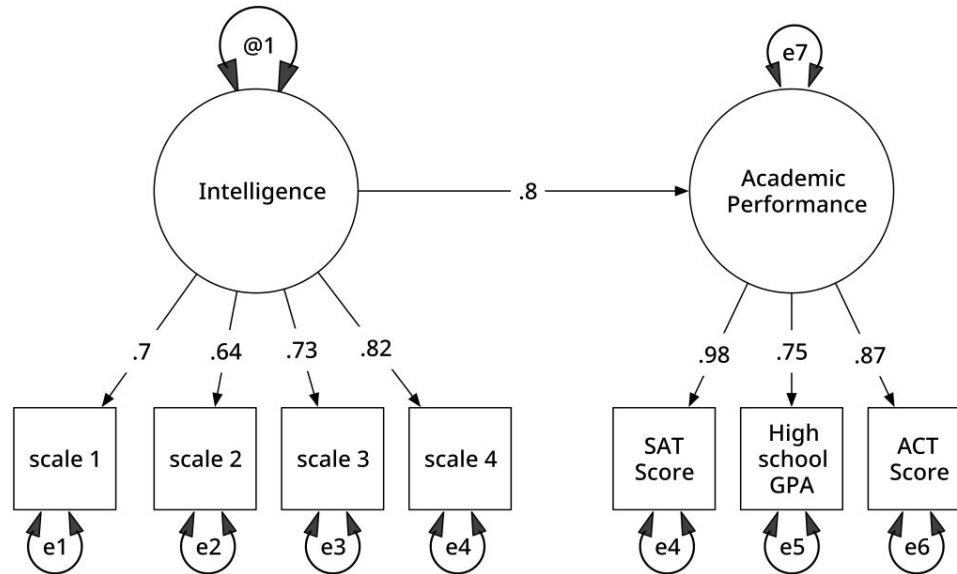


Исследование эффективности методов обучения DAG.

Галкин, Коршунова, Шаповалов

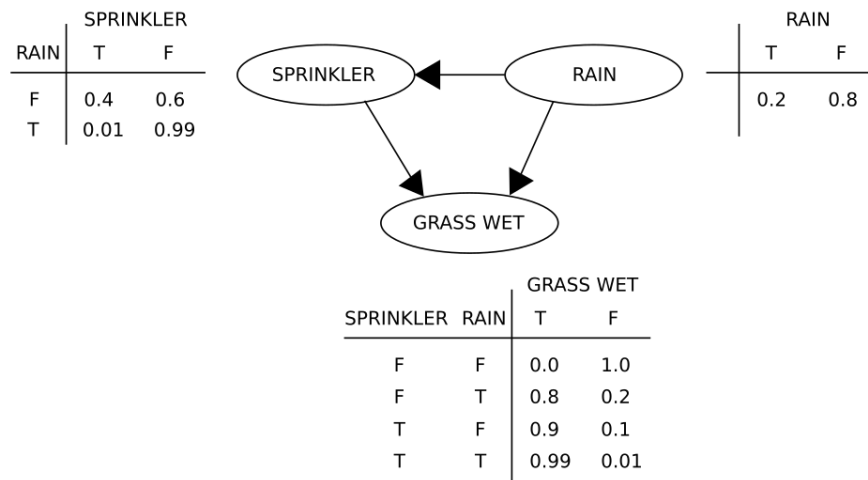
Применение DAG

- Байесовские сети
- Структурные уравнения (SEM)



Постановка задачи обучения DAG

На основе наблюдаемых данных восстановить связи между переменным - структуру графа. В непрерывных БС и SEM - также и веса графа.



