Related Work

- Collect neighborhoods with weights
- Generate a individual explainer
- Use White-box models

- **Globally** train local explainer
- Use **Neural Network** models
- Based mutual information defination

Fixed Number of Selection & Non-Real Time

- LIME (Local Interpretable Model-agnostic Explanations)
- SHAP (SHapley Additive exPlanations)

Fixed Number of Selection & Real Time

- L2X (Learning to Explain)
- MEED (Model-agnostic Effective Efficient Direct)

Variable Number of Selection & Real Time

- INVASE
  (Instance-wise Vari- able Selection using Neural Network)

- **Globally** train local explainer
- Use **Neural Network** models
- Based **Actot-Critic** methodology

Methods	**Neural Network-based Explainer**	**Real-time Explanation**	**Need to decide the number of features to select**	**Use of Unselected Features**
LIME	×	×	✓	×
SHAP	×	×	✓	×
L2X	✓	✓	✓	×
MEED	✓	✓	✓	✓
INVASE	✓	✓	×	×

Related Work

Fixed Number of Selection & Non-Real Tim

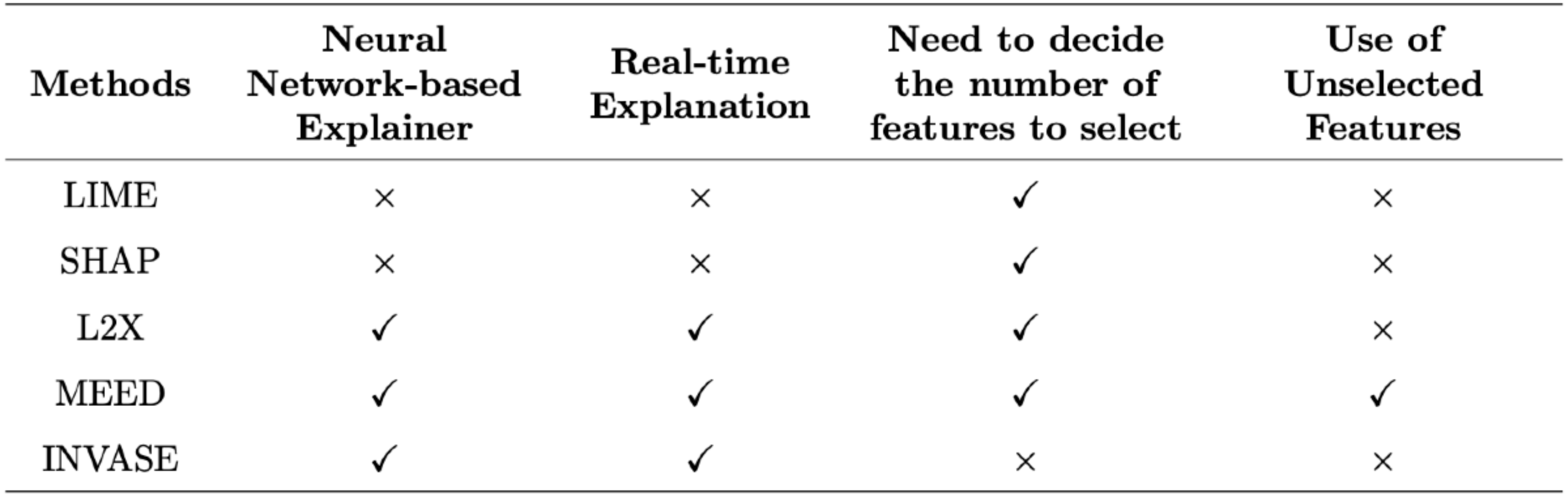
- LIME (Local Interpretable Model-agnostic Explanations
- SHAP (SHapley Additive exPlanations)

Fixed Number of Selection & Real Time

- L2X (Learning to Explain)
- MEED (Model-agnostic Effective Efficient Direct)

