

Байесовский подход к регуляризации задачи обучения сети функций радиального базиса

Аннотация. Рассматривается задача аппроксимации скалярной функции многих переменных с помощью сети функций радиального базиса. Для борьбы с некорректностью задачи аппроксимации используется метод регуляризации Тихонова. Поиск регуляризационного множителя осуществляется по методу Байеса. Предлагаемый алгоритм позволяет существенно сократить вычислительные затраты за счет замены "дорогостоящей" итерационной процедуры поиска экстремума целевой функции аналитическим решением.

Ключевые слова: функции радиального базиса, некорректные задачи, байесовская регуляризация обучения.

Введение

В работе рассматривается задача обучения на примерах – восстановления функции, определяющей зависимость между множествами входных данных X и выходных значений Y , по конечному набору примеров такой зависимости $\{\vec{x}_i, y_i\}_{i=1}^L$, где L – число примеров. Предполагается, что данные зашумлены, т.е. истинное значение функции $f(x_i)$ отличается от y_i на некоторую случайную величину.

Обычно предполагается, что искомая функция аналитически зависит от набора адаптивных параметров $f(\vec{x}, \vec{w})$. В параметрическом подходе, когда число таких параметров невелико, часто для определения их значений можно использовать метод наименьших квадратов:

$$\vec{w} = \arg \min \sum_{i=1}^L (f(\vec{x}_i, \vec{w}) - y_i)^2. \quad (1)$$

Однако в случае большого числа адаптивных параметров, задача минимизации квадратичной ошибки отклонения (1) становится некорректной – имеющей множество решений. Взяв достаточно большое число параметров, мы всегда можем добиться нулевой ошибки отклонения функции на обучающей выборке, но при этом такая функция будет давать большую ошибку на новых, ранее

неизвестных ей данных. В теории машинного обучения такую ситуацию принято называть *переобучением* модели.

Один из основных способов борьбы с переобучением заимствован из теории решения некорректных задач Тихонова [1] – метод регуляризации задачи. В этом подходе целевая функция минимизируется одновременно со стабилизирующим функционалом, который зависит от значений адаптивных параметров $\Omega(\vec{w})$:

$$\vec{w} = \arg \min \sum_{i=1}^L (f(\vec{x}_i, \vec{w}) - y_i)^2 + \lambda \sum_{n=1}^N w_n^2 \quad (2)$$

В выражении (2) стабилизирующий функционал взят в гауссовой форме $\Omega(\vec{w}) = \sum_{n=1}^N w_n^2$, где

N – длина вектора $\vec{w}$, λ – регуляризационный множитель. По сути, правильный выбор значения этого множителя позволяет с одной стороны избежать переобучения модели, с другой – избежать наложения очень сильных ограничений на функцию $f(\vec{x}, \vec{w})$, что может привести к слишком большой ошибке описания данных даже на обучающей выборке (1). В работе используется байесовский подход к выбору регуляризационного параметра λ , который основывается на принципе максимального правдоподобия [2-4].

Далее будет дано описание сетей функций радиального базиса, изложен байесовский подход к задаче регуляризации обучения, приведен метод ортогонализации базиса, позволяющий существенно упростить вычисления.

1. Сеть радиальных базисных функций

Математически сеть радиальных базисных функций (РБФ) [5] можно записать в виде функционального ряда:

$$f(x) = \sum_{n=1}^N w_n \psi_n(\|\vec{x} - \vec{x}_0^n\|),$$

где $\psi_n(\|\vec{x} - \vec{x}_0^n\|)$ – скалярные функции многих переменных, обладающие тем свойством, что их значение зависит только от расстояния до некоторого центра $\vec{x}_0^n$. В качестве примеров представителей данного семейства функций можно привести:

полиномиальные функции

$$\psi_n(\|\vec{x} - \vec{x}_0^n\|) = (\vec{x} - \vec{x}_0^n)^k,$$

мультиквадратичную функцию

$$\psi_n(\|\vec{x} - \vec{x}_0^n\|) = \sqrt{1 + \frac{(\vec{x} - \vec{x}_0^n)^2}{\sigma_n^2}},$$

