

## Deep Feature Selection(DFS)

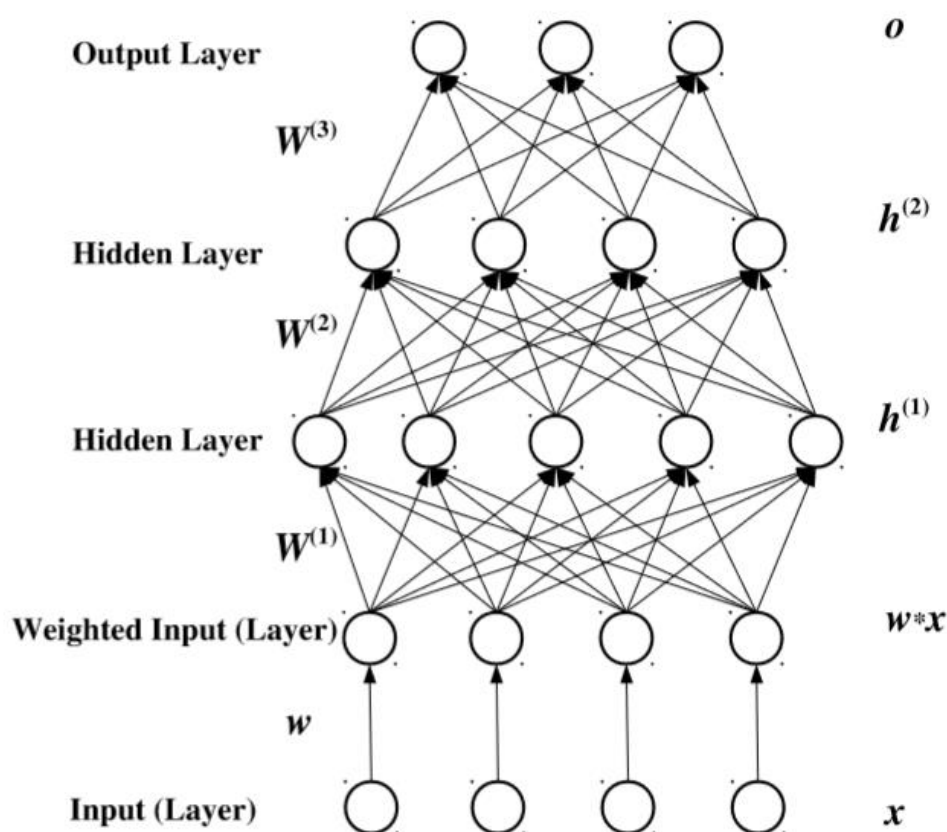
Паненко М.В.

В данном отчете содержится информация о реализации метода (DFS), из статьи [Deep Feature Selection: Theory and Application to Identify Enhancers and Promoters](#).

Суть метода заключается в добавлении некоторого слоя (weighed layer, WL) к стандартной полносвязной нейронной сети, таким образом, что weighted layer учится выбирать наиболее значимые с точки зрения вклада в предсказание целевой переменной признаки.

По данным авторов статьи и людей применявшим этот метод, выбор параметров предсказанных сетью позволяет на больших объемах данных увеличивать точность линейных моделей классификации на 5-7%.

Для задачи ранжирования признаков по метрике 'BAD' выбрал архитектуру сети:



(d) Deep DFS

Извлекая веса из первых слоев с одной связью получаю данные:

	name	value
4	FIRSTLOAN	[[3.2013562]]
0	FAMILYQUANT	[[2.0547101]]
3	MONTHINCOME	[[1.7984327]]
6	IS_ONLINE	[[1.5143607]]
2	CLIENTAGE	[[1.5068781]]
5	issued_sum	[[0.70071745]]
1	FAMILYSTATUS	[[0.18181011]]

Возможно, эти данные позволят улучшить предсказательную силу случайного леса.

Данная полносвязная сеть(очень простая) дает точность предсказания при кросс валидации 78%.

Код метода на githab: [https://github.com/mark-rtb/data\\_mining/blob/master/app/keras\\_DeepFeatureSelection.py](https://github.com/mark-rtb/data_mining/blob/master/app/keras_DeepFeatureSelection.py)