第 30 章 径向基函数网络

30.1 RBF 网络假设

定义 30.1.1 (径向基函数（Radial Basis Function, RBF）)

径向基函数（Radial Basis Function, RBF）是一类实值函数，其函数值仅依赖于输入向量与某一固定中心点之间的欧氏距离，而与方向无关。数学上，存在一元函数 $\varphi : [0,\infty) \to \mathbb{R}$，使得

$$\Phi(x) = \varphi\big(\|x - c\|\big),$$

其中 c 为中心向量，$\|\cdot\|$ 为欧几里得范数。

常见实例

- 高斯径向基函数

$$\varphi(r) = \exp\Big(-\frac{r^2}{2\sigma^2}\Big), \quad \sigma > 0.$$

- 多二次径向基函数

$$\varphi(r) = \sqrt{r^2 + c^2}, \quad c > 0.$$

- 逆多二次径向基函数

$$\varphi(r) = \frac{1}{\sqrt{r^2 + c^2}}, \quad c > 0.$$

- 薄板样条径向基函数

$$\varphi(r) = r^2 \log r.$$

主要特性

- 径向对称，便于高维推广；
- 局部作用：远离中心时函数值迅速衰减；
- 根据 Stone-Weierstrass 定理，RBF 组合可在紧集上任意精度逼近任意连续函数。

应用领域　函数插值、曲面重建、支持向量机核函数、RBF 神经网络隐层激活函数等。

命题 30.1.1 (从神经网络到 RBF 网络)

RBF（径向基函数）网络可视为神经网络的一种特殊结构，二者仅在隐藏层计算方式不同。

1. 隐藏层对比

- 神经网络：内积 + 非线性

$$h_j(x) = \tanh(w_j^\top x + b_j).$$

- **RBF** 网络：距离 + 径向基

$$RBF(x, \mu_m) = h_m(x) = \exp\Big(-\gamma\|x - \mu_m\|^2\Big).$$

2. 输出层统一　均为线性聚合

$$g(x) = \sum_{m=1}^{M} \beta_m \, h_m(x) + b.$$

3.RBF 网络假设

给定中心 μ_m 与权重 β_m，RBF 网络定义为

$$g(x) = \sum_{m=1}^{M} \beta_m \exp\Big(-\gamma\|x-\mu_m\|^2\Big) + b.$$

其中

- μ_m：中心（可随机选、K-means 或 SVM 支持向量）；
- β_m：带符号的投票权重（可最小二乘或 SVM 对偶系数 $\alpha_m y_m$）。

4. 与 **SVM** 的联系

高斯核 SVM 可视为

$$M = \#\mathrm{SV}, \quad \mu_m = \text{支持向量}, \quad \beta_m = \alpha_m y_m.$$

结论　RBF 网络以距离度量替代内积，兼具神经网络与核方法的优点，历史上被视为神经网络的重要分支。

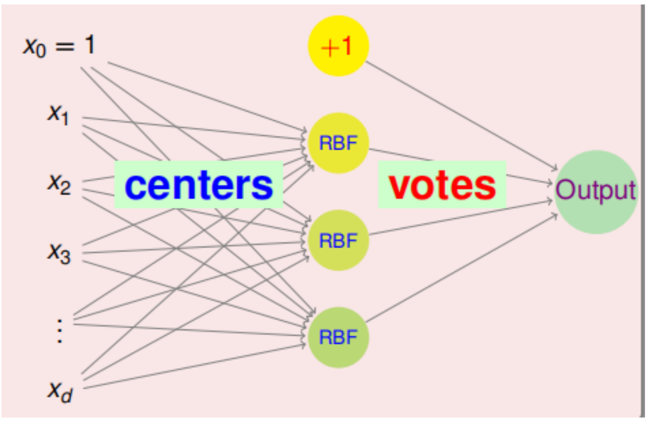


图 **30.1.1:** RBF 网络

命题 30.1.2 (RBF 与相似性)

