Получение простой выборки на выходе слоя нейронной сети*

Панченко С. К.[1]***, Гадаев Т. Т.***[1]***, Грабовой А. В.***[1]***, Стрижов В. В.***
panchenko.sk@phystech.edu
[1]Московский Физико-Технический Институт

Границы применимости многочисленных методов математической статистики, применяемых в анализе, обуславливаются знанием вероятностной природы данных, которая, как правило, заранее неизвестна. В данной статье предлагается объединение целого комплекса статистических критериев в универсальный инструмент исследования распределений, порождающих выборки. Также предлагается алгоритм усовершенствования нейронной сети на основе анализа выборки, полученной на выходе её предпоследнего слоя, с помощью предложенного инструмента.

Ключевые слова: *тестирование гипотез, полносвязная нейросеть, условия Гаусса-Маркова.*

1 Введение

Знание того, какое распределение породило признаковое описание выборки или шум в заданных ответах, позволяет обоснованно применять к построенной модели разнообразные критерии математической статистики - мощный аппарат, позволяющий анализировать вероятностные закономерности в данных. Подобный анализ не только помогает интерпретировать модель и улучшать её предсказательные способности, но также позволяет производить отбор признаков по значимости, строить доверительные интервалы, и многое другое. Выборку, для которой известно семейство и параметры распределений, породивших признаковое описание её объектов и, возможно, шум в заданных ответах, назовём простой. На практике, однако, эти распределения часто заранее неизвестны, и изучать их вероятностный характер при необходимости приходится самостоятельно. Именно построение универсального алгоритма, позволяющего установить вероятностный закон, породивший данные, или, другими словами, алгоритма, исследующего выборку на простоту, и становится целью исследования.

Одно из ключевых применений искомого алгоритма находит своё место в совершенствовании нейронных сетей. Как известно, часто результат работы нейронной сети можно рассмотреть как применение некоторой обобщенно-линейной модели к выходам предпоследнего слоя. Совокупность слоев нейронной сети, вплоть до предпоследнего, можно в такой интерпретации рассматривать как композицию в общем случае нелинейных преобразований признакового описания исходной выборки. В итоге на предпоследнем слое сети формируется преобразованная выборка, которая исследуется на простоту с помощью построенного инструмента. Результаты этого исследования помогут усовершенствовать имеющуюся сеть, к примеру, с помощью отбора признаков по значимости окажется возможным предложить технику прореживания сети без потери качества, что является актуальной задачей глубокого обучения.

Как уже упоминалось выше, проводимое исследование существенно опирается на теорию и приложения математической статистики. В первую очередь рассматриваются такие статистические критерии, как критерий Вальда, тест Уайта, тесты Голдфелда-Кванта и

*Научный руководитель: Стрижов В. В. Задачу поставил: Грабовой А. В. Консультант: Гадаев Т. Т.

Дарби-Ватсона, тесты хи-квадрат, Жарка-Бера и Шапиро-Уилка. На основе этих критериев в применении к различным выборкам и вырабатывается искомый алгоритм. Вычислительный эксперимент проводится на наборе стандартных выборок, а также на синтетических и реальных данных.

В качестве основных источников исследованию послужат такие фундаментальные публикации, как [1], ..., а также следующие статьи: ... .

2 Постановка задачи

Пусть $\mathbf{X} \in \mathbb{R}^{m \times n}$ - признаковое описание заданной обучающей выборки с вектором ответов $\mathbf{y} \in \mathbb{R}^m$. Пусть также имеется исходная модель - полносвязная нейронная сеть-перцептрон

$$f(\mathbf{x}) = f(\mathbf{x}|\mathbf{w}) : \mathbb{R}^n \to \mathbb{R},$$

где $\mathbf{w}$ - совокупность параметров нейронной сети.

В самом общем виде нейронную сеть можно представить в виде последовательности преобразований вида

$$f_i(\mathbf{z}) : \mathbb{R}^{k_{i-1}} \to \mathbb{R}^{k_i}, f_i(\mathbf{z}) = \mathbf{r}_i(\mathbf{W}_i \times \mathbf{z} + \mathbf{b}_i)$$

- преобразования во внутренних слоях сети, где $i = 1, ..., s-1$ (s - количество слоев нейронной сети), $\mathbf{W}_i \in \mathbb{R}^{k_i \times k_{i-1}}$ - матрица весов i-ого слоя, $\mathbf{b}_i \in \mathbb{R}^{k_i}$ - вектор порогов активации i-ого слоя, $\mathbf{r}_i : \mathbb{R}^{k_i} \to \mathbb{R}^{k_i}$ - функция активации i-ого слоя ($k_0 = n$).

