

Аппроксимация функций RBF нейронными сетями

1. Цель РГР

Изучение архитектуры RBF сетей, формирование умений в области расчетов параметров RBF-сетей, приобретение навыков самостоятельного решения прикладной задачи с помощью нейросетей.

2. Основные теоретические положения

2.1 Архитектура RBF-сети

Архитектура RBF сети изображена на рисунке 2.1. Сеть является двухслойной.

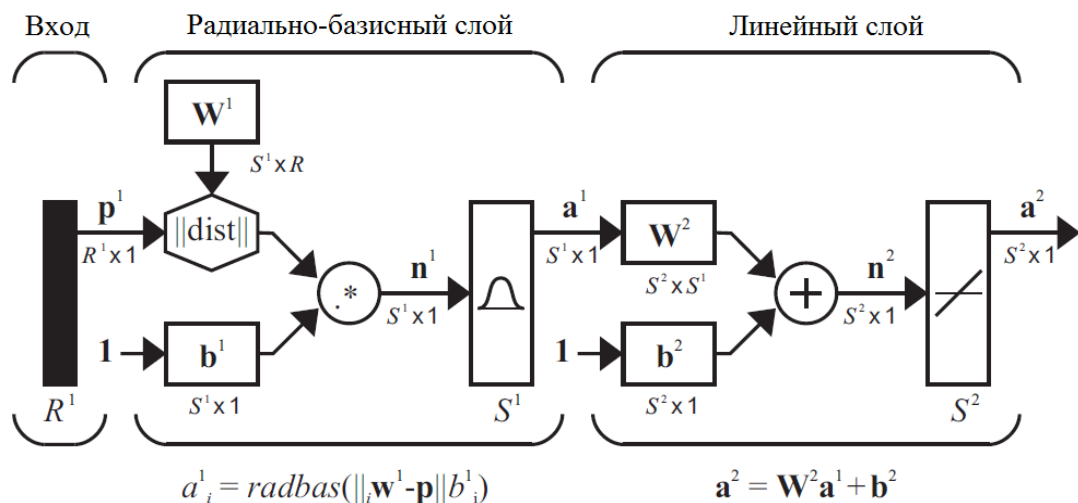


Рисунок 2.1 – Архитектура RBF-сети

В первом слое вычисляется *расстояние* между входным вектором \mathbf{p} и строкой весовой матрицы \mathbf{W}^1 и выполняется *умножение на смещение* \mathbf{b} . Поэтому сетевой вход i -го нейрона 1-го слоя равен:

$$n_i^1 = \|\mathbf{p} - \mathbf{w}_i^1\| b_i^1.$$

Каждая строка матрицы \mathbf{W}^1 соответствует некоторому *центральному значению*, при котором сетевой вход будет равен нулю. Смещение \mathbf{b}^1 задает масштаб, который называют *шириной или радиусом* функции преобразования. В качестве функции преобразования *radbas* может использоваться гауссова функция (рисунок 2.2), тогда $b=1/(\sigma\sqrt{2})$. Это *локальная функция* – значения, удаленные от 0 имеют малые ординаты.

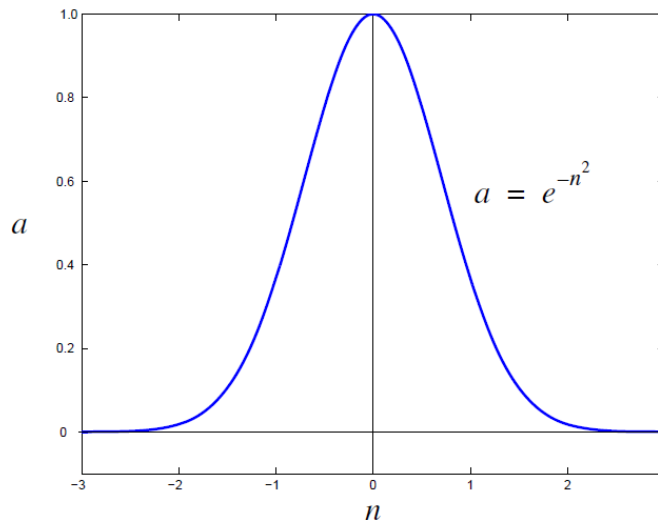


Рисунок 2.2. – Пример гауссовой радиальной функции

RBF-сети, как и MLP, являются универсальными аппроксиматорами. Если имеется достаточное число нейронов скрытого слоя, то сеть может аппроксимировать любую функцию. В отличие от MLP радиальные функции активны только в небольшой области пространства входных значений, а поэтому обладают локальными свойствами. Если входы удалены от центров радиальных базисных функций, то выходные значения будут нулевыми. Поэтому при обучении сети необходимо управлять расположением центров нейронов скрытого слоя и выбирать смещения (масштабы) так, чтобы обеспечивалось значительное перекрытие базисных функций.

