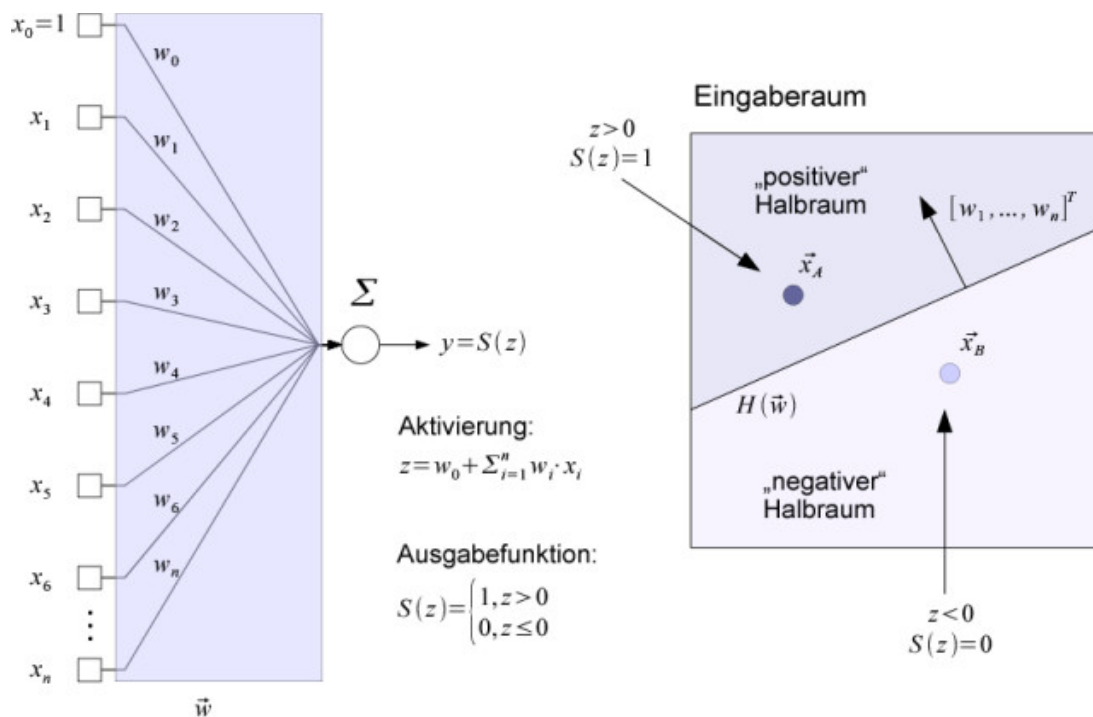


## 7. RBF-Netze

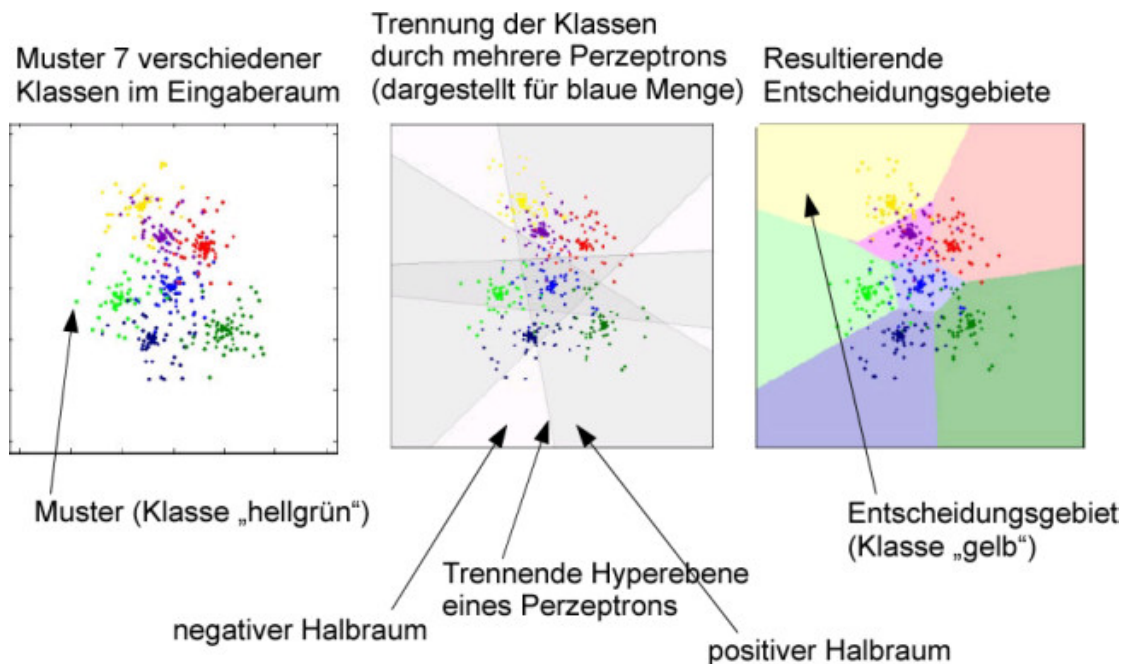
### 7.1 Einleitung: Klassifizierung mit Perzeptrons

Zur Lösung von Klassifikationsproblemen haben wir uns zu Beginn das sogenannte Perzeptron angesehen, mit dem es möglich war, zu einer Reihe von Mustern eine trennende Hyperebene zu lernen, die zwei zu unterscheidende Klassen voneinander separiert. Der Eingaberaum wurde durch diese Hyperebene so in zwei Teile geteilt, da das Perzeptron in einem Teilraum eine positive und im restlichen Teilraum eine negative Aktivierung  $z$  besitzt (vergleiche dazu die nachfolgende Abbildung 7.1), was es uns ermöglichte, anhand der Ausgabe des Perzeptrons eine Aussage darüber zu treffen, in welchem Teilraum sich ein bestimmtes Eingabemuster befindet.



**Abbildung 7.1** Perzeptron (links) und Aufteilung des Eingaberaums in zwei Bereiche unterschiedlicher Neuronenaktivität (rechts)

Verwenden wir nun mehrere Perzeptrons, so lassen sich auf diese Weise komplexere Bereiche aus einem Eingaberaum "herausschneiden". In Abbildung 7.2, links, ist ein Eingaberaum mit Mustern aus 7 verschiedenen Klassen dargestellt. Abbildung 7.2, mitte, enthlt die Hyperebenen einiger Perzeptrons, wobei dunkle Bereiche Halbrume positiver Aktivierung und helle Bereiche Halbrume negativer Aktivierung angeben sollen. Der Bereich in der Mitte, in dem sich die blaue Menge befindet, liegt ausschlielich auf der positiven Seite jeder Hyperebene, so da die Perzeptrons, deren Gewichtsvektoren diese Hyperebenen definieren, hier ausschlielich eine Ausgabe von +1 besitzen. Damit lsst sich die blaue Menge (bis auf einige wenige Ausnahmen) von den brigen Mengen unterscheiden.