обратную мультиквадратичную функцию

$$\psi_n(\|\vec{x} - \vec{x}_0^n\|) = \frac{1}{\sqrt{1 + \frac{(\vec{x} - \vec{x}_0^n)^2}{\sigma_n^2}}},$$

функцию Гаусса

$$\psi_n(\|\vec{x} - \vec{x}_0^n\|) = \exp\left(-\frac{(\vec{x} - \vec{x}_0^n)^2}{\sigma_n^2}\right)$$

Большинство из часто используемых сетей функций радиального базиса имеют три типа адаптивных параметров – центры функций $\vec{x}_0^n$, параметры, отвечающие за “ширины” функций σ_n и коэффициенты разложения решения по набору базисных функций w_n . Для задания центров $\vec{x}_0^n$ часто используются кластерные методы, такие, например, как K-means или кар-

ты Кохонена [6]. В результате кластеризации каждый входной вектор обучающей выборки $\vec{x}$ будет приписан к своему кластеру, который характеризуется центром $\vec{x}_0^n$ (вектор относится к тому кластеру, чей центр ближе всего к нему по евклидову расстоянию). Взяв в качестве центра радиальной базисной функции центр кластера, параметр ширины можно задать исходя из других характеристик кластера. Так, например, для сети гауссовых функций параметр σ_n часто берут равным дисперсии точек, относящихся к n -му кластеру.

Эти два параметра задают локализацию и ширину радиальной базисной функции. После их определения задача сводится к нахождению коэффициентов разложения w_n . Поиск этих коэффициентов, основанный на идее метода наименьших квадратов (1), как было ранее сказано, является некорректной задачей и может привести к переобучению. Для борьбы с некорректностью можно провести регуляризацию задачи (2), а регуляризационный множитель определить по методу Байеса, который будет описан ниже.

2. Байесовский подход к регуляризации обучения

В Байесовском подходе выбор решения f ведется в рамках некоторой модели H – набора ограничивающих условий. Решение выбирается как наиболее вероятное по формуле Байеса:

$$P(f|D, H) = \frac{P(D|f, H)P(f|H)}{P(D|H)} \quad (3)$$

Здесь $P(f|D, H)$ – вероятность выбора функции f для описания данных D в рамках модели H , $P(D|f, H)$ – вероятность того, что функция f генерирует данные D ; $P(f|H)$ – некоторая априорная вероятность выбора f в рамках модели H . Коэффициент $P(D|H)$ необходим, чтобы нормировать множество значений произведения $P(D|f, H)P(f|H)$ на единицу, придав ему, таким образом, смысл вероятности.

Насколько подходит тот или иной набор ограничивающих условий H для описания данных определяет как раз знаменатель формулы Байеса [2]:

$$P(D|H) = \sum_f P(D|f, H) P(f|H) = \sum_f P(D, f|H)$$

В литературе он получил название *Evidence*. Чем больше *Evidence*, тем соответствующая модель более предпочтительна [2].

Сделав несколько предположений о внешнем виде вероятностей в формуле (3), мы сможем найти параметр регуляризации исходя из принципа максимального правдоподобия. Предположим, что данные зашумлены гауссовым шумом, тогда вероятность порождения функцией $f(x)$ данных $D = \{y_i, \bar{x}_i\}_{i=1}^L$ можно оценить следующим выражением:

$$P(D|f) = \frac{1}{Z_X} \exp\left(-\beta \sum_{i=1}^L (y_i - f(\bar{x}_i))^2\right).$$