Методы обучения



- Score-based: FGS
 - Локальные минимумы
 - Вычислительная сложность
 - Чувствительность к выбору функции стоимости
- Constraint-based: PC - алгоритм
 - Чувствительность к шуму
 - Сложность при большом числе переменных
 - Точность зависит от тестов

Проблема



Из-за комбинаторного ограничения ацикличности обучение DAG - NP -трудная задача. Вышеперечисленные методы из-за того же комбинаторного поиска обладают недостатками:

- чувствительность к шуму
- высокая вычислительная сложность
- неэффективны на больших размерностях

В следующих методах предлагается перейти к задаче непрерывной оптимизации некоторой score-функции. Так же, как и в score-based методах.

$$\begin{array}{ll} \min_{\mathbf{G}} & Q(\mathbf{G}) \\ \text{subject to} & \mathbf{G} \in \mathbb{D} \end{array}$$

Метод NOTEARS

В данном методе предлагается свести задачу оптимизации с предыдущего слайда к задаче непрерывной оптимизации некоторой функции потерь $F : \mathbb{R}^{d \times d} \rightarrow \mathbb{R}$

$$\begin{array}{ll} \min_{W \in \mathbb{R}^{d \times d}} F(W) \\ \text{subject to } G(W) \in \text{DAGs} \end{array} \iff \begin{array}{ll} \min_{W \in \mathbb{R}^{d \times d}} F(W) \\ \text{subject to } h(W) = 0, \end{array}$$

где

$$h(W) = \text{tr}(e^{W \circ W}) - d = 0,$$

И для линейных SEM

$$F(W) = \ell(W; \mathbf{X}) + \lambda \|W\|_1 = \frac{1}{2n} \|\mathbf{X} - \mathbf{X}W\|_F^2 + \lambda \|W\|_1$$

Метод DAGMA



DAGMA использует лог-детерминант для проверки ацикличности

$$h(W) = -\log \det(sI - W \circ W) + d \log s$$

И в общем виде задача оптимизации выглядит так:

$$\min_{\theta} Q(f_{\theta}; X) + \beta_1 \|\theta\|_1 \quad \text{subject to } h_{\text{ldet}}^s(W(\theta)) = 0$$

Где Q - некоторая score-функция и

$$[W(f)]_{i,j} \stackrel{\text{def}}{=} \|\partial_i f_j\|_{L^2}$$

Область определения h

$$\mathbb{W}^s = \{W \in \mathbb{R}^{d \times d} \mid s > \rho(W \circ W)\}$$

Основные свойства методов

В NOTEARS $h(W)$ - невыпукла, поэтому задачу преобразуют к следующей

$$\min_W F(W) + \frac{\rho}{2} h(W)^2 \quad \nabla h(W) = (e^{W \circ W})^T \circ 2W,$$

В DAGMA $h(W)$ выпуклая, что гарантирует нахождение глобального минимума, задачу также преобразуют к безусловной и используют $h(W)$ в качестве регуляризатора.

$$\bar{\theta}^{(t+1)} = \arg \min_{\theta} \mu^{(t)} (Q(f_{\theta}; X) + \beta_1 \|\theta\|_1) + h_{\text{ldet}}^s(W(\theta))$$

$$\nabla h_{\text{ldet}}^s(W) = 2(sI - W \circ W)^{-T} \circ W,$$

Асимптотическая сложность обоих методов кубическая от размерности графа, однако на практике DAGMA работает быстрее.

Эксперименты на синтетических данных



Генерировалась G , а затем из нее получалась W .

С помощью W генерировались данные линейного SEM:

$$X = W^T X + z \in \mathbb{R}^d$$

На которых методы и восстанавливали искомый DAG.

Было показано, что метод NOTEARS, работает точнее FGS для различных n .