**Abbildung 7.2** Klassifizierung von Mengen durch mehrere Perzeptrons

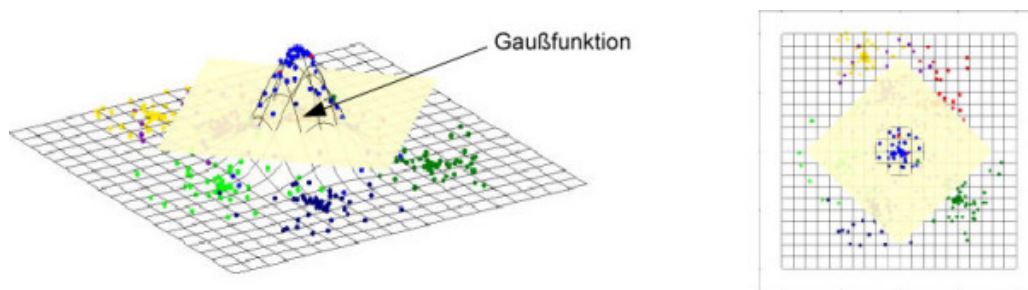
In Abbildung 7.2, rechts, wurden weitere Perzeptrons verwendet und deren Ausgaben geschickt kombiniert, um so eine Klasseneinteilung fr jede der sieben Mengen vorzunehmen. Die Grenzen der Entscheidungsgebiete, die sich so ergeben, sind dabei ausschlielich durch Hyperebenen gegeben, was im 2-dimensionalen Fall sehr schn daran zu erkennen ist, da die einzelnen Entscheidungsgebiete linear voneinander abgegrenzt sind.

Fr einfach strukturierte (beispielsweise linear separierbare) Mengen funktioniert dies recht gut, allerdings haben die Entscheidungsgebiete in der Regel weitaus komplexere, gekrmmte, Formen. Die Abbildung 7.2 lsst schon erahnen, da eine lineare Trennung der Klassen nicht optimal ist, da sich mehrere Muster einer Klasse in unterschiedlichen Entscheidungsgebieten befinden und damit falsch klassifiziert werden.

## 7.2 Klassifizierung mit Radialen Basisfunktionen (RBFs)

Aus diesem Grund wäre ein Verfahren nützlich, das eine Klassifikation mit gekrümmten Entscheidungsgebieten vornimmt. Eine Möglichkeit dies zu tun, besteht in der Verwendung sogenannter *Radialer Basisfunktionen*.

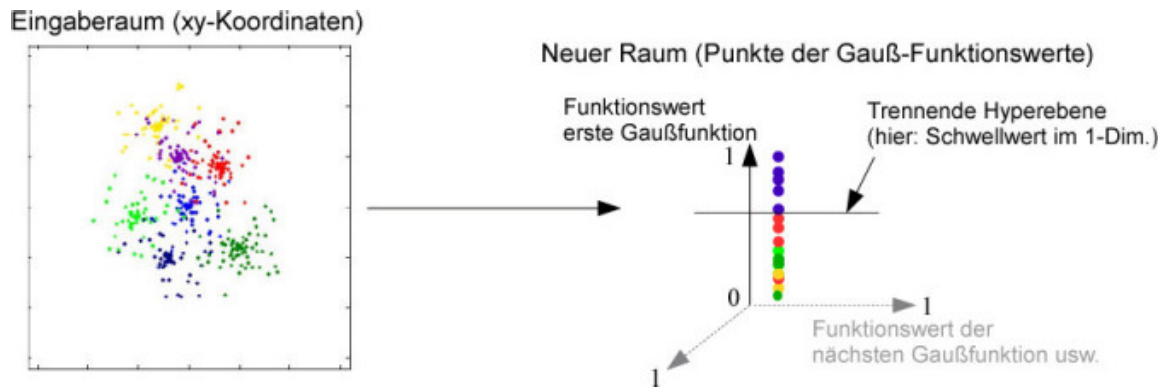
Die grundsätzliche Idee besteht darin, verschiedene Gaußfunktionen in den Eingaberaum zu legen und die einzelnen Muster damit so in einen neuen Raum abzubilden, da die einzelnen Klassen dort mit einer höheren Wahrscheinlichkeit linear separierbar sind. Siehe dazu Abbildung 7.3.



**Abbildung 7.3** Gaußfunktion im *Eingaberaum* (links), die jedem Muster einen Funktionswert zuordnet. Die blaue Menge lässt sich dann mit einem Schwellwert von den übrigen Mengen abtrennen. Im Eingaberaum ergibt sich dadurch ein gekrümmtes (kreisförmiges) Entscheidungsgebiet.