Возьмем стабилизирующий функционал в гауссовой форме

$$\Omega = \sum_{n=1}^N w_n^2$$

Ему соответствует априорная вероятность [3,4]

$$P(f|H) = \frac{1}{Z_A} \exp\left(-\alpha \sum_{n=1}^N w_n^2\right).$$

Максимизация вероятности

$$P(f|D, H) = \frac{1}{Z_M} \exp(-M),$$

где $M = \beta \sum_{i=1}^L (y_i - f_i)^2 + \alpha \sum_{n=1}^N w_n^2$ сводится к решению системы линейных алгебраических уравнений (СЛАУ)

$$\sum w_m A_{mn} = B_n, \quad (4)$$

где $A_{mn} = \beta \sum_{i=1}^L \psi_{mi} \psi_{ni}, n \neq m,$

$$A_{nn} = \beta \sum_{i=1}^L \psi_{ni}^2 + \alpha, B_n = \beta \sum_{i=1}^L y_i \psi_{ni}$$

В качестве значений α и β берутся те, при которых *Evidence* максимально. Таким образом, чтобы их определить, можно, например, получить аналитическое выражение для *Evidence*, а

затем градиентными методами найти его максимум. Легко показать, что

$$\ln P(D|H) = \ln Z_M - \ln Z_A - \ln Z_X,$$

где

$$Z_A = \int_{-\infty}^{+\infty} \exp\left(-\alpha \sum_{n=1}^N w_n^2\right) d^N w_n = \left(\frac{2\pi}{\alpha}\right)^{\frac{N}{2}},$$

$$Z_X = \left(\frac{\pi}{\beta}\right)^{\frac{L}{2}}$$

$$Z_M = \int d^N w_n \times \exp\left(-\beta \sum_{n=1}^N \left(\bar{y} - \sum_{n=1}^N w_n \bar{\psi}_n\right)^2 - \alpha \sum_{n=1}^N w_n^2\right),$$

здесь через $\bar{\psi}_n$ обозначены векторы значений n -ой базисной функции в точках $\{\bar{x}_i\}_{i=1}^L$, а через $\bar{y}$ – вектор значений искомой функции в этих точках.

Последнее выражение считается приближенно по методу перевала. Подробности вычисления этого интеграла описаны в работе [3]. В результате приходим к следующему приближенному выражению для *Evidence*:

$$\begin{aligned} \ln P(D|H) = & \\ = \sum_{n=1}^N \left(\frac{(\beta \bar{y} \bar{\psi}_n)^2}{A_{nn}} - \frac{1}{2} \ln A_{nn} - \beta \sum_{m \neq n} \bar{\psi}_n \bar{\psi}_m w_n w_m \right) + & \\ + \frac{N}{2} \ln \alpha - \beta \bar{y}^2 + \frac{L}{2} \ln \beta - \frac{L}{2} \ln \pi & \quad (5) \end{aligned}$$

Данный функционал содержит коэффициенты разложения $\{w_n\}_{n=1}^N$, для нахождения которых из (4), в свою очередь, необходимо знать значения параметров α и β . В таких случаях обычно пользуются ЕМ-алгоритмом [7]:

- На первом шаге некоторым образом (возможно, случайным) фиксируют параметры модели α, β и вычисляют значения $\{w_n\}_{n=1}^N$ из (4).

- На втором шаге находят новые параметры α и β путем поиска максимума функционала (5) при текущих значениях коэффициентов разложения $\{w_n\}_{n=1}^N$.

Далее, берут параметры модели, полученные на втором шаге, и используют их для получения новых коэффициентов разложения и т.д. При

многократном повторении этих шагов алгоритм сходится к некоторому решению $f(\bar{x})$ [7].

