

КЛАССИФИКАЦИЯ ПАРАМЕТРОВ НА ОСНОВЕ ОБЪЕДИНЕНИЯ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ И НАИВНОГО БАЙЕСОВСКОГО КЛАССИФИКАТОРА

А.С. Кузнецова, Д.С. Световец

Томский государственный университет систем управления и радиоэлектроники 634050, г.

Томск, пр. Ленина, 40

E-mail: lehpflehp@gmail.com, ryu77@bk.ru

Введение

Классификация в данном случае рассматривается, как определение класса, к которому принадлежит набор исходных параметров.

Искусственная нейронная сеть – это модель, построенная по принципу работы биологической нейронной сети. В основе нейронной сети лежат нейроны, которые взаимодействуют друг с другом.

В каждый нейрон подаётся набор параметров, каждому параметру присваивается синаптический вес, после полученные веса суммируются, и данная сумма подаётся на вход в активационную функцию и на выходе из функции мы получаем выходной сигнал из этого нейрона. (рис.1) [1]

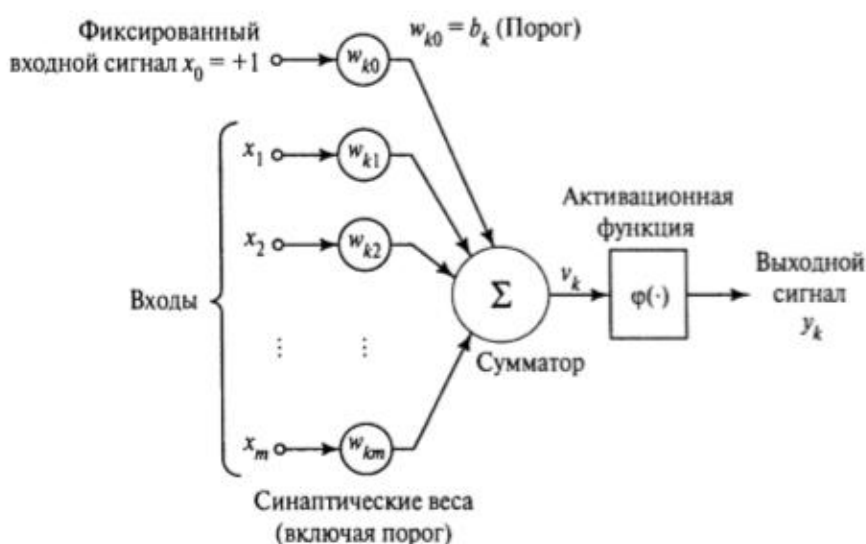


Рисунок 1

В наше время использование искусственной нейронной сети

Наивный классификатор Байеса – это простой вероятностный классификатор, основанный на применении теоремы Байеса со строгими (наивными) предположениями о независимости. [2]

Формула Байеса:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

где $P(A)$ – вероятность гипотезы A , $P(A|B)$ – вероятность гипотезы A при наступлении события B , $P(B|A)$ – вероятность наступления события B при истинности гипотезы A , $P(B)$ – полная вероятность наступления события B .

Постановка задачи

Целью данной работы является эффективное объединение искусственной нейронной сети и наивного байесовского классификатора.

Объединять нейронную сеть и наивный байесовский классификатор, мы будем с помощью методов оптимизации в MATLAB R2016a.

Реализация

При проверке нашей программы, мы использовали базы данных с репозитория KEEL, реализовывали на них работу искусственной нейронной сети и наивного байесовского классификатора, а также их объединения на основе методов оптимизации. Затем сравнивая с оригиналом провели расчёт ошибки для каждого метода классификации. В итоге получилась таблица:

Таблица 1

Название базы данных	Среднее значение доли ошибок при классификации с помощью нейронной сети	Среднее значение доли ошибок при классификации с помощью наивного байесовского классификатора	Среднее значение доли ошибок при объединении полученных классификаций
Banana	0.25	0,4687	0.23
Vowel	0.3423	0.4214	0.2951
Magic	0.2533	0.5325	0.2493
Pima	0.241	0.4924	0.239
Appendicitis	0.314	0.391	0.2621

Теперь проверим значимость объединения и имеет ли оно смысл. Принимаем нулевую гипотезу, такую что доля ошибок при объединении полученных классификаций неотличима от доли ошибок при классификации с помощью нейронной сети, следовательно, альтернативная гипотеза заключается в том, что доля ошибок при объединении полученных классификаций меньше чем доля ошибок при классификации с помощью нейронной сети.

Определяем наблюдаемое значение $Z_{\text{набл}}$ по формуле:

$$Z_{\text{набл}} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{D_{\text{в}}(X)}{n} + \frac{D_{\text{в}}(Y)}{m}}} \quad (2)$$

где, \bar{x} – среднее значение доли ошибок при классификации с помощью нейронной сети, \bar{y} – среднее значение доли ошибок при объединении полученных классификаций, $D_{\text{в}}(X)$ – дисперсия выборочная от значений доли ошибок при классификации с помощью нейронной сети, $D_{\text{в}}(Y)$ – дисперсия выборочная от значений доли ошибок при объединении полученных классификаций, n – количество итераций при вычислении доли ошибок при классификации с помощью нейронной сети, m – количество итераций при вычислении доли ошибок при объединении полученных классификаций. В нашем случае $n=m$, следовательно формулу (2) можно записать в виде:

$$Z_{\text{набл}} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{D_{\text{в}}(X) + D_{\text{в}}(Y)}{n}}} \quad (3)$$

После вычисления наблюдаемых значений мы вычисляем значение функции Лапласа соответствующее нашему уровню значимости по формуле:

$$\Phi(Z_{\text{кр}}) = \frac{1 - 2\alpha}{2} \quad (4)$$

где, α – уровень значимости, который в нашем случае равен 0,95, $\Phi(Z_{\text{кр}})$ – функция Лапласа от критического значения $Z_{\text{кр}}$, затем, чтобы найти критическое значение $Z_{\text{кр}}$ воспользуемся таблицей значений функции Лапласа.

$$\Phi(Z_{\text{кр}}) = 0,45, Z_{\text{кр}} = 0,6.$$

После вычисления наблюдаемого и критического значений сравним их, чтобы проверить нулевую гипотезу, то есть если неравенство $Z_{\text{набл}} \leq Z_{\text{кр}}$ выполняется, то нулевая

гипотеза считается верной, в обратном случае она отвергается и принимается альтернативная гипотеза, которая утверждает, что доля ошибок при объединении полученных классификаций меньше, чем доля ошибок при классификации с помощью нейронной сети.

Таблица 2

Название базы данных	Наблюдаемое значение $Z_{\text{набл}}$	Критическое значение $Z_{\text{кр}}$	Принимаемая гипотеза в результате сравнения
Banana	0.8557	0.6	Альтернативная
Vowel	1.84		Альтернативная
Magic	0,97		Альтернативная
Pima	0,0018		Нулевая
Appendicitis	0.103		Нулевая

Заключение

По полученным значениям видно, что при объединении ошибка классификации уменьшается, следовательно, объединение с целью уменьшения ошибки прошло успешно.

Но применение данного объединения, не всегда целесообразно, потому что в некоторых случаях уменьшение ошибки не значительно.

Список использованной литературы

1. Саймон Хайкин, Нейронные сети полный курс, второе издание // Вильямс – 2006 – Г.1.3 стр.42-45
2. Гмурман В. Е. Теория вероятностей и математическая статистика, — М.: Высшее образование. 2005 — 52 с