Variable Number of Selection & Real Time

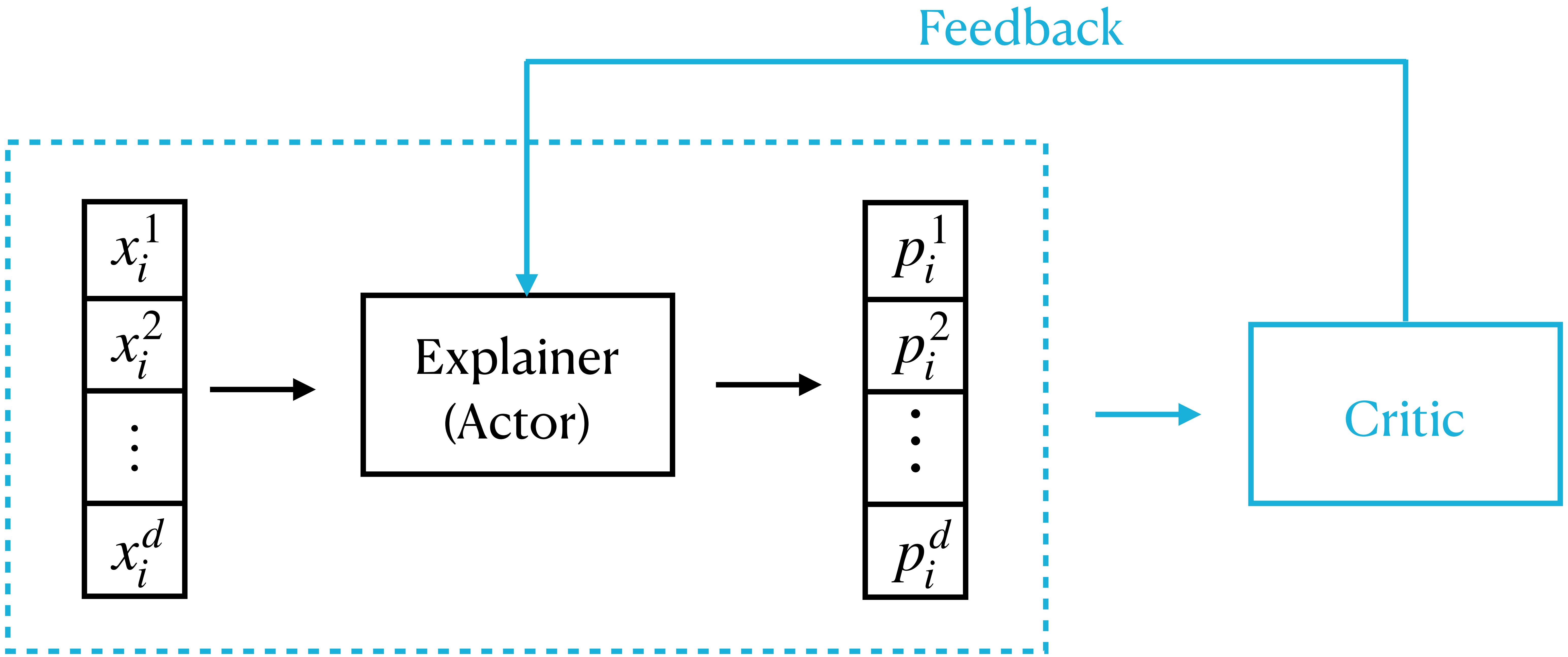
- INVASE
  (Instance-wise Vari- able Selection using Neural Network)



Methods	Neural Network-based Explainer	Real-time Explanation	Need to decide the number of features to select	Use of Unselected Features
LIME	×	×	✓	×
SHAP	×	×	✓	×
L2X	✓	✓	✓	×
MEED	✓	✓	✓	✓
INVASE	✓	✓	×	×

Related Work

Actor-Critic Based IFS Methods



- Actor (Explainer): Pick features for a given instance
- Critic: Evaluate the goodness of current Actor