3. Случай ортонормированных сверток базисных функций

Вычислительная процедура существенно упростится, в случае, когда свертки значений базисных функций будут ортонормированы:

$$\sum_{i=1}^L \psi_{mi} \psi_{ni} = \delta_{m,n} \quad (6)$$

При этом:

1. Дорогостоящая процедура решения системы линейных алгебраических уравнений (4) (число операций пропорционально числу базисных функций в третьей степени $\sim N^3$) заменяется более простой, трудоемкость которой растет линейно от числа базисных функций $\sim N$:

$$w_n = \frac{\beta}{\beta + \alpha} \sum_{i=1}^L y_i \psi_{ni} = \frac{\beta}{\beta + \alpha} \bar{y} \bar{\psi}_n \quad (7)$$

2. Выражение для *Evidence* (5) вычисляется точно [4].

$$\ln P(D|H) = \frac{\beta^2 S}{\beta + \alpha} + \frac{N}{2} \ln \frac{\alpha}{(\beta + \alpha)} - \beta \bar{y}^2 + \frac{L}{2} \ln \frac{\beta}{\pi}, \quad (8)$$

где $S = \sum_{n=1}^N (\bar{y} \bar{\psi}_n)^2$,

3. Логарифм *Evidence* (8) перестает зависеть от коэффициентов разложения, что позволяет не прибегать к ЕМ-алгоритму, а провести обучение в одну итерацию.

4. Функционал (8) имеет только одну точку максимума, и соответствующие ей значения параметров находятся аналитически [4]:

$$\alpha = \frac{1}{2} \frac{(L - N)}{\left(\frac{L}{N} S - \bar{y}^2\right)}, \quad \beta = \frac{1}{2} \frac{(L - N)}{\bar{y}^2 - S},$$

$$\lambda = \frac{\alpha}{\beta} = \frac{\bar{y}^2 - S}{\left(\frac{L}{N} S - \bar{y}^2\right)} \quad (9)$$

Параметры α и β должны быть положительными, иначе выражения для вероятностей теряют свой смысл. Это условие выполняется, когда $L > N$, что всегда существует при линейной независимости векторов $\{\bar{\psi}_n\}_{n=1}^N$. Требование линейной независимости в нашем случае выражено в условии (6).

4. Метод ортогонализованных базисных функций

Общий случай неортонормированных сверток базисных функций всегда можно свести к ортогональному базису, при условии их линейной независимости (6). В этом случае векторы значений функции представляют собой базис в N -мерном пространстве.

Обозначим через $\hat{\Psi}$ матрицу, столбцы которой составлены из векторов $\{\bar{\psi}_n\}_{n=1}^N$. В этом случае матрица $\hat{\Psi}^T \hat{\Psi}$ будет квадратной симметричной и положительно определенной.

Решим задачу на собственные векторы $\bar{u}_n$ и собственные значения Λ_n матрицы $\hat{\Psi}^T \hat{\Psi}$:

$$\hat{\Psi}^T \hat{\Psi} \bar{u}_n = \Lambda_n \bar{u}_n$$

Множество векторов

$$\bar{\omega}_n = \frac{\bar{u}_n \hat{\Psi}}{\sqrt{\Lambda_n}} \quad (10)$$

также будут базисом в том же пространстве, где и $\{\bar{\psi}_n\}_{n=1}^N$ (предполагается, что собственные векторы нормированы на единицу). При этом эти векторы будут отвечать условию

$$\sum_{i=1}^L \omega_{mi} \omega_{ni} = \delta_{m,n} \quad (11)$$

Таким образом, мы можем найти коэффициенты разложения искомого решения по базисным функциям, представленным значениями в точках обучающей выборки $\{\bar{\omega}_n\}_{n=1}^N$, используя формулы (7) и (9). После этого, через полученные коэффициенты разложения $\tilde{w}$ можно выразить коэффициенты разложения искомой функции через исходный базис:

$$\vec{w} = \vec{\tilde{w}} \hat{U}^T, \quad (12)$$

где $\hat{U}$ – матрица, столбцы которой составлены из следующих векторов $\frac{\vec{u}_n}{\sqrt{\Lambda_n}}$

5. Описание алгоритма

Теперь можно описать последовательность действий необходимую для построения функции, аппроксимирующей точки $\{\vec{x}_i, y_i\}_{i=1}^L$, с помощью сети радиальных базисных функций:

1. Задаем число радиальных базисных функций как внешний параметр модели N .