Dargestellt ist eine einzelne Gaußfunktion, die sich direkt über der blauen Menge befindet und jedem Muster im Eingaberaum einen Funktionswert zuordnet. Wie man sieht, werden die blauen Muster dadurch angehoben und können von den übrigen Mustern durch einen Schwellwert (symbolisiert durch die gelb dargestellte Ebene) abgetrennt werden.

Der Eingaberaum wird durch eine einzelne Gaußfunktion daher in einen Raum der Dimension 1 abgebildet, dargestellt in Abbildung 7.4. Es ist zu erkennen, da sich in dem neuen Raum leicht eine trennende Hyperebene angeben lässt (im 1-dimensionalen Fall entspricht die Hyperebene einfach einem Schwellwert).

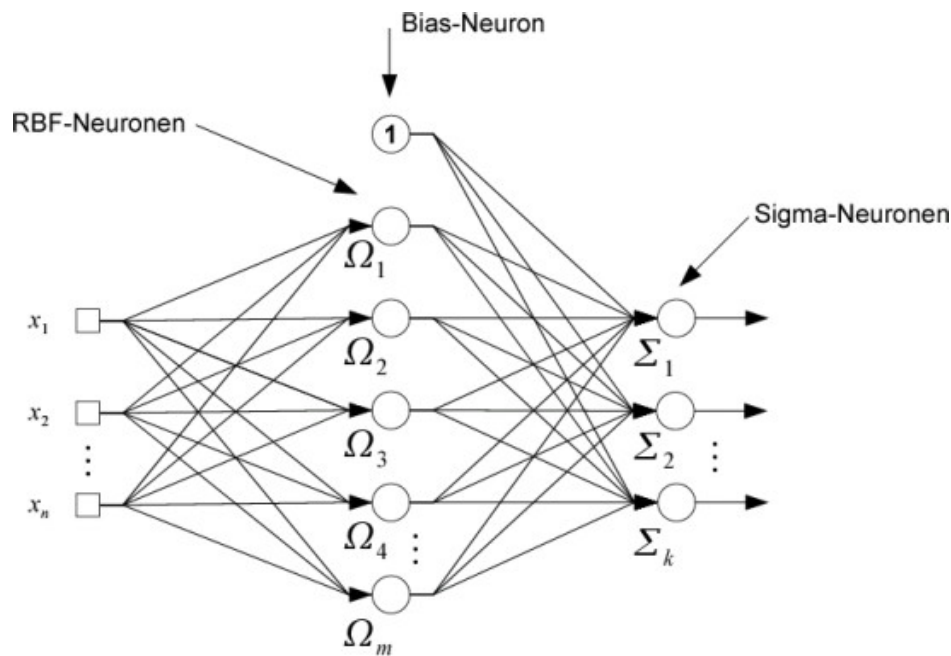


**Abbildung 7.4** Abbildung eines Eingaberaumes in einen neuen Zielraum, in dem sich die einzelnen Klassen durch eine Hyperebene leichter voneinander trennen lassen, als im ursprünglichen Eingaberaum.

Um komplexere Entscheidungsgebiete zu bilden, können wir nun mehrere Gaußfunktionen in den Eingaberaum legen. Wir erhalten dann entsprechend einen neuen Raum, dessen Dimensionalität genau der Anzahl verwendeter Gaußfunktionen entspricht, da jedem Muster im Eingaberaum genau so viele verschiedene Funktionswerte zugeordnet werden, wie wir Gaußfunktionen verwendet haben (angedeutet in Abbildung 7.4, rechts).

