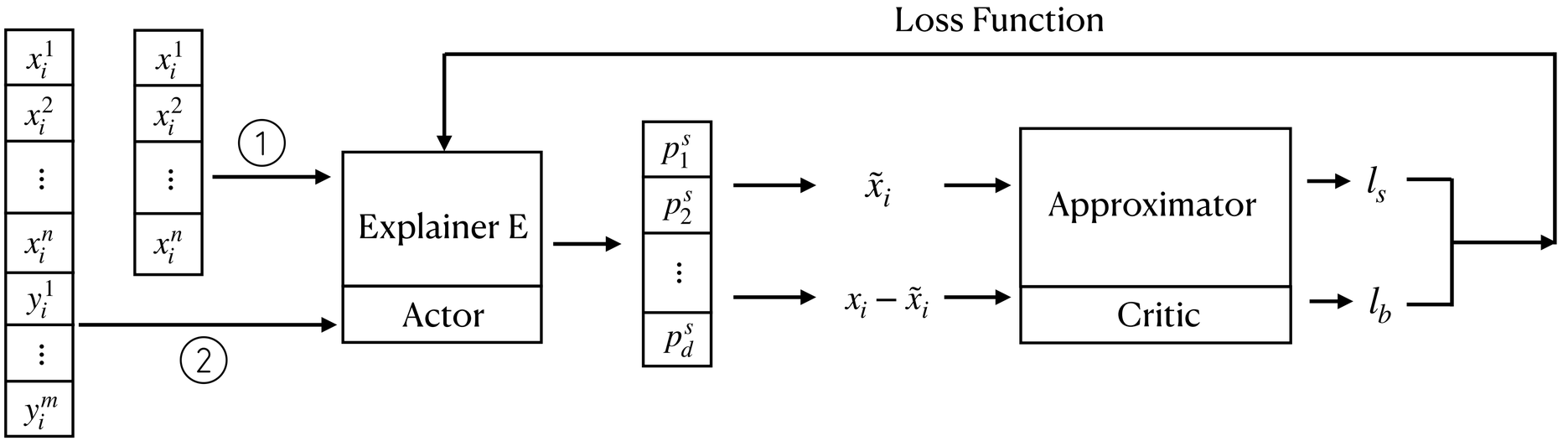
- $\mathbb{E}_{(x,y)\sim p}[\mathbb{E}_s\ [(l_s - l_b) + \lambda\|s\|_0]]$

✤ Loss Function Alternatives

- Direct replacement version: $\mathbb{E}_{(x,y)\sim p}[\mathbb{E}_s\ [(l_s - l_u) + \lambda^*\|s\|_0]]$
- Percentage version: $\mathbb{E}_{(x,y)\sim p}[\mathbb{E}_s\ [\frac{l_s - l_u}{l_u} + \lambda^*\|s\|_0]]$

