

## 8. ЛЕКЦИЯ. Стратегии обучения на основе RBF

Процесс обучения сети на основе радиальных базисных функций (RBF) без учета его теоретического обоснования можно рассматривать следующим образом. Линейные веса, связанные с выходным узлом (узлами) сети, могут изменяться во «временном масштабе», отличном от используемого при работе нелинейных функций активации скрытых элементов. Если функции активации скрытых нейронов изменяются медленно, то веса выходных элементов изменяются довольно быстро с помощью линейной стратегии оптимизации. Здесь важно отметить, что разные слои сети RBF выполняют разные задачи, поэтому будет целесообразным отделить процесс оптимизации скрытого и выходного слоев друг от друга и использовать для них разные методы и, возможно, даже разные масштабы времени.

Существует множество различных стратегий обучения сети, зависящих от способа определения центров радиальных базисных функций. Первые три стратегии применимы к сетям RBF, описание которых основано на теории интерполяции (*способ нахождения промежуточных значений величины по имеющемуся дискретному набору известных значений*). Последняя стратегия создания сетей сочетает в себе элементы теории регуляризации и оценки регрессии ядра.

### ***Случайный выбор фиксированных центров***

Простейший из подходов предполагает использование фиксированных радиальных базисных функций, определяющих функции активации скрытых элементов. Размещение центров может быть выбрано случайным образом из множества данных примеров.

Такой подход считается «чувствительным» и требует представительного распределения множества обучающих данных с учетом рассматриваемой задачи. Что же касается самих радиальных базисных функций, то для их реализации можно задействовать изотропные функции Гаусса, стандартное отклонение которых является фиксированным относительно разброса центров. В частности, (нормализованные) радиальные базисные функции с центром в точке  $t_i$  определяются выражением

$$G(\|x - t_i\|^2) = \exp\left(\frac{m_1}{d_{\max}^2} \|x - t_i\|^2\right), i = 1, 2, \dots, m_1$$

где  $m_1$  – количество центров;  $d_{\max}$  – максимальное расстояние между выбранными центрами. В результате стандартное отклонение (т.е. ширина) всех радиальных базисных функций Гаусса будет фиксированным:

$$\sigma = \frac{d_{\max}}{\sqrt{2m_1}}$$

Эта формула гарантирует, что отдельные радиальные базисные функции не будут слишком гладкими или слишком остроконечными. Обоих этих крайних случаев следует избегать. В качестве альтернативы последнему выражению в наиболее разреженных областях можно использовать центры с большей шириной, что требует эксперимент с данными обучения.

При использовании этого подхода единственными параметрами, которые настраиваются в процессе обучения сети, являются синоптические веса ее выходного слоя. Их проще всего настроить с помощью метода псевдоинверсии. В частности:

$$w = G^+ d$$

где  $d$  – вектор желаемого отклика для множества примеров. Матрица  $G^+$  является псевдообратной матрице  $G$ , которая, в свою очередь, определяется следующим образом:

$$G = \{g_{ji}\}$$

где

$$g_{ji} = \exp\left(\frac{m_1}{d^2} \|x_j - t_i\|^2\right), j = 1, 2, \dots, N; i = 1, 2, \dots, m_1$$

где  $x_j$  –  $j$ -й входной вектор множества примеров обучения.

Основой всех алгоритмов вычисления псевдообратных матриц является сингулярная декомпозиция (*определённого типа разложение прямоугольной матрицы, имеющее широкое применение, в силу своей наглядной геометрической интерпретации, при решении многих прикладных задач.*).

Если  $G$  – действительная матрица размерности  $N \times M$ , то существуют ортогональные матрицы

$$U = \{u_1, u_2, \dots, u_N\} \text{ и } V = \{v_1, v_2, \dots, v_M\},$$

такие, что

$$U^T C V = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_K), K = \min(N, M),$$

$\text{diag}$  – диагональная матрица, квадратная матрица, все элементы которой, стоящие вне главной диагонали, равны нулю.