2.2 Алгоритмы обучения RBF-сетей

Алгоритмы обучения RBF являются 2-х этапными:

- 1) настройка центров и смещений нелинейного первого слоя;
- 2) обучение линейного второго слоя.

Варианты реализации 1-го этапа.

1. *Настройка центров нейронов 1-го слоя с использованием сетки значений входных переменных:*
 - не оптимально, так как в точках, где аппроксимируемая функция наиболее сложная, требуется больше базисных функций;
 - «проклятие размерности», если узлы выбираются на сетке, то при увеличении размерности входного пространства, число базисных функций увеличивается геометрически.
2. *Использование в качестве центров подмножества входных векторов из обучающей последовательности:*
 - достоинство — центры будут помещены в те области, где располагаются значения входных переменных;
 - не оптимально, из-за случайности входных векторов.
3. *Кластеризация входного пространства и выбор центров кластеров в качестве центров базисных функций:*

- достоинство — центры базисных функций размещаются в областях входного пространства со значительной активностью.

2.3 Обучение линейного слоя на основе алгоритма LS

Будем полагать, что центры 1-го слоя выбраны случайным образом. Тогда все смещения 1-го слоя можно вычислить по формуле

$$b_i^1 = \sqrt{S^1} / d_{\max},$$

где d_{\max} — максимальное расстояние между соседними центрами. Это обеспечивает хорошее перекрытие базисных функций.

Выход 1-го слоя для каждого входного вектора \mathbf{p}_q обучающего множества:

$$\{\mathbf{p}_1, \mathbf{t}_1\}, \{\mathbf{p}_2, \mathbf{t}_2\}, \dots, \{\mathbf{p}_Q, \mathbf{t}_Q\}, n_{i,q}^1 = \|\mathbf{p}_q - \mathbf{w}_i^1\| b_i^1, \mathbf{a}_q^1 = \text{radbas}(\mathbf{n}_q^1).$$

Для случая *скалярного выхода* a_q 2-го слоя с учетом обозначений получим:

$$\mathbf{x} = \begin{bmatrix} {}^1\mathbf{w}^2 \\ b^2 \end{bmatrix}, \quad \mathbf{z}_q = \begin{bmatrix} \mathbf{a}_q^1 \\ 1 \end{bmatrix}, \quad a_q = \mathbf{x}^T \mathbf{z}_q.$$

Тогда целевую функцию $F(\mathbf{x})$ нейросети можно записать в виде:

$$\mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_Q \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} {}^1\mathbf{u}^T \\ {}^2\mathbf{u}^T \\ \vdots \\ {}^Q\mathbf{u}^T \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1^T \\ \mathbf{z}_2^T \\ \vdots \\ \mathbf{z}_Q^T \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_Q \end{bmatrix}. \quad \mathbf{e} = \mathbf{t} - \mathbf{U}\mathbf{x}, \quad F(\mathbf{x}) = (\mathbf{t} - \mathbf{U}\mathbf{x})^T (\mathbf{t} - \mathbf{U}\mathbf{x}).$$

Если целевую функцию дополнить регуляризационной составляющей, то

$$F(\mathbf{x}) = (\mathbf{t} - \mathbf{U}\mathbf{x})^T (\mathbf{t} - \mathbf{U}\mathbf{x}) + \lambda \mathbf{x}^T \mathbf{x} = \mathbf{t}^T \mathbf{t} - 2\mathbf{t}^T \mathbf{U}\mathbf{x} + \mathbf{x}^T [\mathbf{U}^T \mathbf{U} + \lambda \mathbf{I}] \mathbf{x}$$

Сравнивая с квадратичной функцией, получаем

$$F(\mathbf{x}) = c + \mathbf{d}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x}.$$

$$c = \mathbf{t}^T \mathbf{t}, \quad \mathbf{d} = -2\mathbf{U}^T \mathbf{t}, \quad \text{Гессиан } \mathbf{A} = 2[\mathbf{U}^T \mathbf{U} + \rho \mathbf{I}].$$