### 7.3 Klassifikation mit RBF-Netzen

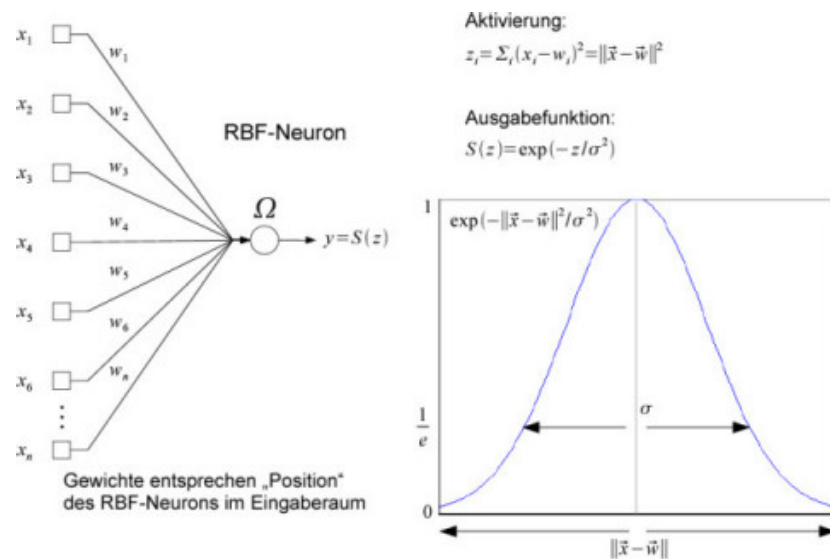
Eine Möglichkeit, eine derartige Klassifikation mit Gaußfunktionen vorzunehmen, besteht in der Verwendung eines sogenannten *RBF-Netzes*. Wie der Name andeutet, verwendet es Radiale Basisfunktionen (im konkreten Fall Gaußfunktionen), um wie oben angegeben eine Klassifikation von Mustern selbstständig zu lernen. Der Netzaufbau ist in Abbildung 7.5 dargestellt.



**Abbildung 7.5** Struktur eines RBF-Netzes

Im Unterschied zu den üblichen Netzen, die wir bisher betrachtet haben, besteht ein RBF-Netz aus zwei unterschiedlichen Typen von Neuronen, den *Omega*- und den *Sigma*-Neuronen. Die Sigma-Neuronen sind die üblichen Neuronen, die wir bisher betrachtet haben. Sie summieren einfach ihre Eingabe gewichtet auf und produzieren das so erhaltene Ergebnis als Ausgabe (daher die Bezeichnung mit dem Summenzeichen "Sigma").

Die Omega-Neuronen, die wir im folgenden auch als *RBF-Neuronen* bezeichnen werden, funktionieren etwas anders. Siehe dazu Abbildung 7.6. Ihre Aktivierung besteht aus einer Summe quadrierter Differenzen zwischen den einzelnen Eingabewerten und den zugehörigen Gewichten. Dies entspricht gerade dem quadrierten euklidischen Abstand zwischen einem Eingabevektor  $x$  und dem Gewichtsvektor  $w$  eines RBF-Neurons. Anschließend wird eine Gaußfunktion als Ausgabefunktion auf die Aktivierung des Neurons angewendet.