2. Проводим кластерный анализ входного множества обучающих данных X , разбиваем его на N кластеров, из которых вычисляем N центров радиальных базисных функций и N параметров, определяющих их “ширины”.

3. Вычисляем значения радиальных базисных функций $\{\psi_n(x)\}_{n=1}^N$ в точках обучающей выборки $\{\vec{x}_i\}_{i=1}^L$.

4. Проверяем базис на линейную независимость. Если полученная система векторов является линейно зависимой, то необходимо избавиться от некоторых функций – исключить из базиса минимально достаточное число функций, чтобы он стал линейно независим.

5. Переходим к ортогональному базису (10)

6. Находим значения α и β по формулам (9), где значение S вычисляется через новые ортогональные векторы:

$$S = \sum_{n=1}^N (\vec{y} \vec{\omega}_n)^2$$

7. Подставляя α и β в выражение для коэффициентов разложения

$$\vec{w}_n = \frac{\beta}{\beta + \alpha} \vec{y} \vec{\omega}_n$$

Находим, таким образом, коэффициенты разложения $\vec{w} = \{\vec{w}_n\}_{n=1}^N$ и выражение для функции $f(\vec{x})$ через ортогонализированный базис $\{\vec{\omega}_n\}_{n=1}^N$.

8. По формуле (12) получаем коэффициенты $\vec{w}$ и переходим к выражению для $f(\vec{x})$ через исходный базис:

$$f(x) = \sum_{n=1}^N \vec{w}_n \psi_n(x)$$

6. Эквивалентность ортогонализированных базисов

Выше был предложен вариант построения матрицы перехода к ортогональному базису через решение задачи на собственные векторы и собственные значения. Однако можно построить бесконечное число линейных преобразований матрицы $\hat{\Psi}$, приводящих к ортогональным базисам. Можно показать, что конечный результат (коэффициенты разложения по исходным функциям) не зависит от выбора того или иного преобразования.

Действительно, пусть $\hat{\tilde{U}}$ – другое линейное преобразование, приводящее исходный базис к ортонормированному. Эта матрица будет линейно выражаться через $\hat{U}$

$$\hat{\tilde{U}} = \hat{U} \hat{Q}$$

Поскольку матрица $\hat{Q}$ одновременно является матрицей перехода от одного ортонормированного базиса к другому, то она должна подчиняться условию:

$$\hat{Q}^T \hat{Q} = \hat{I}$$

Проверим неизменность параметров регуляризации. Матрица поворота входит в формулы

(9) через величину $S = \sum_{n=1}^N (\vec{y} \vec{\omega}_n)^2$. Пусть S –

величина, полученная с использованием преобразования $\hat{U}$, а $\tilde{S}$ – полученная с использованием преобразования $\hat{\tilde{U}}$.

$$\begin{aligned} S &= \sum_{n=1}^N (\vec{y} \vec{\omega}_n)^2 = \vec{y} (\hat{\Psi} \hat{U}) (\hat{\Psi} \hat{U})^T \vec{y}^T = \vec{y} \hat{\Psi} \hat{U} \hat{U}^T \hat{\Psi}^T \vec{y}^T \\ \tilde{S} &= \sum_{n=1}^N (\vec{y} \vec{\tilde{\omega}}_n)^2 = \vec{y} \hat{\tilde{\Psi}} \hat{\tilde{U}} \hat{\tilde{U}}^T \hat{\tilde{\Psi}}^T \vec{y}^T = \\ &= \vec{y} \hat{\tilde{\Psi}} \hat{U} \hat{Q} \hat{Q}^T \hat{U}^T \hat{\tilde{\Psi}}^T \vec{y}^T = \vec{y} \hat{\tilde{\Psi}} \hat{U} \hat{U}^T \hat{\tilde{\Psi}}^T \vec{y}^T = S \end{aligned} \quad (13)$$

Здесь $\vec{\tilde{\omega}}_n$ – ортогонализированные базисные функции, выраженные через матрицу $\hat{\tilde{U}}$, а $\vec{\tilde{\omega}}_n$ – соответственно, через матрицу $\hat{U}$.