где

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_K > 0$$

Векторы-столбцы матрицы  $U$  называют левыми сингулярными векторами) матрицы  $G$ , а векторы-столбцы матрицы  $V$  – правыми сингулярными векторами. Числа  $\sigma_1, \sigma_2, \dots, \sigma_K$  называют сингулярными значениями матрицы  $G$ . Согласно теореме о декомпозиции по сингулярным значениям, матрица, псевдообратная

матрице  $G$ , размерности  $N \times M$  определяется следующим образом:

$$G^+ = V\Sigma^+U^T$$

где  $\Sigma^+$  - матрица размерности  $N \times N$ , выраженная в терминах сингулярных значений матрицы  $G$ :

$$\Sigma^+ = \text{diag}\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \dots, \frac{1}{\sigma_k}, 0, \dots, 0\right)$$

Опыт использования методики случайного выбора центров показал, что этот метод относительно нечувствителен к использованию регуляризации (метод добавления некоторых дополнительных ограничений к условию с целью решить некорректно поставленную задачу или предотвратить переобучение). Случайный выбор центров можно использовать в качестве метода построения сетей RBF на основе множества примеров большого объема с возможным применением регуляризации.

### ***Выбор центров на основе самоорганизации***

Основной проблемой метода случайного выбора фиксированных центров является тот факт, что для обеспечения удовлетворительного уровня эффективности он требует большого множества примеров. Одним из способов обхода этой проблемы является использование гибридного процесса обучения, состоящего из двух этапов.

- *Этап обучения на основе самоорганизации.* Его целью является оценка подходящих положений центров радиальных базисных функций скрытого слоя.
- *Этап обучения с учителем.* На этом этапе создание сети завершается оценкой линейных весов выходного слоя.

Хотя для реализации этих двух этапов обучения можно применить пакетную обработку, все-таки лучше использовать адаптивный (итеративный) подход. Для процесса обучения на основе самоорганизации требуется разработать алгоритм кластеризации, разбивающий заданное множество точек данных на две подгруппы, каждая из которых должна быть максимально однородной. Один из таких алгоритмов называется алгоритм кластеризации  $k$  – средних. Согласно этому алгоритму центры радиальных базисных функций размещаются только в тех областях входного пространства  $X$ , в которых имеются информативные данные. Пусть  $m_1$  – количество радиальных базисных функций. Определение подходящего значения для  $m_1$  требует проведения некоторых экспериментов. Пусть  $\{t_k(n)\}_{k=1}^{m_1}$  – центры радиальных базисных функций на  $n$ -й итерации алгоритма. Тогда алгоритм кластеризации по  $k$ -средним можно описать следующим образом.

1. Инициализация. Выбираем случайные значения для исходных центров  $t_k(0)$ . Единственным требованием к их выбору на данном шаге является различие всех начальных значений. При этом значения Евклидовой нормы по возможности должны быть небольшими.

2. Выборка. Выбираем вектор  $x$  из входного пространства  $X$  с определенной вероятностью. Этот вектор будет входным для алгоритма на итерации  $n$ .

3. Проверка подобия. Обозначим  $k(x)$  индекс наиболее подходящего (победившего) центра для данного вектора  $x$ . Находим  $k(x)$  на итерации  $n$  используя критерий минимального Евклидова расстояния:

$$k(x) = \operatorname{argmin}_k \|x(n) - t_k(n)\|, \quad k = 1, 2, \dots, m_1.$$

где  $t_k(n)$  – центр  $k$ -й радиально базисной функции на итерации  $n$ .

4. Корректировка. Корректируем центры радиальных базисных функций, используя следующее правило:

$$t_k(n+1) = \begin{cases} t_k(n) + \eta[x(n) - t_k(n)], & k = k(x) \\ t_k(n) & \text{в противном случае} \end{cases}$$

$\eta$  – параметр скорости обучения, выбранный из диапазона  $0 < \eta < 1$ .

5. Продолжение. Увеличиваем на единицу значение  $n$  и возвращаемся к шагу 2, продолжая процедуру до тех пор, пока положение центров  $t_k$  существенно изменяется.