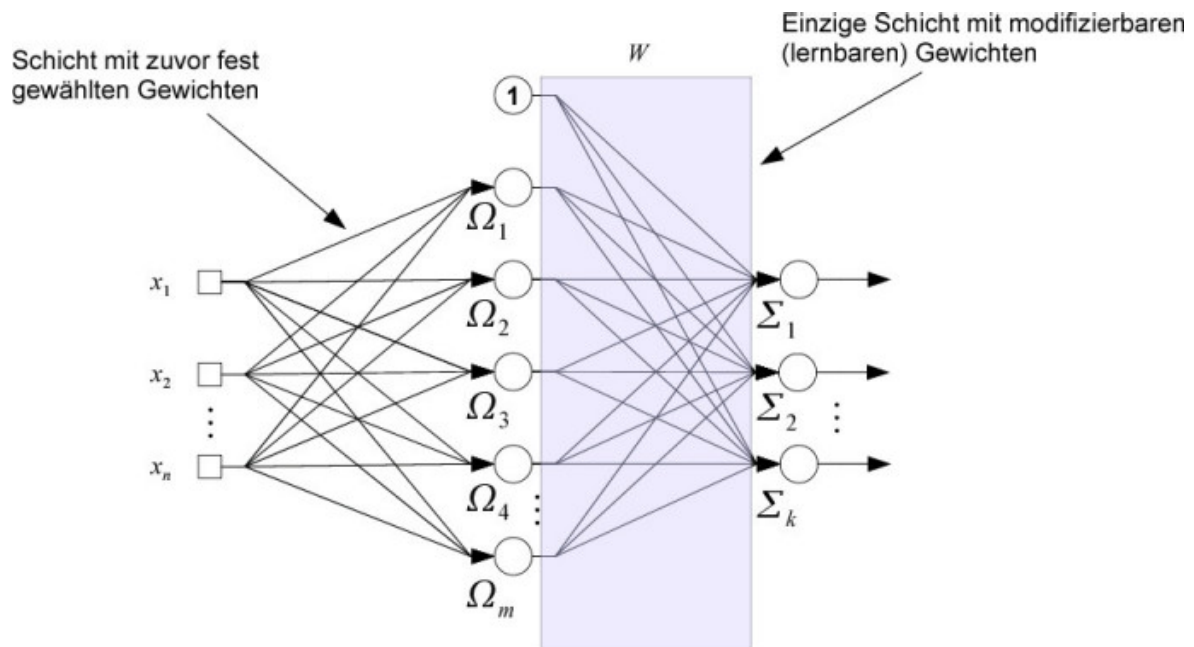


**Abbildung 7.6** Aufbau eines RBF-Neurons

Bei näherer Betrachtung stellt sich heraus, daß das so definierte RBF-Neuron eine einzelne Gaußfunktion im Eingaberaum definiert, deren Position im Eingaberaum gerade durch den Gewichtsvektor des betreffenden Neurons gegeben ist.

## 7.4 Training von RBF-Netzen

Die Gewichte der ersten Schicht des RBF-Netzes, die zu den RBF-Neuronen gehören, werden üblicherweise zu Anfang fest gewählt, bevor mit dem Training des Netzes begonnen wird und werden im Verlauf des Trainings auch nicht mehr verändert. Einzig die Gewichte der Sigma-Neuronen werden angepaßt (siehe Abbildung 7.7).



**Abbildung 7.7** Struktur eines RBF-Netzes. Die Gewichte der ersten Schicht werden zu Anfang fest gewählt, während die Gewichte der zweiten Schicht gelernt werden können.

Das Training eines solchen Netzes ist dabei sehr einfach, da wir nur eine einzelne lineare Schicht trainieren müssen, und kann durch einen der Standardalgorithmen erfolgen, die wir bisher kennengelernt haben.