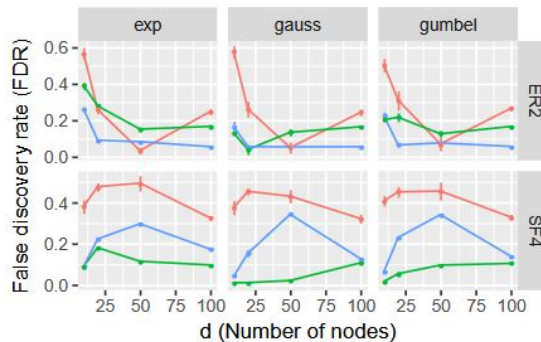
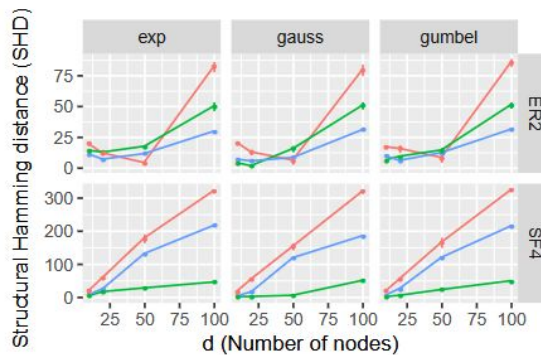
Нами было реализовано сравнение методов NOTEARS и DAGMA.

Метрики точности, не включая время:

$$\text{FDR} = \frac{\text{FP}}{\text{FP} + \text{TP}} \quad \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{SHD} = \text{FP} + \text{FN} + \text{Wrong directions}$$

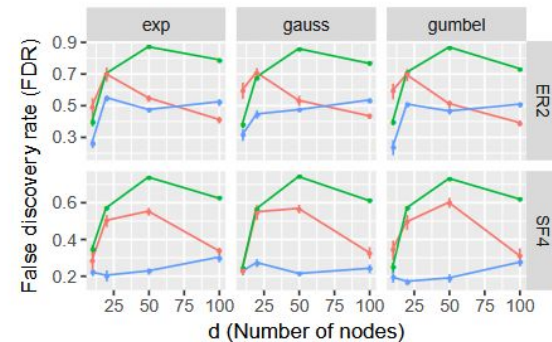
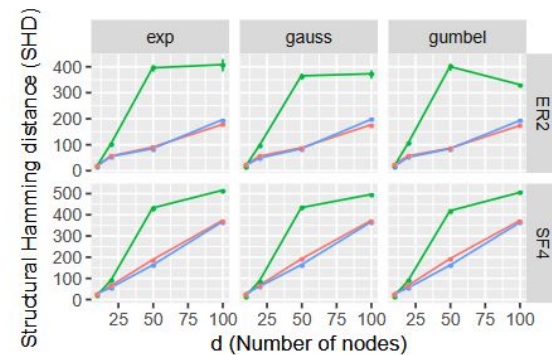
Эксперименты на синтетических данных

Было показано, что метод
NOTEARS, работает точнее
FGS для различных n .