Таким образом, из (13) видно, что величины S и $\tilde{S}$ эквивалентны, следовательно, параметры (9) не зависят от способа ортогонализации базиса.

Рассмотрим выражение для коэффициентов разложения (7). Выразим коэффициенты разложения по ортогональному базису через исходные функции:

$$\tilde{w} = \frac{\beta}{\beta + \alpha} \bar{y}(\hat{\Psi}\hat{U})$$

Подставим это выражение в формулу (12)

$$\bar{w} = \frac{\beta}{\beta + \alpha} \bar{y}\hat{\Psi}\hat{U}\hat{U}^T$$

Аналогично, получаем

$$\begin{aligned} \bar{w} &= \frac{\beta}{\beta + \alpha} \bar{y}\hat{\Psi}\hat{U}\hat{U}^T = \frac{\beta}{\beta + \alpha} \bar{y}\hat{\Psi}\hat{U}\hat{Q}\hat{Q}^T\hat{U}^T = \\ &= \frac{\beta}{\beta + \alpha} \bar{y}\hat{\Psi}\hat{U}\hat{U}^T \end{aligned}$$

Таким образом, приходим к эквивалентности всех линейных преобразований, приводящих к ортогональному базису.

7. Эксперименты

В качестве примера рассмотрим сеть гауссовых функций радиального базиса

$$\psi_n(\|\bar{x} - \bar{x}_0^n\|) = \exp\left(-\frac{(\bar{x} - \bar{x}_0^n)^2}{\sigma_n^2}\right)$$

Проведем аппроксимацию модельных данных с помощью сети таких функций. Для удобства иллюстрации возьмем функцию двух переменных $F(x^1, x^2)$. Сгенерируем в квадрате $-\frac{A}{2} \leq x^1 \leq \frac{A}{2}$; $-\frac{A}{2} \leq x^2 \leq \frac{A}{2}$ L пар значений $\{x_i^1, x_i^2\}_{i=1}^L$. С помощью функции $F(x^1, x^2)$ вычислим значения аппроксимируемой функции $\{y\}_{i=1}^L$ в точках $\{x_i^1, x_i^2\}_{i=1}^L$.

Зададим число базисных функций N , как внешний параметр модели. Далее, пошагово проделаем процедуру, описанную в пункте 5.

- Задаем центры радиальных базисных функций как центры кластеров, полученных путем кластеризации данных по Кохонену.

- В качестве параметра σ_n^2 , берем значение дисперсии данных, попавших в соответствующий кластер.

- Находим коэффициенты w_n .

В качестве функции генерирующей данные было взято следующее выражение:

$$F = \exp\left(-\frac{(x^1)^2 + (x^2)^2}{2A^2}\right) \cos^2\left(5\frac{x^1}{A}\right) \cos^2\left(5\frac{x^2}{A}\right),$$

где параметр $A = 5$. С помощью этой функции было сгенерировано $L = 1000$ точек $\{x_i^1, x_i^2, y_i\}_{i=1}^L$.

Первоначальное число базисных функций $N = 500$.

Параметр ширины для каждой функции вычислялся по формуле:

$$\sigma_n^2 = \frac{1}{L_n} \sum_{i_n} (\bar{x}_{i_n} - \bar{x}_0^n)^2,$$

где i_n – индексы точек, попавших в n -ый кластер, а L_n – число точек в этом кластере.

Ниже на Рис.1 показан результат аппроксимации двумерной функции по модельным данным. Символом + на нем обозначены исходные данные – точки, по которым проводилась аппроксимация функции. В узлах регулярной сетки на рисунке приведены значения восстановленной функции $f(x)$.

Заключение

В работе рассматривалась задача обучения сети функций радиального базиса на примерах. Для выбора центров и параметров “ширины” функций использовался кластерный анализ.