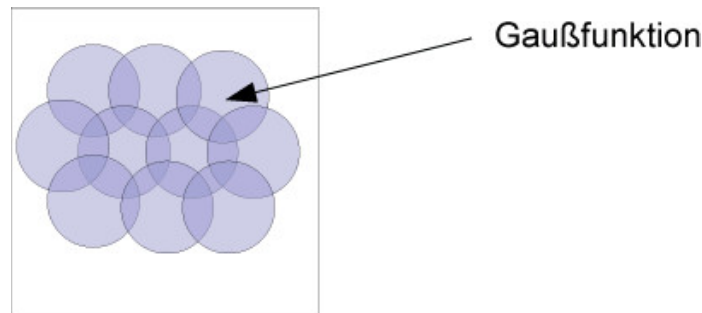
### Training eines RBF-Netzes

1.  $n \leftarrow$  ANZAHL AN RBF-NEURONEN
2.  $m \leftarrow$  ANZAHL AN TRAININGSMUSTERN
  - A. Wahl der RBF-Zentren und -Radien
3.  $\forall i, 1 \leq i \leq n: \underline{c}_i \leftarrow$  Position von Neuron  $i$  im Eingaberaum
4.  $\forall i, 1 \leq i \leq n: \sigma_i \leftarrow$  RBF-Radius für Neuron  $i$ 
  - B. Berechnung des Ausgabevektors für jedes RBF-Neuron bei Eingabe der einzelnen Trainingsmuster  $\underline{x}_j$
5. für  $j=1,2,\dots,m$ :
  6.  $\forall i, 1 \leq i \leq n: y_i \leftarrow \exp(-||\underline{c}_i - \underline{x}_j||^2 / \sigma_i^2)$
  7.  $\underline{y}_j = [y_1, \dots, y_n]^T$
8. end
  - C. Training der Ausgabeschicht des RBF-Netzes
9. Trainiere Gewichtsmatrix  $W^{(2)}$  der Ausgabeschicht unter Verwendung der Trainingsmuster  $(\underline{y}_j, \underline{\ell}_j), j = 1, 2, \dots, m$  (z.B. mit Hilfe von Backpropagation oder der Delta-Regel)

Das einzige Problem an dieser Stelle besteht darin, wie die Gewichte der RBF-Neuronen, also die Position der Gaußfunktionen im Eingaberaum (die sogenannten *Zentren* der RBF-Neuronen) und die Ausdehnung der Gaußfunktionen (die sogenannten *RBF-Radien* - siehe Abbildung 6) gewählt werden sollen.

Was die RBF-Radien betrifft, so wählt man diese meist so, daß sich die Gaußfunktionen benachbarter RBF-Neuronen leicht überlappen, siehe Abbildung 8. Dadurch wird es dem Netz in der Regel erleichtert, eine gute Approximation der Klasseneinteilung im Eingaberaum zu lernen. Meist verwendet man (da es einfacher ist) dabei eine feste Breite für alle Gaußfunktionen, so daß zumindest keine Neuronen "für sich" stehen.



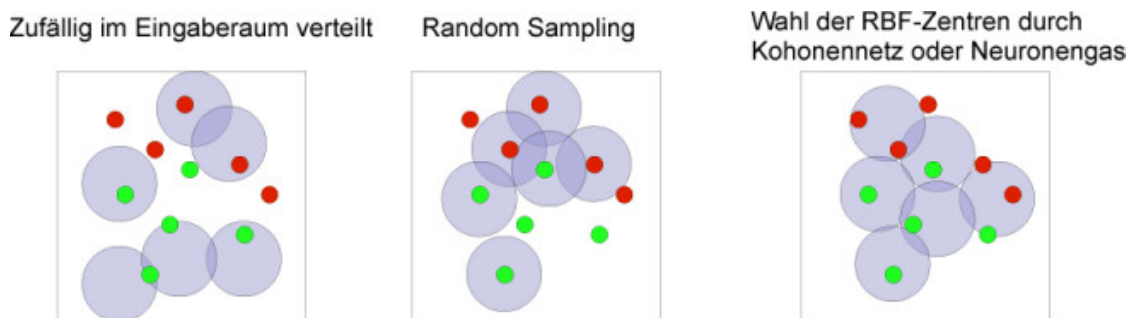


**Abbildung 7.8**  
 Übliche Wahl der RBF-Radius

Die genaue Wahl der RBF-Zentren ist schon etwas kniffliger, da sie, zusammen mit den RBF-Radius, stark die Qualität des RBF-Netzes beeinflusst. Dabei gibt es mehrere Heuristiken, die sich in der Praxis bewährt haben, siehe Abbildung 7.9. Die einfachste Methode besteht in einer zufälligen Positionierung der RBF-Zentren (Abbildung 7.9, links), was gute Ergebnisse liefern kann, solange man sich auf einen kleinen Eingaberaum mit sehr wenigen Dimensionen (ca. 1 bis 3) beschränkt.