RBF 网络通过“空间距离”度量输入 x 与中心 μ_m 的相似度，从而构造特征变换。

常见相似度函数

- 神经网络神经元

$$\mathrm{Neuron}(x,x') = \tanh(\gamma x^\top x' + 1).$$

- 字符串相似度

$$\mathrm{DNASim}(x,x') = \mathrm{EditDistance}(x,x').$$

- 核函数（Mercer 条件）

$$\mathrm{Poly}(x,x') = (1+x^\top x')^2, \qquad \mathrm{Gaussian}(x,x') = \exp(-\gamma\|x-x'\|^2).$$

- 截断相似度

$$\mathrm{Truncated}(x,x') = \max\big(0, 1-\|x-x'\|\big)^2.$$

RBF 视角

RBF 网络将"距离到中心"的单调不增函数作为相似度，进而

$$\Phi(x) = \left[\text{Gaussian}(x, \mu_1), \ldots, \text{Gaussian}(x, \mu_M)\right]^\top$$

完成特征变换，随后线性组合即可。

结论　RBF 网络把"距离 $\to$ 相似度 $\to$ 特征"作为统一框架，兼容核方法与神经网络思想。

例题 30.1 选择题：径向基函数的判定

以下哪一项不是径向基函数（radial basis function）？

1) $\phi(\mathbf{x}, \boldsymbol{\mu}) = \exp(-\gamma\|\mathbf{x} - \boldsymbol{\mu}\|^2)$
2) $\phi(\mathbf{x}, \boldsymbol{\mu}) = -\sqrt{\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\boldsymbol{\mu} + \boldsymbol{\mu}^T\boldsymbol{\mu}}$
3) $\phi(\mathbf{x}, \boldsymbol{\mu}) = [\![\mathbf{x} = \boldsymbol{\mu}]\!]$
4) $\phi(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\mu}^T\boldsymbol{\mu}$

解答　正确选项为 $\boxed{4}$。径向基函数的核心是函数值仅依赖于输入 $\mathbf{x}$ 与中心 $\boldsymbol{\mu}$ 的距离（即 $\phi(\mathbf{x}, \boldsymbol{\mu}) = \phi(\|\mathbf{x} - \boldsymbol{\mu}\|)$）。

- 选项 1 是高斯径向基函数，依赖 $\|\mathbf{x} - \boldsymbol{\mu}\|^2$；
- 选项 2 化简后依赖 $\|\mathbf{x} - \boldsymbol{\mu}\|$；
- 选项 3 是指示函数，依赖 $\|\mathbf{x} - \boldsymbol{\mu}\| = 0$ 的条件；
- 选项 4 $\mathbf{x}^T\mathbf{x} + \boldsymbol{\mu}^T\boldsymbol{\mu}$ 与 $\mathbf{x}$ 和 $\boldsymbol{\mu}$ 的距离无关，不满足径向基函数的定义。

因此，选项 4 不是径向基函数。 ■

30.2 RBF 网络学习

命题 30.2.1 (Full RBF 网络与最近邻)

当 RBF 网络将全部训练样本设为基函数中心时，可导出两种经典模型。

1. **Full RBF** 网络

   令中心数 $M = N$ 且 $\mu_m = x_m$，则

   $$g_{\text{uniform}}(x) = \text{sign}\Big(\sum_{m=1}^{N} y_m \exp\big(-\gamma\|x - x_m\|^2\big)\Big).$$

   物理意义：每个样本 x_m 依据相似度 $\exp(-\gamma\|x - x_m\|^2)$ 对 x 投出带符号的"一票"。

2. 最近邻（**1-NN**）

   若仅保留相似度最大的一项，则投票退化为选择：

   $$\text{nbor}(x) = y_m, \quad \text{其中} \, m = \arg\min_i \|x - x_i\|.$$

   进一步推广到 k 个最近邻可得到 k-NN 模型：

   $$g_{k\text{-NN}}(x) = \text{sign}\Big(\sum_{x_i \in \mathcal{N}_k(x)} y_i\Big), \quad \mathcal{N}_k(x) \text{ 为} x \text{ 的} k \text{ 个最近邻。}$$