Этот алгоритм целесообразно реализовывать на стадии обучения без учителя (на основе самоорганизации).

Ограничением алгоритма кластеризации по  $k$ -средним является нахождение только локального оптимального решения, зависящего от исходного выбора центров кластера. Следовательно, вычислительные ресурсы могут расходоваться напрасно: отдельные центры изначально попадут в те области входного пространства  $X$ , где количество точек данных невелико и откуда не будет шанса переместиться в области, требующие большего количества центров. В результате можно получить неоправданно большую сеть. Чтобы обойти этот недостаток обычного алгоритма кластеризации по  $k$ -средним, в 1995 году был предложен улучшенный алгоритм кластеризации по  $k$ -средним, который основан на понятии взвешенной переменной меры принадлежности кластеру, обеспечивающем сходимость алгоритма к оптимальной или квазиоптимальной (близкий к оптимальному; такой, который может использоваться в качестве оптимального в практических целях) конфигурации, независимо от исходного положения центров.

Определив отдельные центры гауссовых радиальных базисных функций и

их общий вес с помощью алгоритма кластеризации по k-средним или его улучшенной версии, можно перейти к следующему (и последнему) этапу процесса гибридного обучения – оценке весов выходного слоя. Простейшим методом такой оценки является алгоритм LMS (least-mean-square, алгоритм минимизации среднеквадратической ошибки).

Напомню краткое описание алгоритма минимизации среднеквадратической ошибки (рис. 1).

<i>Обучающий пример</i>	Вектор входного сигнала = $\mathbf{x}(n)$ Желаемый отклик = $d(n)$
<i>Выбираемый пользователем параметр</i>	$\eta$
<i>Инициализация весов</i>	$\hat{\mathbf{w}}(0) = 0$
<i>Вычислительная схема</i>	Для $n = 1, 2, \dots$ $e(n) = d(n) - \mathbf{x}^T(n)\hat{\mathbf{w}}(n),$ $\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \eta \mathbf{x}(n)e(n)$

Рис. 1. Алгоритм минимизации среднеквадратической ошибки

Вектор выходного сигнала, сгенерированного скрытыми узлами, является входным вектором алгоритма LMS. Обратите внимание на то, что алгоритм кластеризации по k-средним для скрытых узлов и алгоритм LMS для выходных узлов могут выполнять вычисления параллельно. Таким образом, процесс обучения ускоряется.

### ***Выбор центров с учителем***

В рамках третьего подхода центры радиальных базисных функций и все остальные свободные параметры сети настраиваются в процессе обучения с учителем. Другими словами, сеть RBF принимает самый общий вид. Естественным выбором для такой ситуации является обучение на основе коррекции ошибок, которое удобнее всего реализовывать с помощью процедуры градиентного спуска, являющейся обобщением алгоритма LMS.

Первым шагом в разработке такой процедуры обучения является определение значения функции стоимости:

$$E = \frac{1}{2} \sum_{j=1}^N e_j^2$$

где  $N$  – размер выборки, использованной для обучения;  $e_j$  – сигнал ошибки следующего вида:

$$e_j = d_j - F^*(x_j) = d_j - \sum_{i=1}^M \omega_i G(\|x_j - t_i\|_{C_i})$$

Требуется найти свободные параметры  $\omega_i$ ,  $t_i$  и  $\sigma_i^{-1}$  (последний связан с

матрицей взвешивания нормы  $C_i$ ), минимизирующие функцию стоимости  $E$ . Результаты минимизации приведены на рис. 2.

