

# Автоматическое построение нейросети оптимальной сложности

Маркин Валерий, Забазнов Антон, Горян Николай,  
Сергей Губанов, Сергей Таранов, Товкес Артём, Улитин  
Александр, Криницкий Константин

Московский физико-технический институт

10 декабря, 2018г.

## Иследуется

Задача выбора структуры нейронной сети.

## Требуется

Найти нейросеть оптимальной сложности.

## Проблемы

- Большое количество параметров,
- Высокая вычислительная сложность оптимизации,
- Невозможность использования эвристических и переборных алгоритмов выбора структуры модели

- *LeCun Y., Denker J. , Solla S.*  
Optimal Brain Damage // Advances in Neural Information Processing Systems, 1989. Vol. 2. P. 598–605.
- *Graves A.*  
Practical Variational Inference for Neural Networks // Advances in Neural Information Processing Systems, 2011. P. 2348–2356.
- *Bishop C.*  
Pattern Recognition and Machine Learning. — Berlin: Springer, 2006. 758 p.
- *Grunwald P. A*  
Tutorial introduction to the minimum description length principle. 2005.

# Постановка задачи

$$\mathcal{D}^{\text{train}} = \{\mathbf{x}_i, y_i\}, \quad i = 1, \dots, m^{\text{train}},$$

$$\mathcal{D}^{\text{valid}} = \{\mathbf{x}_i, y_i\}, \quad i = 1, \dots, m^{\text{valid}},$$

где  $\mathbf{x}_i \in \mathbf{X} \subset \mathbb{R}^n$ ,  $y_i \in \mathbf{Y} \subset \mathbb{R}$ .

$y \in \mathbf{Y} = \{1, \dots, Z\}$ , где  $Z$  - количество классов.

Модель задаётся ориентированным графом  $\mathbf{G} = (\mathbf{V}, \mathbf{E})$

$\mathbf{g}^{i,j}$ — базовые функции ребра  $(i, j)$  с весами  $\gamma^{i,j}$

Требуется построить такую модель  $\mathbf{f}$  с параметрами  $\mathbf{W} \in \mathbb{R}^n$ :

$$\mathbf{f}(\mathbf{x}, \mathbf{W}) = \{\mathbf{f}_i(\mathbf{x}, \mathbf{w}_i)\}_{i=1}^{|\mathbf{V}|}$$

где  $\mathbf{f}_i(\mathbf{x}, \mathbf{w}_i)$  - подмодель с параметрами  $\mathbf{w}_i$  задаётся как:

$$\mathbf{f}_i(\mathbf{x}, \mathbf{w}_i) = \sum_{j \in \text{adj}(i)} \langle \gamma^{i,j}, \mathbf{g}^{i,j} \rangle \mathbf{f}_j(\mathbf{x}, \mathbf{w}_j)$$

# Постановка задачи

Функция потерь на обучении  $L$  и функция потерь на валидации  $Q$  задаются как:

$$L(\mathbf{W}, \mathbf{A}, \Gamma) = \log p(\mathbf{Y}^{\text{train}} | \mathbf{X}^{\text{train}}, \mathbf{W}, \Gamma) + e^{\mathbf{A}} \|\mathbf{W}\|^2,$$

$$Q(\mathbf{W}, \Gamma) = \log p(\mathbf{Y}^{\text{valid}} | \mathbf{X}^{\text{valid}}, \mathbf{W}, \Gamma) + \lambda p(\Gamma),$$

где  $\mathbf{A}$  и  $\lambda$  — регуляризационные слагаемые,  $p(\Gamma)$  — произведение всех произведение вероятностей всех  $\gamma^{i,j} \in \Gamma$ .

Требуется решить задачу двухуровневой оптимизации, оптимизируя параметры модели по обучающей выборке, а структуру модели по валидационной:

$$\mathbf{W}^*(\Gamma) = \arg \min_{\mathbf{W}} L(\mathbf{W}, \Gamma)$$

$$\Gamma^*, \mathbf{A}^* = \min_{\Gamma, \mathbf{A}} Q(\mathbf{W}^*(\Gamma), \Gamma)$$