结论　Full RBF 网络由"全体投票"演变为"最近邻选择"，揭示了 RBF、核方法与实例学习之间的内在联系。

命题 30.2.2 (全连接 RBF 网络的插值性质)

考虑平方误差意义下的回归任务，采用全连接 RBF 网络

$$h(x)=\sum_{m=1}^{N}\beta_m\,\underbrace{\mathrm{RBF}(x,x_m)}_{\text{Gaussian}},\quad \text{其中中心 } \mu_m=x_m,\ M=N.$$

线性回归视角

将输入 x 通过径向基变换映射为

$$z_n=\left[\mathrm{RBF}(x_n,x_1),\ \mathrm{RBF}(x_n,x_2),\ \dots,\ \mathrm{RBF}(x_n,x_N)\right]^\top\in\mathbb{R}^N,\quad Z=[z_1,\dots,z_N]^\top\in\mathbb{R}^{N\times N}.$$

最优系数由最小二乘给出

$$\beta=(Z^\top Z)^{-1}Z^\top y,\quad y=[y_1,\dots,y_N]^\top.$$

插值定理

若所有样本 $\{x_n\}_{n=1}^N$ 互不相同，则采用高斯 RBF 的矩阵 Z 正定、可逆，于是

$$\beta=Z^{-1}y,\quad \text{网络恰好插值所有训练点：}\ h(x_n)=y_n,\ \forall n.$$

结论　全连接 RBF 网络在高斯核下天然具备插值能力，其系数可通过一次性求解线性方程组获得。

命题 30.2.3 (正则化全连接 RBF 网络与中心剪枝)

全连接高斯 RBF 网络在插值情形下满足

$$g_{\mathrm{RBF}}(x_n)=y_n,\quad \forall n=1,\dots,N,\quad \Longrightarrow\quad E_{\text{in}}=0,$$

称为精确插值，但易导致过拟合。

1. 正则化系数估计

采用岭回归（ridge regression）代替最小二乘，得

$$\beta=(Z^\top Z+\lambda I)^{-1}Z^\top y,$$

其中

$$Z^\top Z=[\mathrm{Gaussian}(x_n,x_m)]_{n,m=1}^N=K$$

即高斯核矩阵，于是

$$\beta=(K+\lambda I)^{-1}y,$$

等价于核岭回归。

2. 中心剪枝正则化

用 $M<N$ 个中心 $\{\mu_m\}_{m=1}^M$ 替代全部 N 个样本中心：

$$g(x)=\sum_{m=1}^{M}\beta_m\,\mathrm{Gaussian}(x,\mu_m)+b,\quad \mu_m\ \text{为原型/质心}.$$

- 降低参数数量，抑制过拟合；
- 物理意义：μ_m 作为数据的代表性“原型”；
- 与 SVM 类比：支持向量仅为一组有效中心。

结论　通过正则化或中心剪枝，全连接 RBF 网络从“精确插值”过渡到“稳健学习”，兼顾拟合

与泛化。

例题 30.2 选择题：全高斯 RBF 网络 Z 矩阵变化

若 $\mathbf{x}_1 = \mathbf{x}_2$，全高斯径向基函数（RBF）网络的 Z 矩阵会发生什么变化？

1) 矩阵的前两行相同
2) 矩阵的前两列不同
3) 矩阵可逆
4) 前两行和前两列交叉处的子矩阵包含常数 0

解答 正确选项为 $\boxed{1}$。全高斯 RBF 网络中，Z 矩阵元素定义为 $Z_{ij} = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$（以训练样本为中心）。

当 $\mathbf{x}_1 = \mathbf{x}_2$ 时，对任意 j，有：

$$Z_{1j} = \exp(-\gamma\|\mathbf{x}_1 - \mathbf{x}_j\|^2), \quad Z_{2j} = \exp(-\gamma\|\mathbf{x}_2 - \mathbf{x}_j\|^2) = \exp(-\gamma\|\mathbf{x}_1 - \mathbf{x}_j\|^2)$$

因此，Z 矩阵的第一行与第二行元素完全相同。 ■

30.3 k-均值算法

命题 30.3.1 (聚类原型：K-means 交替优化)