Решив уравнение

$$\nabla F(\mathbf{x}) = \nabla \left(c + \mathbf{d}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} \right) = \mathbf{d} + \mathbf{A} \mathbf{x} = -2\mathbf{U}^T \mathbf{t} + 2[\mathbf{U}^T \mathbf{U} + \rho \mathbf{I}] \mathbf{x} = 0,$$

получим единственный минимум, если Гессиан положительно полуопределен

$$\mathbf{x}^* = [\mathbf{U}^T \mathbf{U} + \rho \mathbf{I}]^{-1} \mathbf{U}^T \mathbf{t}$$

Последняя формула позволяет вычислить \mathbf{x}^* (вектор параметров второго слоя RBF-сети) при известных значениях активности слоя.

Пример.

Требуется выполнить аппроксимацию функции с использованием RBF-сети структурой 1-3-1, обучающее множества состоит из 6 точек

$$g(p) = 1 + \sin\left(\frac{\pi}{4}p\right), \quad p = \{-2, -1.2, -0.4, 0.4, 1.2, 2\}, \quad t = \{0, 0.19, 0.69, 1.3, 1.8, 2\}.$$

Выберем центры базисных функций, разбив диапазон входа на 3 части в соответствии с числом нейронов скрытого слоя: -2, 0 и 2. Тогда веса и смещения первого слоя примут значения:

$$\mathbf{W}^1 = [-2 \ 0 \ 2], \quad \mathbf{b}^1 = [0.5 \ 0.5 \ 0.5].$$

Выполнив вычисления в соответствии с алгоритмом LS, получим решение

$$\mathbf{x}^* = [\mathbf{U}^T \mathbf{U} + \rho \mathbf{I}]^{-1} \mathbf{U}^T \mathbf{t} = [-1.03 \ 0 \ 1.03 \ 1], \text{ т.е. } \mathbf{W}^2 = [-1.03 \ 0 \ 1.03], \mathbf{b}^2 = [1]$$

На рисунке 2.3. изображены базисные функции и результирующая аппроксимация функции $g(p)$ RBF-сетью с 3-мя нейронами в скрытом слое.

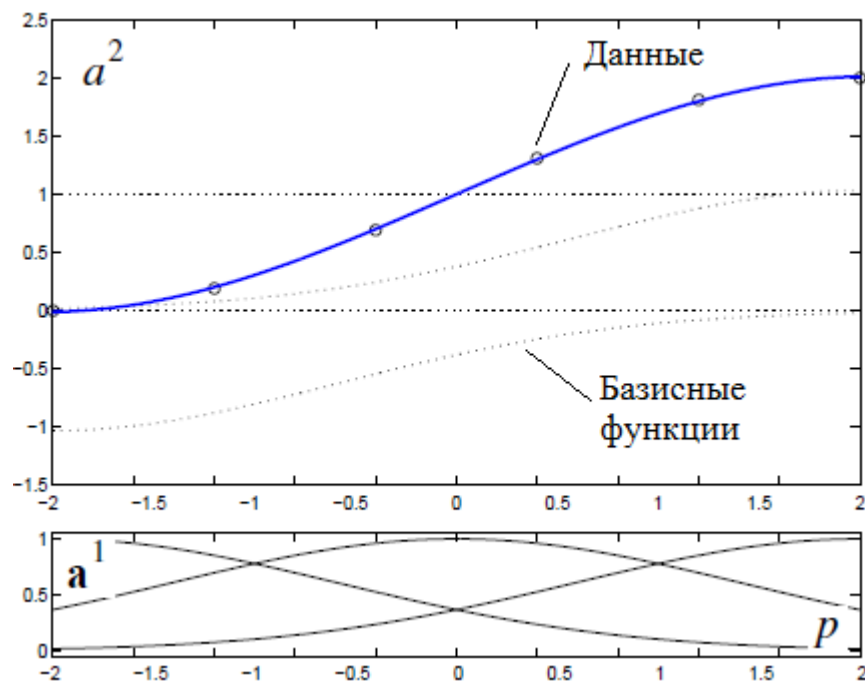


Рисунок 2.3 – Аппроксимация функции RBF-сетью

2.4 Кластеризация и вычисление центров RBF-сети

Для вычисления весов (центров) и смещений 1-го слоя можно применить сети Кохонена, которые выполняют кластеризацию входных векторов

обучающего множества. Пусть имеется обучающее множество $\{\mathbf{p}_1, \mathbf{t}_1\}, \{\mathbf{p}_2, \mathbf{t}_2\}, \dots, \{\mathbf{p}_Q, \mathbf{t}_Q\}$.