Exploration of Potential Improvements to INVASE

Loss Function Revision



✤ INVASE's original loss function

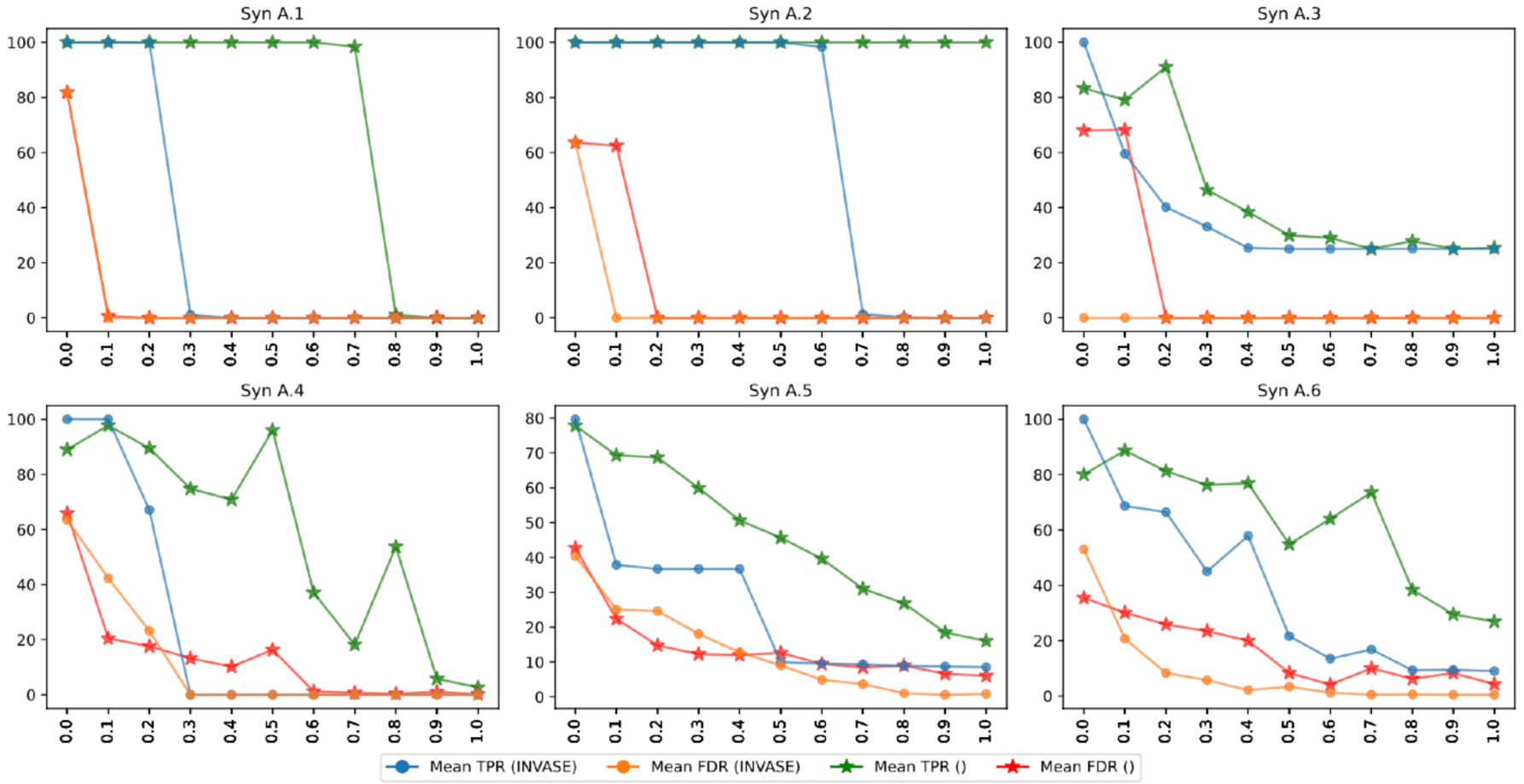
- $\mathbb{E}_{(x,y)\sim p}[\mathbb{E}_s\ [(l_s - l_b) + \lambda\|s\|_0]]$

✤ Loss Function Alternatives

- Direct replacement version: $\mathbb{E}_{(x,y)\sim p}[\mathbb{E}_s\ [(l_s - l_u) + \lambda^*\|s\|_0]]$
- Percentage version: $\mathbb{E}_{(x,y)\sim p}[\mathbb{E}_s\ [\frac{l_s - l_u}{l_u} + \lambda^*\|s\|_0]]$

Exploration of Potential Improvements to INVASE

Application of Direct Replacement Version Loss



Individual Settings: **Activation**: Choose between SeLU and ReLU to determine which yields better results. (*INVASE*: A.1–A.5 use ReLU, A.6 uses SeLU; Proposed: A.1–A.3 use ReLU, A.4–A.6 use SeLU.); **Policy**: Early Stopping Policy ($\delta = 0.3\%$, patience $T = 5$).