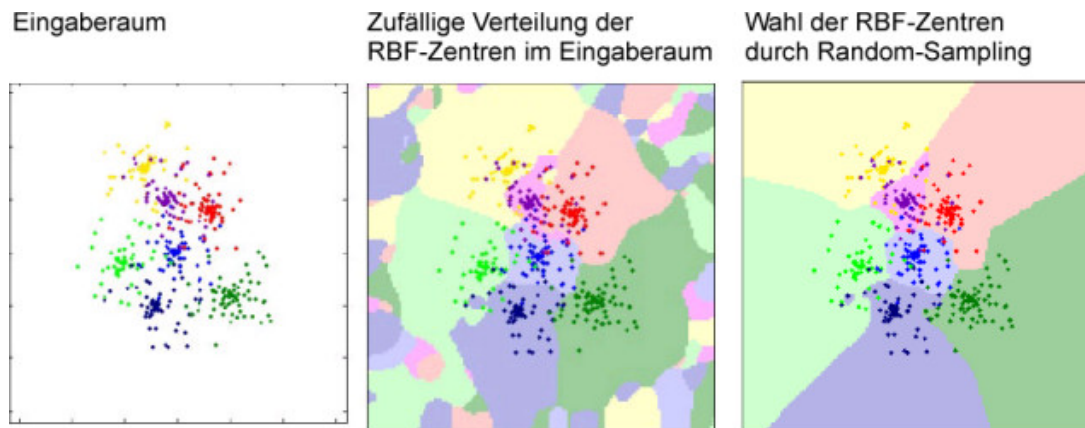
Bessere Ergebnisse in höherdimensionalen Eingabebereichen liefert das sogenannte *Random Sampling*. Dabei werden die einzelnen RBF-Neuronen auf die Position zufällig gewählter Muster gesetzt. Diese Methode hat den Vorteil, dass keine RBF-Neuronen für leere Bereiche des Eingaberaumes "verschwendet" werden.

Man kann die Position der RBF-Neuronen auch mit Kohonenkarten oder Neuronengas lernen. Diese Methode liefert in der Regel die besten Ergebnisse, ist allerdings auch etwas aufwendiger und wird erst in der nächsten Woche behandelt.



**Abbildung 7.9** Verschiedene Möglichkeiten, die Zentren von RBF-Neuronen zu wählen. Meist verwendet man das Random-Sampling, da es sowohl sehr einfach ist, als auch zu guten Ergebnissen führt. Die beste Positionierung lässt sich mit Kohonenkarten oder Neuronengas erreichen.

Abschließend ist in Abbildung 7.10 einmal das Ergebnis dargestellt, da man mit einem RBF-Netz für die 7 Klassen aus Abbildung 7.2 erzielen kann.



**Abbildung 10** Klassifikationsergebnis der Menge aus Abbildung 7.2 mit einem RBF-Netz. Zu erkennen sind die geklammerten Entscheidungsgebiete, die wesentlich besser an die Verteilung der einzelnen Klassen angepaßt sind.

Die getrennte Optimierung der Parameter der ersten und zweiten Schicht hat verschiedene Vorteile. So lassen sich zum einen verschiedene, voneinander unabhängige Methoden zur Optimierung der ersten und der zweiten Schicht einsetzen, zum anderen ist die Konvergenz dieser Verfahren für die Anpassung einer einzelnen Schicht deutlich besser als bei der gleichzeitigen Anpassung beider Schichten. Dies hängt damit zusammen, dass der Suchraum bei beiden Schichten (die Anzahl aller möglichen Parameterwertkombinationen) exponentiell größer ist als bei einer Beschränkung auf die Parameter nur einer Schicht; die Dimensionszahl beider Suchräume addieren sich.