Мы хотим определить кластеры входных векторов $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_Q\}$.

Будем обучать веса 1-го слоя RBF-сети с помощью правила Кохонена

$${}_{i^*}\mathbf{w}^1(q) = {}_{i^*}\mathbf{w}^1(q-1) + \alpha(\mathbf{p}(q) - {}_{i^*}\mathbf{w}^1(q-1)),$$

где $\mathbf{p}(q)$ - входной вектор, ${}_{i^*}\mathbf{w}^1(q-1)$ - вектор весов ближайший к $\mathbf{p}(q)$.

После обучения веса нейронов 1-го слоя будут соответствовать центрам кластеров. Таким образом, базисные функции будут располагаться в точках наиболее вероятного появления входных векторов.

Для каждого центра кластера определим расстояние до n_c ближайших векторов

$$dist_i = \frac{1}{n_c} \left(\sum_{j=1}^{n_c} \|\mathbf{p}_j - {}_i\mathbf{w}^1\|^2 \right)^{\frac{1}{2}}$$

Тогда смещение для нейронов первого слоя можно определить по формуле

$$b_i^1 = \frac{1}{\sqrt{2}dist_i}.$$

Можно также использовать и другие алгоритмы кластеризации такие, как самоорганизующиеся карты признаков или k-средних. При этом если кластер широкий, то и соответствующая базисная функция будет так же широкой. В этом случае смещение каждого нейрона в первом слое будет различным. Это делает сеть более эффективной по сравнению со случаем одинаковых смещений. Однако потенциальным недостатком применения методов кластеризации для определения параметров первого слоя RBF-сети является учет только распределения входных векторов. Метод не учитывает особенности самой аппроксимируемой функции. При этом возможна ситуация, при которой аппроксимируемая функция является более сложной на тех участках, где мало входных обучающих векторов.

3. Варианты заданий и порядок выполнения работы

1. Ознакомиться с особенностями построения и обучения RBF-сетей [2,3,4].
2. Определить номер вариант задания, вычислив остаток от деления номера студента в списке группы на 15.
3. Выполнить с помощью RBF-сети со структурой 1-S-1 аппроксимацию одномерной функции $f(x)$, номер которой соответствует

варианту задания и которая задана в N точках, равномерно распределенных на отрезке:

1. Функция $f(x) = \frac{\sin x \cos x}{x + \cos x}$, N=15, на отрезке $[0, 2\pi]$.
2. Функция $f(x) = \ln(2x - 1)\sqrt{e^{-x} + 4e^x}$, N=10, на отрезке $[0.7, 4]$.
3. Функция $f(x) = (\sin x - 1)\sqrt{e^{-\cos x} + \operatorname{tg} x e^x}$, N=20, на отрезке $[0.05, 1]$.
4. Функция $f(x) = x \sin x + \frac{e^{-x} - e^x}{e^{-x} + e^x}$, N=30, на отрезке $[0, 1]$.
5. Функция $f(x) = x (\sin x + \cos x) \frac{x - \operatorname{ctg} 2x}{1 + \sin^2 x}$, N=10, на отрезке $[0, \pi/2]$.
6. Функция $f(x) = \frac{x}{1 + \sqrt[3]{x+1}}$, N=30, на отрезке $[10, 100]$.
7. Функция $f(x) = \frac{\operatorname{sh} x + \operatorname{ch} x}{\operatorname{th} x + 5}$, N=20, на отрезке $[-3, 3]$.
8. Функция $f(x) = x^3 |\sin x + \cos x| + x - \operatorname{tg} 2x$, N=10, на отрезке $[0, \pi/2]$.
9. Функция $f(x) = \log_2(\sin x + 1)\sqrt{e^{-\cos x}}$, N=30, на отрезке $[0, \pi/2]$.
10. Функция $f(x) = \frac{\sqrt[3]{x^2+1}}{\sqrt{|x|+0.5}} \frac{x}{1 + \sin^2 x}$, N=10, на отрезке $[-\pi/2, \pi/2]$.
11. Функция $f(x) = \frac{\sin^2(2x + \frac{\pi}{2})}{\sqrt{\frac{|x|+1}{4}}} + \frac{x \ln x}{1+x^2}$, N=20, на отрезке $[-\pi/2, \pi/2]$.
12. Функция $f(x) = \frac{\sqrt{(3x+1)^3}}{x+2} + \sin^2(\frac{2x-1}{x+1})$, N=15, на отрезке $[0, 1]$.
13. Функция $f(x) = \sqrt{\log_2(\sin x + 1)} + \sqrt{1 + x^{3.5}}$, N=30, на отрезке $[0, \pi/2]$.
14. Функция $f(x) = 4 \cos^2\left(\frac{x + \frac{\pi}{3}}{x+2}\right) - \ln\left(\frac{e^{-x} - e^x}{e^{-x} + e^x}\right)$, N=20, на отрезке $[0, 1]$.
15. Функция $f(x) = \ln\left(\sqrt{\frac{2x-1}{x+1}}\right) + \sin(\sqrt{2+x})$, N=10, на отрезке $[0.7, 4]$.