---

1 Линейные веса (выходной слой)

$$\frac{\partial E(n)}{\partial w_i(n)} = \sum_{j=1}^N e_j(n) G(\|\mathbf{x}_j - \mathbf{t}_i(n)\|_{C_i}),$$

$$w_i(n+1) = w_i(n) - \eta_1 \frac{\partial E(n)}{\partial w_i(n)}, \quad i = 1, 2, \dots, m_1$$

2 Позиции центров (скрытый слой)

$$\frac{\partial E(n)}{\partial \mathbf{t}_i(n)} = 2w_i(n) \sum_{j=1}^N e_j(n) G'(\|\mathbf{x}_j - \mathbf{t}_i(n)\|_{C_i}) \sigma_i^{-1} [\mathbf{x}_j - \mathbf{t}_i(n)],$$

$$\mathbf{t}_i(n+1) = \mathbf{t}_i(n) - \eta_2 \frac{\partial E(n)}{\partial \mathbf{t}_i(n)}, \quad i = 1, 2, \dots, m_1$$

3 Распределение центров (скрытый слой)

$$\frac{\partial E(n)}{\partial \sigma_i^{-1}(n)} = -w_i(n) \sum_{j=1}^N e_j(n) G'(\|\mathbf{x}_j - \mathbf{t}_i(n)\|_{C_i}) \sigma_i^{-1} \mathbf{Q}_{ji}(n),$$

$$\mathbf{Q}_{ji}(n) = [\mathbf{x}_j - \mathbf{t}_i(n)][\mathbf{x}_j - \mathbf{t}_i(n)]^T,$$

$$\sigma_i^{-1}(n+1) = \sigma_i^{-1}(n) - \eta_3 \frac{\partial E(n)}{\partial \sigma_i^{-1}(n)},$$

где  $e_j(n)$  – сигнал ошибки выходного узла  $j$  в момент времени  $n$ ;  $G'(\cdot)$  – первая производная функции Грина  $G(\cdot)$  по своему аргументу

---

Рис.2. Формулы настройки линейных весов, положений и распределения центров для сетей RBF

При рассмотрении рис.2 особого внимания заслуживают следующие моменты.

- Функция стоимости  $E$  является выпуклой по линейному параметру  $\omega_i$ , однако не выпуклой по отношению к центрам  $t_i$  и матрице  $\sigma_i^{-1}$ . В последних случаях поиск оптимального значения  $t_i$  и матрицы  $\sigma_i^{-1}$  может остановиться в точке локального минимума пространства параметров.

- В формулах для модификации значений  $\omega_i$ ,  $t_i$  и  $\sigma_i^{-1}$  в общем случае можно использовать разные параметры скорости обучения  $\eta$ :  $\eta_1$ ,  $\eta_2$  и  $\eta_3$ .

- В отличие от алгоритма обратного распространения, процедура градиентного спуска (см. рис. 2) для сети RBF не предполагает обратного распространения сигнала ошибки.

- Вектор градиента  $\partial E / \partial t_i$  обладает свойством, аналогичным свойству алгоритма кластеризации, – его значение зависит от конкретной задачи.

Для инициализации процедуры градиентного спуска поиск в пространстве параметров желательно начинать с некоторого структурированного начального условия, которое ограничивает область поиска уже известной полезной областью. Этого можно достичь с помощью стандартного метода

классификации. Таким образом, вероятность сходимости к нежелательному локальному минимуму в пространстве весов уменьшается. Например, можно начать с гауссова классификатора, использование которого предполагает, что примеры всех классов имеют распределение Гаусса.

Возникает вопрос: чего можно добиться с помощью настройки положения центров радиальных базисных функций? Ответ на этот вопрос зависит от конкретной задачи. Тем не менее на основе опубликованных в литературе результатов можно сделать вывод, что идея настройки местоположения центров имеет некоторые преимущества. Исследования показывают:

- Сети RBF с самонастройкой (без учителя) положения центров и адаптацией (с учителем) весов выходного слоя не обеспечивают такой эффективности обобщения, как многослойный персептрон, обучаемый с помощью алгоритма обратного распространения.
- Обобщенные сети RBF (в которых обучение с учителем применяется как для настройки положения центров скрытого слоя, так и для адаптации весов выходного слоя) могут достичь уровня производительности многослойного персептрона.

### *Строгая интерполяция с регуляризацией*

Метод построения сетей RBF, объединяющий в себе элементы теории регуляризации и теории оценки регрессии ядра подразумевает использование четырех составляющих.