Method FGS NOTEARS NOTEARS-L1

(a) SHD with $n = 1000$

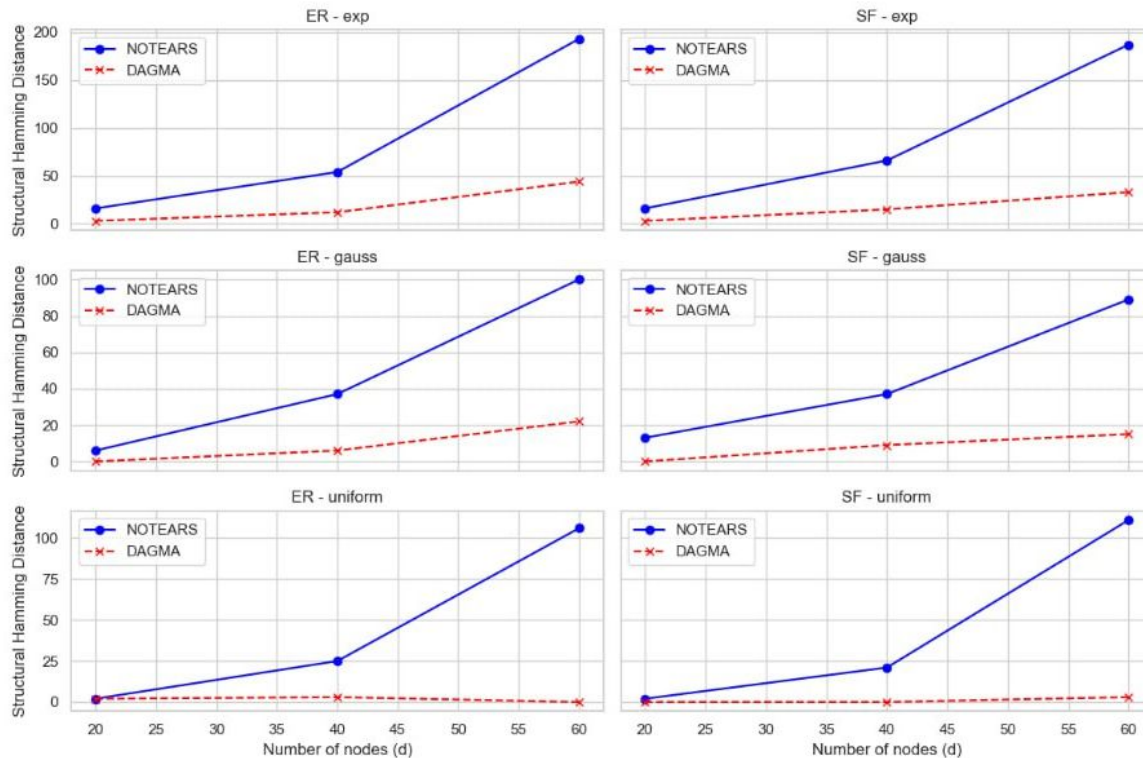


Method FGS NOTEARS NOTEARS-L1

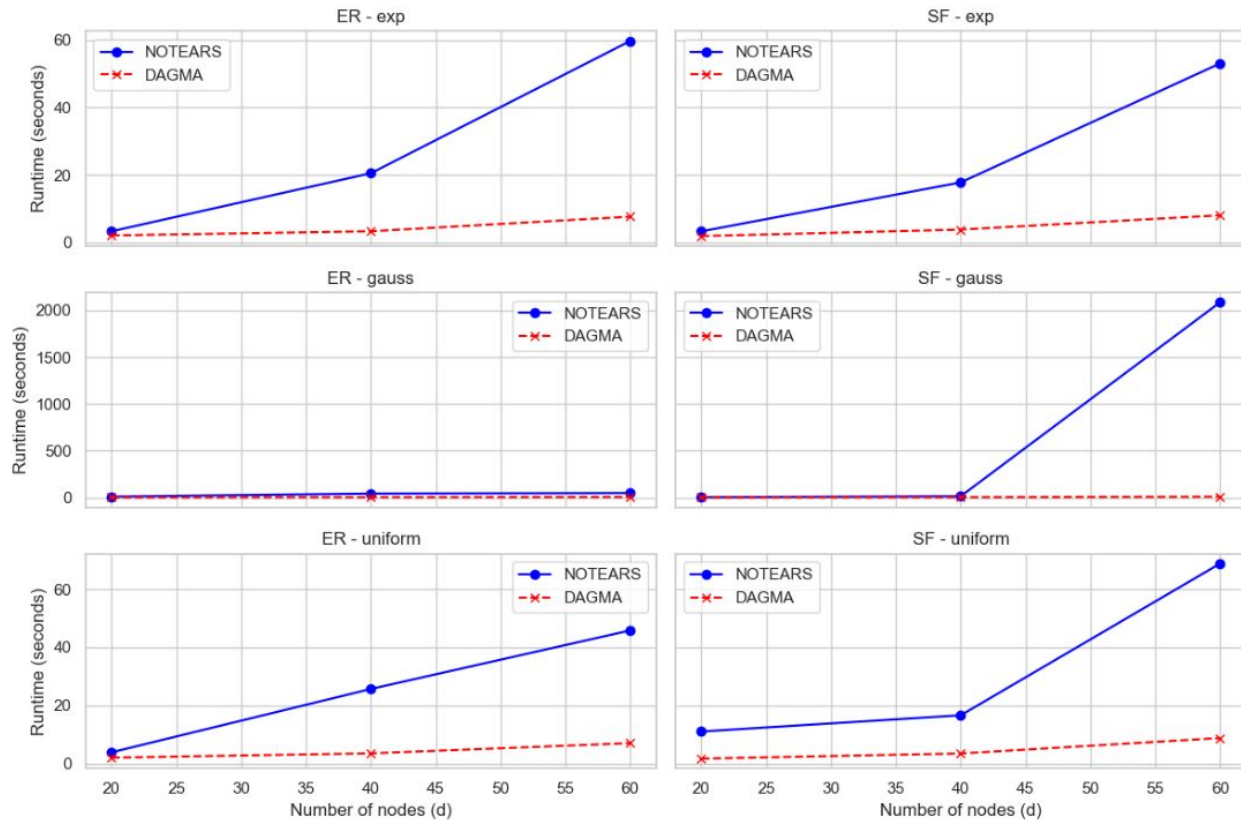
(b) SHD with $n = 20$

Эксперименты на синтетических данных

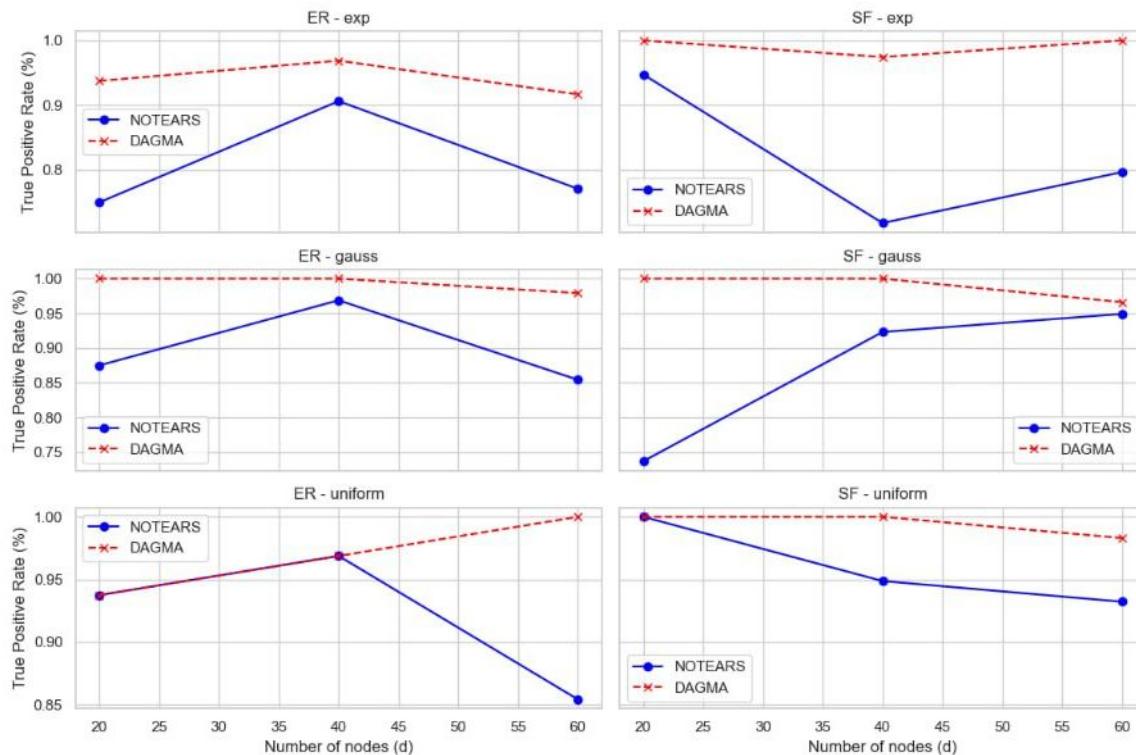
Сравнение DAGMA и
NOTEARS по SHD



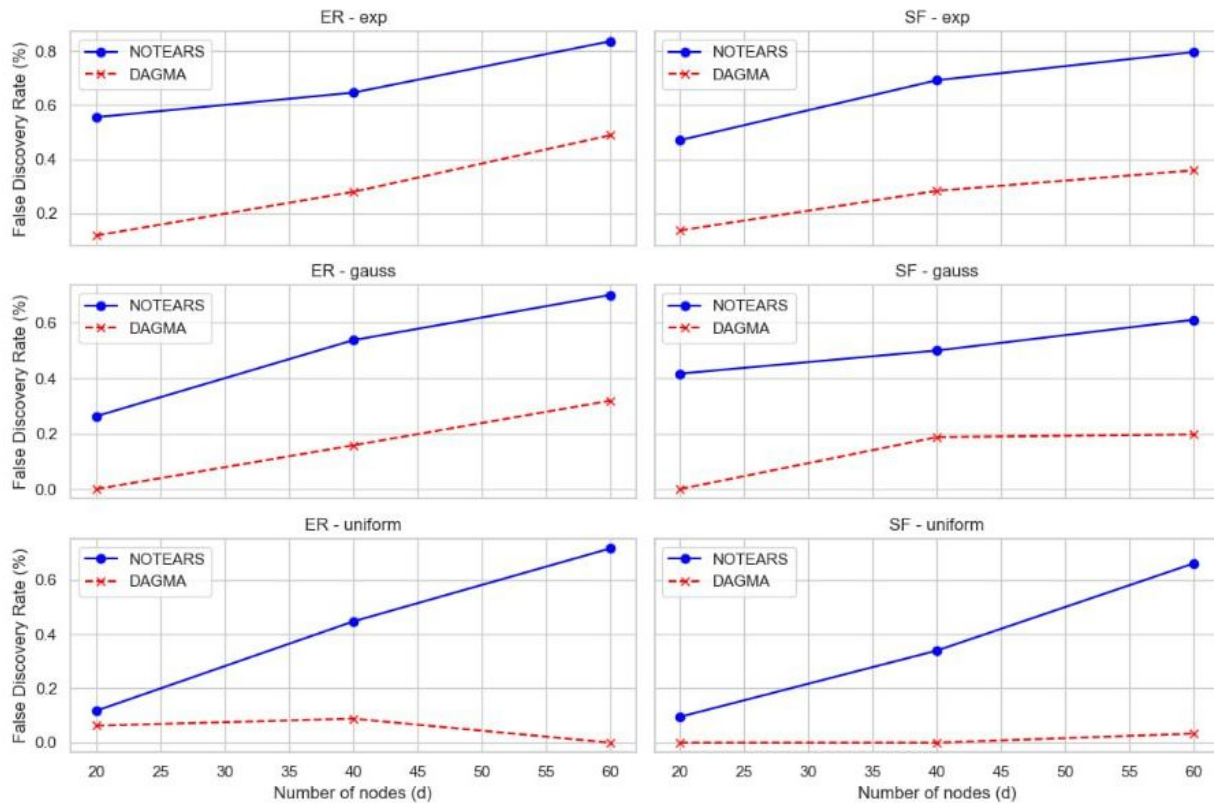
Эксперименты на синтетических данных



Эксперименты на синтетических данных



Эксперименты на синтетических данных



Итоги



Была показана эффективность методов NOTEARS и DAGMA, основанных на решении непрерывной оптимизационной задачи:

- NOTEARS показал себя точнее и эффективнее FGS на различных размерностях
- DAGMA в свою очередь оказался более эффективным и точным, чем NOTEARS.

Есть и ограничения:

- Зависимость от порогового параметра, который определяет, какие из элементов матрицы W следует занулить
- Довольно высокая вычислительная сложность
- Отсутствие гарантии, что был восстановлен именно истинный граф и действительно ли ребро соответствует причинно следственной связи

Ссылки



- <https://paperswithcode.com/paper/dagma-learning-dags-via-m-matrices-and-a-log> - DAGMA
- <https://arxiv.org/abs/1803.01422> - NOTEARS
- <https://github.com/korshunovamm/Learning-Dags-NLA-course/tree/main> - код с экспериментами и их более подробное описание.