Завершает цепочку преобразований функция активации последнего слоя, представляющая собой (в простейшем случае) линейную модель над выходами предпоследнего слоя:

$$f_s(\mathbf{z}) : \mathbb{R}^{k_{s-1}} \to \mathbb{R}, f_s(\mathbf{z}) = \langle \mathbf{w}_s, \mathbf{z} \rangle + b_s,$$

где $\mathbf{w}_s \in \mathbb{R}^k_{s-1}$ - вектор весов, а $b_s \in \mathbb{R}$ - свободный член линейной модели последнего слоя.

Итого, результат действия нейронной сети на объект представляется в виде композиции:

$$f(\mathbf{x}) = (f_s \circ f_{s-1} \circ ... \circ f_2 \circ f_1)(\mathbf{x}) = f_s((f_{s-1} \circ ... \circ f_2 \circ f_1)(\mathbf{x})) = g(h(\mathbf{x})) = (g \circ h)(\mathbf{x}),$$

где $g = f_s, g(\mathbf{x}) : \mathbb{R}^l \to \mathbb{R}$ (здесь $l := k_{s-1}$), $h = f_{s-1} \circ ... \circ f_2 \circ f_1, h(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^l$.

Рассмотрим теперь $\mathbf{Z} = [h(\mathbf{x}_1)^\mathsf{T}, ..., h(\mathbf{x}_m)^\mathsf{T}]^\mathsf{T} \in \mathbb{R}^{m \times l}$ - совокупность преобразованных внутренними слоями сети исходных векторов признаков, или, другими словами, преобразованное сетью признаковое описание заднной обучающей выборки. Результат действия исходной модели $f(\mathbf{x})$ на объект $\mathbf{x}_i \in \mathbf{X}$ эквивалентен результату действия последнего слоя сети, т.е. линейной модели $g(\mathbf{z}) = \langle \mathbf{w}_s, \mathbf{z} \rangle + b_s$, на преобразованный объект $\mathbf{z}_i = h(\mathbf{x}_i) \in \mathbf{Z}$.

Итак, обратимся теперь к исследованию вышеупомянутой модели $g(\mathbf{z})$ применительно к выборке $\mathfrak{D} = \{(\mathbf{z}_i, y_i)|\ \mathbf{z}_i = h(\mathbf{x}_i), \mathbf{x}_i \in X, i = 1, ..., m\}$. В первую очередь имеющаяся модель иследуется на удовлетворение условиям Гаусса-Маркова. Сформулируем их в виде следующей теоремы:

Теорема 1. *Рассматривается модель регрессии, в которой наблюдения* $\mathbf{Y}$ *связаны с* $\mathbf{X}$ *следующей зависимостью:* $\mathbf{Y}_i = \beta_1 + \langle \beta_2, \mathbf{X}_i \rangle + \varepsilon_i$. *На основе* n *выборочных наблюдений оценивается уравнение регрессии* $\hat{\mathbf{Y}}_i = \hat{\beta}_1 + \langle \hat{\beta}_2, \mathbf{X}_i \rangle$. *Теорема Гаусса—Маркова гласит:*

Если данные обладают следующими свойствами:

– *Модель данных правильно специфицирована;*

- *Все* $\mathbf{X}_i$ *детерминированы и не все равны между собой;*
- *Ошибки не носят систематического характера, то есть* $\mathbb{E}(\varepsilon_i) = 0\ \forall i$*;*
- *Дисперсия ошибок одинакова и равна некоторой* σ^2*;*
- *Ошибки некоррелированы, то есть* $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0\ \forall i, j$*;*

— то в этих условиях оценки метода наименьших квадратов оптимальны в классе линейных несмещённых оценок.

В случае, когда модель $g(\mathbf{z})$ с выборкой $\mathbf{Z}$ не удовлятворяет условиям Гаусса-Маркова, предлагается усовершенствовать исходную сеть, модифицировав параметры её внутренних слоев и получив, таким образом, новую функцию $h'(\mathbf{x})$, такую, что преобразованная выборка $h'(\mathbf{X})$ была простой, а линейная модель уже удовлетворяла бы условиям Гаусса-Маркова. Эта функция также должна быть выбрана из следующих соображений: имеющаяся новая модель - модифицированная нейронная сеть $f'(\mathbf{x}) = g(h'(\mathbf{x}))$ должна иметь меньшую среднеквадратичную ошибку на тестовой выборке (предполагая разделение исходной выборки на обучение и тест). Другими словами, ищется такая модификация h', что

$$Q(\mathbf{X}|f') \leqslant Q(\mathbf{X}|f),$$

где $Q(\mathbf{X}|f) = \frac{1}{m}\sum_{j=1}^{m}(f(\mathbf{x}_m) - \mathbf{y}_m)^2$ - среднеквадратичкая ошибка алгоритма на выборке. Поиску такой функции h' будет посвящен следующий раздел.

Литература

[1] Christopher M. Bishop. *Pattern recognition and machine learning, 5th Edition.* Information science and statistics. Springer, 2007.

Поступила в редакцию