1. Радиальные базисные функции  $G$ , рассматриваемые (иногда с масштабированием) как ядро состоятельной оценки регрессии Надарайа-Ватсона (NWRE).
2. Диагональная матрица взвешивания нормы  $\Sigma^{-1}$ , общая для всех центров:

$$\Sigma = \text{diag}(h_1, h_2, \dots, h_{m_o}),$$

где  $h_1, h_2, \dots, h_{m_o}$  – ширина полос по отдельным измерениям NWRE с (масштабированным) ядром  $G$ ;  $m_o$  – размерность входного пространства. Например, можно принять  $h_i = \alpha_i \sigma_i^2$ ; где  $\sigma_i^2$  – дисперсия  $i$ -й входной переменной, вычисленная на множестве входных данных. Затем можно определить параметры масштабирования  $\alpha_1, \alpha_2, \dots, \alpha_{m_o}$ , спользуя подходящую процедуру перекрестной проверки.

3. Регуляризованная процедура строгой интерполяции, включающая обучение линейных весов.

$$w = (G + \lambda I)^{-1} d$$

Где  $G(x_i, x_j)$  – набор функций Грина, где  $i = 1, 2, \dots, N$ ;  $I$  – единичная матрица размерности  $N \times N$ ;  $w$  – вектор весов  $w$ , соответствующего вектору желаемого отклика  $d$  и данному значению параметра регуляризации  $\lambda$ .

4. Выбор параметра регуляризации  $\lambda$  и множителей  $\alpha_1, \alpha_2, \dots, \alpha_{m_0}$  с помощью некоторого асимптотически-оптимального метода (например, метода перекрестной проверки). Выбранные параметры можно интерпретировать следующим образом.

- Чем больше значение параметра  $\lambda$ , тем сильнее шум влияет на входные параметры.

- Если радиальная базисная функция  $G$  имеет унимодальное (т.е. гауссово) ядро, то чем меньше значения  $\alpha_i$ , тем более «чувствителен» выход сети к соответствующей размерности входа. И наоборот, чем больше значения  $\alpha_i$ , тем менее «адекватно» соответствующая входная размерность отражает влияние изменения входа на общий выход сети. Таким образом, выбранные значения  $\alpha_i$  можно использовать для ранжирования важности соответствующих входных переменных и, следовательно, для определения потенциальных кандидатов на удаление при снижении размерности в случае необходимости.

В контексте рассматриваемой задачи такой выбор можно мотивировать следующим образом. Несложно показать, что сети NRWE составляют специальный класс регуляризованных сетей RBF в том смысле, что любую сеть NWRE можно аппроксимировать правильно построенной последовательностью регуляризованных сетей RBF, для которой последовательность параметров регуляризации  $\{\lambda_N\}$  может возрасти до бесконечности с увеличением размера  $N$  обучающего множества (при этом среднеквадратическая или абсолютная ошибка аппроксимации будет стремиться к нулю). С другой стороны, при стремлении  $N$  к бесконечности риск, при определенных условиях будет стремиться к (глобальной) среднеквадратической ошибке. Если для построения последовательности параметров регуляризации используется процедура выбора асимптотически оптимальных параметров, то (по определению) результирующая последовательность RBF-сетей должна иметь (асимптотически) минимальную среднеквадратическую ошибку на всем множестве возможных последовательностей параметров (включая ту, которая соответствует сети NWRE). Если известно, что сеть NWRE является состоятельной в смысле среднеквадратической ошибки, то этому же условию удовлетворяет регуляризованная сеть RBF, построенная согласно такой же процедуре. Другими словами, регуляризованная сеть RBF, построенная с помощью этой процедуры, может унаследовать свойства состоятельности сети



NWRE. Такая взаимозависимость позволяет распространить известные результаты, касающиеся состоятельности сетей NWRE, на такие области, как регрессия временных рядов, в которых часто встречаются зависимые и нестационарные процессы и где допущения обычных нейронных сетей о независимости данных обучения и стационарности процессов не выполняются. Подводя итог, можно сказать, что описанная процедура, объединяющая элементы теории регуляризации и теории оценки регрессии ядра, обеспечивает практическую методику создания и применения теоретически обоснованных регуляризованных сетей RBF.