4. Порядок выполнения работы:

- 1) определить на языке Scilab [1] заданную функцию и построить её график;
- 2) изобразить архитектуру сети при $S=N$;
- 3) подготовить обучающее множество с числом элементов, равным N;
- 4) выбрать значения весов первого слоя в соответствии с входными данными;
- 5) определить смещения первого слоя, обеспечив адекватное перекрытие базисных гауссовых функций;

6) выполнить вычисление весов и смещений второго слоя в соответствии с алгоритмом LS, используя систему Scilab.

7) выполнить моделирование RBF сети при вычисленных значениях параметров. Для этого определить соответствующую функцию на языке Scilab, назвав её `ann_rbf_run`.

8) сравнить значения функции, вычисляемые по заданной формуле и с помощью модели RBF-сети как в заданных точках N , так и в тех точках, которые не использовались для обучения. Для этого построить соответствующие графики.

9) вычислить среднюю относительную ошибку аппроксимации функции с помощью модели RBF-сети.

10) подготовить отчет по работе в соответствии с пунктами выполнения, указанными выше, защитить отчет.

4. Содержание отчета по работе

4.1 Цель работы.

4.2 Вариант задания, описание используемых расчетных формул.

4.3 Листинги программ с комментариями.

4.4 Архитектура сети, результаты вычислений и графики функций.

4.5 Выводы по результатам работы.

5 Контрольные вопросы

5.1 Зарисуйте архитектуру RBF-сети и объясните назначение слоёв.

5.2 Запишите формулу для вычисления сетевого входа 1-го слоя RBF-сети. Объясните её.

5.3 Чему будет равен сетевой вход RBF нейрона, если его вектор весов равен входному вектору?

5.4 Изобразите и запишите выражение для гауссовой радиальной базисной функции.

5.5 Дайте интерпретацию смещению RBF-нейрона.

5.6 Что можно утверждать об аппроксимирующих свойствах RBF-сети

5.7 Назовите этапы обучения RBF-сети.

5.8 Какие Вы знаете варианты реализации 1-го этапа обучения RBF-сети. Их достоинства и недостатки.

5.9 Запишите целевую функцию RBF-сети со скалярным выходом.

5.10 Приведите формулу для вычисления смещений первого слоя

5.11 Выполните вывод LS алгоритма для нахождения весов второго слоя.

5.12 Запишите формулу LS-решения для RBF-сети и объясните её.

5.13 Как выполнить поиск весов и смещений первого слоя с использованием алгоритмов кластеризации.

5.14 Достоинства и недостатки, использования алгоритмов кластеризации для поиска весов и смещений первого слоя.

Список рекомендованной литературы

1. Ерин С.В. Scilab – примеры и задачи: практическое пособие / С.В. Ерин – М.: Лаборатория «Знания будущего», 2017. – 154 с.
2. Медведев, В.С. Нейронные сети. MATLAB 6 / В.С. Медведев, В.Г. Потемкин; под общ. ред. В.Г. Потемкина. — М.: ДИАЛОГ-МИФИ, 2002. — 496 с.
3. Хайкин С. Нейронные сети: Полный курс. Пер. С англ. / С. Хайкин. — М.: Изд. «Вильямс», 2006. — 1104 с.
4. Hagan M.T. Neural Network Design. The 2nd edition [Электронный ресурс] /М.Т.Hagan, Н.В.Demuth, М.Н.Beale, О.Д. Jesus. . — Frisco, Texas, 2014 . — 1012 p. Режим доступа: <https://www.hagan.okstate.edu/NNDesign.pdf>. —Последний доступ: 14.01.2019. —Название с экрана.