Для борьбы с некорректностью задачи выбора коэффициентов разложения искомого решения по радиальным базисным функциям использовался метод регуляризации Тихонова. Выбор регуляризационного множителя осуществлялся по методу Байеса. Предложенный в работе метод Байесовской регуляризации обладает рядом существенных преимуществ.

1. Не требует итерационных процедур обучения наподобие градиентного спуска или ЕМ-алгоритма.

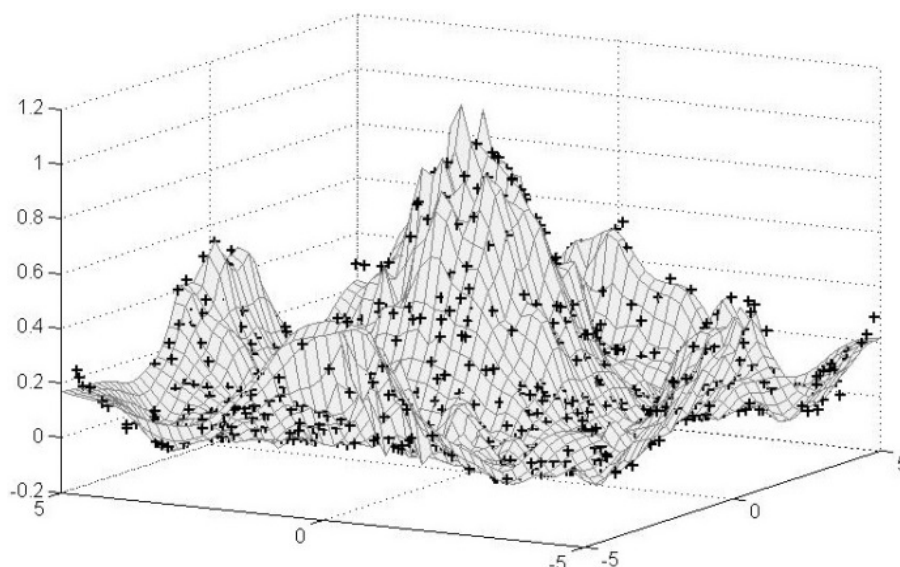


Рис.1. Аппроксимация двумерных данных сетью гауссовых радиальных базисных функций

2. Дает единственное точное аналитическое решение для параметра регуляризации.

3. Экономичен с вычислительной точки зрения.

Для иллюстрации алгоритма была проведена аппроксимация скалярной функции двух переменных. При этом использовались гауссовы функции радиального базиса.

Литература

1. А.Н. Тихонов, В.Я. Арсенин. Методы решения некорректных задач. М.: Наука, 1979. Изд. 2-е.
2. MacKay D. Bayesian interpolation//Neural Computation. – 1992. – v.4. – pp.415–447.

3. А.С. Нужный, С.А. Шумский. “Регуляризация Байеса в задаче аппроксимации функции многих переменных”, Мат. моделирование, 15:9 (2003), 55–63.
4. А.С. Нужный. “Байесовская регуляризация в задаче аппроксимации функции по точкам с помощью ортогонализованного базиса”, Мат. моделирование, 23:9 (2011), 33–42.
5. Buhmann M.D. (2003). Radial Basis Functions: Theory and Implementations Cambridge University Press: Cambridge.
6. Kohonen T. "Self-Organizing Maps" (2-nd edition), Springer, 1997.
7. Dempster A.P., Laird N.M., Rubin D.B. (1977) “Maximum likelihood from incomplete data via the EM algorithm”, Journal of the Royal Statistical Society (B). vol. 39. No.1. pp. 1-38.

Нужный Антон Сергеевич. Научный сотрудник Института проблем безопасного развития атомной энергетики РАН. Окончил Московский инженерно-физический институт в 2002 году. Кандидат физико-математических наук. Автор 33 публикаций. Область научных интересов: методы машинного обучения, теория некорректных задач. E-mail: nuzhny@inbox.ru