### ***Применение метода обратного распространения ошибки для радиально-базисных сетей***

В рамках следующего подхода центры радиальных базисных функций и все остальные свободные параметры сети настраиваются в процессе обучения с учителем. Другими словами, сеть RBF принимает самый общий вид.

Естественным выбором для такой ситуации является обучение, которое составляют градиентные алгоритмы обучения с учителем, где используется алгоритм обратного распространения ошибки. Их основу составляет целевая функция, которая для одного обучающего примера имеет вид:

$$E = \frac{1}{2} \left[ \sum_{i=1}^K w_i \varphi_i(x) - d \right]^2 \quad (2)$$

Предположим, что применяется гауссовская радиальная функция вида:

$$\varphi_i(x(t)) = \exp\left(-\frac{1}{2}u_i(t)\right) \quad (3)$$

где  $i$  – индекс нейрона скрытого слоя,  $j$  – индекс компонента входного вектора,  $t$  – индекс обучающего примера в выборке.

Обучение сети с использованием алгоритма обратного распространения ошибки проводится в два этапа. На первом этапе предъявляется обучающий пример и рассчитываются значения сигналов выходных нейронов сети и значение целевой функции, заданной выражением (2). На втором этапе минимизируется значение этой функции. Подбор значений параметров можно осуществлять, используя градиентные методы оптимизации независимо от объекта обучения – будь то вес или центр. Независимо от выбираемого метода градиентной оптимизации, необходимо, прежде всего, получить вектор градиента целевой функции относительно всех параметров сети. В результате дифференцирования этой функции получим:

$$\frac{\partial E(t)}{\partial w_0(t)} = y(t) - d(t) \quad (4)$$

$$\frac{\partial E(t)}{\partial w_i(t)} = \exp\left(-\frac{1}{2}u_i(t)\right)(y(t) - d(t)) \quad (5)$$

$$\frac{\partial E(t)}{\partial c_{ij}(t)} = (y(t) - d(t))w_i(t)\exp\left(-\frac{1}{2}u_i(t)\right)\frac{(x_j(t) - c_{ij}(t))}{\sigma_{ij}^2(t)} \quad (6)$$

$$\frac{\partial E(t)}{\partial \sigma_{ij}(t)} = (y(t) - d(t))w_i(t)\exp\left(-\frac{1}{2}u_i(t)\right)\frac{(x_j(t) - c_{ij}(t))}{\sigma_{ij}^3(t)} \quad (7)$$

Если в выходном слое содержится несколько нейронов, то формулы (4), (5), (6), (7) соответственно примут следующий вид:

$$\frac{\partial E(t)}{\partial w_{0s}(t)} = y_s(t) - d_s(t) \quad (4a)$$

$$\frac{\partial E(t)}{\partial w_{is}(t)} = \exp\left(-\frac{1}{2}u_i(t)\right)(y_s(t) - d_s(t)) \quad (5a)$$

$$\frac{\partial E(t)}{\partial c_{ij}(t)} = (y(t) - d(t))w_{is}(t)\exp\left(-\frac{1}{2}u_i(t)\right)\frac{(x_j(t) - c_{ij}(t))}{\sigma_{ij}^2(t)} \quad (6a)$$

$$\frac{\partial E(t)}{\partial \sigma_{ij}(t)} = (y(t) - d(t))w_{is}(t)\exp\left(-\frac{1}{2}u_i(t)\right)\frac{(x_j(t) - c_{ij}(t))}{\sigma_{ij}^3(t)} \quad (7a)$$

где  $s$  – номер нейрона выходного слоя.

При использовании метода наискорейшего спуска формулы для корректировки параметров радиально-базисной сети примут следующий вид:

$$w_i(t+1) = w_i(t) - \eta \frac{\partial E(t)}{\partial w_i(t)}, \quad (8)$$

$$c_{ij}(t+1) = c_{ij}(t) - \eta \frac{\partial E(t)}{\partial c_{ij}(t)}, \quad (9)$$

$$\sigma_{ij}(t+1) = \sigma_{ij}(t) - \eta \frac{\partial E(t)}{\partial \sigma_{ij}(t)}. \quad (10)$$

Если в выходном слое содержится несколько нейронов, то формула (8) примет следующий вид:

$$w_{is}(t+1) = w_{is}(t) - \eta_s \frac{\partial E(t)}{\partial w_{is}(t)} \quad (8a)$$

### ***Компьютерное моделирование: классификация образов***

Рассмотрим компьютерный эксперимент, иллюстрирующий методику построения регуляризированной сети RBF на основе использования строгой интерполяции. В этом эксперименте решается задача бинарной классификации данных, полученных на основе двух равновероятных перекрывающихся двумерных распределений Гаусса, соответствующих классам  $C_1$  и  $C_2$ .

Класс  $C_1$  характеризуется вектором среднего значения  $[0,0]^T$  и дисперсией, равной единице; класс  $C_2$  характеризуется вектором среднего значения  $[0,2]^T$  и дисперсией, равной 4. Для двух классов,  $C_1$  и  $C_2$ , строится регуляризированная сеть RBF с двумя выходными функциями – по одной для каждого класса. Кроме того, в качестве желаемых используются следующие бинарные значения:

$$d_k^{(p)} = \begin{cases} 1, & \text{если пример } p \text{ принадлежит классу } C_k, \\ 0 & \text{– в противном случае,} \end{cases}$$

где  $k = 1, 2$ .

Выходы регуляризированной сети RBF обеспечивают оценку апостериорной вероятности принадлежности классу. Однако это истинно только при условии, что сеть обучается с помощью двоичных векторных индикаторов желаемого отклика класса. Тогда для этого класса сетей можно применить следующее решающее правило:

*Выбирается класс, соответствующий максимальному выходному значению функции.*

Метод строгой интерполяции для выбора центров проверялся для различных значений параметра регуляризации  $\lambda$ . При заданном  $\lambda$  для вычисления вектора весов выходного слоя сети RBF использовалось выражение:

$$w = (G + \lambda I)^{-1}d$$

Для каждого из параметров регуляризации строились 50 независимых сетей, которые тестировались на одном и том же множестве из 1000 примеров.

В табл. 1 приведена обобщенная статистическая информация о вероятности корректной классификации  $P_c$ , вычисленной для случая  $m_1 = 20$  центров.

*Таблица 1. Вероятность корректной классификации  $P_c(\%)$  для разных значений параметра регуляризации  $\lambda$  и размера скрытого слоя  $m_1 = 20$*

<i>Вид статистики</i>	<i>Значение параметра регуляризации <math>\lambda</math></i>					
	0	0,1	1	10	100	1000
Среднее	57,49	72,42	74,42	73,80	72,46	72,14
Стандартное отклонение	7,47	4,11	3,51	4,17	4,98	5,09
Минимум	44,20	61,60	65,80	63,10	60,90	60,50
Максимум	72,70	78,30	78,90	79,20	79,40	79,40

Статистика по массиву вычислялась для различных значений параметра регуляризации  $\lambda$ . В табл. 2 представлены соответствующие результаты, вычисленные для случая регуляризированной сети RBF большего размера с  $m_1 = 100$  центрами.

Таблица 2. Вероятность корректной классификации  $P_c(\%)$  для разных значений параметра регуляризации  $\lambda$  и размера скрытого слоя  $m_1 = 100$

Вид статистики	Значение параметра регуляризации $\lambda$					
	0	0,1	1	10	100	1000
Среднее	50,58	77,03	77,72	77,87	76,47	75,33
Стандартное отклонение	4,70	1,45	0,94	0,91	1,62	2,25
Минимум	41,00	70,60	75,10	75,10	72,10	70,10
Максимум	61,30	79,20	79,80	79,40	78,70	78,20

На рис. 3. показаны границы решений, сформированные выходами сети для значения параметра регуляризации  $\lambda = 10$  (при таком значении параметра статистические характеристики оказались наилучшими).

На рис. 3, а показана сеть с наилучшими показателями, а на рис. 3, б – с наихудшими статистическими показателями в рамках массива. Обе части этого рисунка представлены для случая сети с  $m_1 = 100$  центрами.

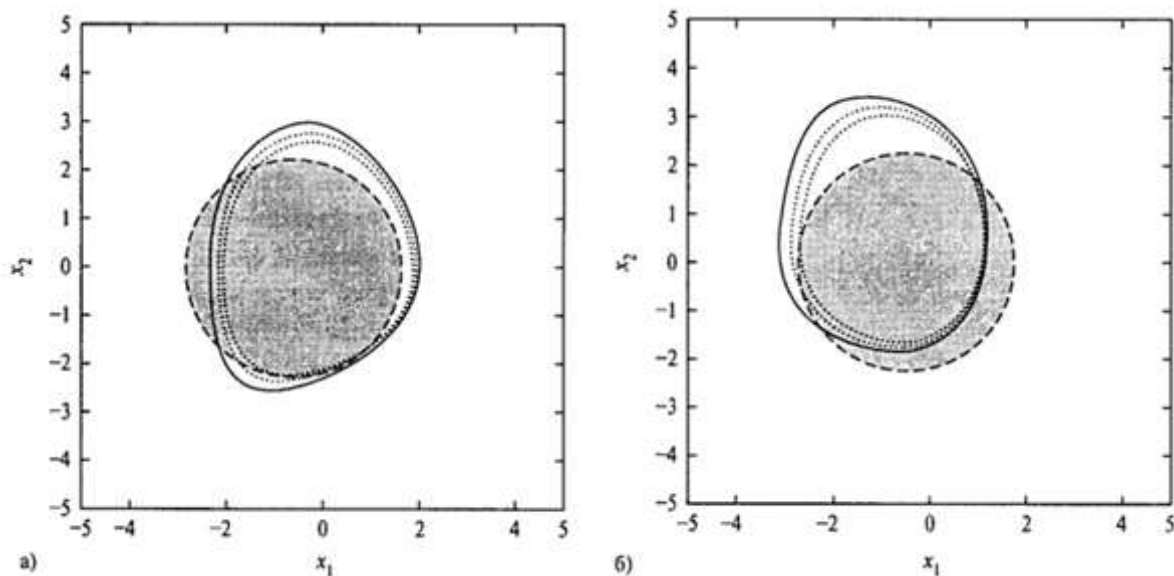


Рис. 5.7. Результаты компьютерного моделирования при решении задачи классификации на основе строгой интерполяции в регуляризируемых сетях RBF: наилучшее решение (а); наихудшее решение (б). Закрашенная область представляет собой оптимальное решение Байеса

Сравнивая табл. 1. и 2, можно сделать следующие выводы.

1. Как в случае  $m_1 = 20$  центров, так и в случае  $m_1 = 100$  центров качество классификации в сети для параметра регуляризации  $\lambda = 0$  остается относительно плохим.

2. Использование регуляризации оказывает большое влияние на

качество классификации в сетях RBF.

3. Для  $\lambda \geq 0,1$  качество классификации зависит от значения  $\lambda$ . Для случая  $m_1 = 20$  центров лучшие показатели были достигнуты при  $\lambda = 1$ , а для случая  $m_1 = 100$  центров – при  $\lambda = 10$ .

4. Увеличение количества центров с 20 до 100 улучшает качество классификации приблизительно на 4,5%.

Структура сетей RBF является необычной в том смысле, что архитектура скрытых элементов в корне отличается от структуры выходных. Поскольку основой функционирования нейронов скрытого слоя являются радиальные базисные функции, теория сетей RBF тесно связана с теорией радиальных базисных функций, которая в настоящее время является одной из основных областей изучения в численном анализе. Интересным также является тот факт, что настройка линейных весов выходного слоя позволяет обеспечить хорошее качество классификации.