给定数据集 $\{\boldsymbol{x}_n\}_{n=1}^N$，通过聚类选取 M 个原型 $\{\boldsymbol{\mu}_m\}_{m=1}^M$，用于构造正则化 RBF 网络。其核心逻辑与步骤如下：

1. 聚类误差定义

定义由 $\{S_1, \ldots, S_M\}$（数据集的不交并划分，满足样本无重叠、全覆盖）和原型 $\{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_M\}$ 共同决定的 平方误差：

$$E_{\text{in}}(S_1, \ldots, S_M; \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_M) = \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M}\mathbb{I}[\boldsymbol{x}_n \in S_m]\|\boldsymbol{x}_n - \boldsymbol{\mu}_m\|^2$$

该误差刻画“样本到所属簇原型的距离总和”，越小代表原型对簇内样本的代表性越强。

2. K-means 交替最小化算法

因直接联合优化划分 S_m 与原型 $\boldsymbol{\mu}_m$ 计算复杂，采用 交替优化 策略，分两步迭代逼近最小误差：

1. 固定原型，优化划分（样本指派）

   若原型 $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_M$ 已确定，对每个样本 $\boldsymbol{x}_n$，将其指派到 距离最近原型 对应的簇：

   $$S_m = \left\{\boldsymbol{x}_n \mid \|\boldsymbol{x}_n - \boldsymbol{\mu}_m\| \le \|\boldsymbol{x}_n - \boldsymbol{\mu}_j\|, \forall j \in \{1, \ldots, M\}\right\}$$

   此操作保证样本“唯一归属”到当前最匹配簇，最小化距离误差。

2. 固定划分，优化原型（均值更新的数学推导）

   若划分 $S_1, \ldots, S_M$ 已确定，需更新每个簇的原型 $\boldsymbol{\mu}_m$。此时对单个原型 $\boldsymbol{\mu}_m$，可通过无约束优化 推导其最优解：

   (1) 梯度计算：

对聚类误差 E_{in} 中与 $\boldsymbol{\mu}_m$ 相关的项求偏导（仅 S_m 内样本贡献非零项）：

$$\nabla_{\boldsymbol{\mu}_m} E_{\text{in}} = -2\sum_{n=1}^{N} [\![\boldsymbol{x}_n \in S_m]\!](\boldsymbol{x}_n - \boldsymbol{\mu}_m)$$

利用指示函数筛选性，简化为：

$$\nabla_{\boldsymbol{\mu}_m} E_{\text{in}} = -2\left(\sum_{\boldsymbol{x}_n \in S_m} \boldsymbol{x}_n \;-\; |S_m|\boldsymbol{\mu}_m\right)$$

(2) 最优原型求解：

令梯度为 **0**（无约束优化极值条件），解得：

$$\boldsymbol{\mu}_m = \frac{1}{|S_m|}\sum_{\boldsymbol{x}_n \in S_m} \boldsymbol{x}_n$$

即最优原型是簇内样本均值，数学上可证明这是当前划分下平方误差最小的唯一解。

结论　K-means 交替优化通过迭代执行“样本指派 → 原型更新”，同时达成：

- 聚类紧凑性：数据被划分为内部距离紧凑的簇；
- 原型实用性：$\boldsymbol{\mu}_m$ 作为簇的“共识中心”，可直接用于正则化 RBF 网络的中心构造，简化模型同时保留数据簇特征。

算法 30.3.1: k-Means 聚类算法

输入: 数据集 $\{x_n\}_{n=1}^N$；聚类数 k；最大迭代次数 $T_{\max}$

输出: 簇划分 $\{S_1, \ldots, S_k\}$；聚类中心 $\{\mu_1, \ldots, \mu_k\}$

初始化

随机选取 k 个样本作为初始中心 $\mu_1^{(0)}, \ldots, \mu_k^{(0)}$.

repeat

分配步骤

对每个样本 x_n 执行

$$S_j \leftarrow \{x_n \mid \|x_n - \mu_j\| \le \|x_n - \mu_i\|,\ \forall i\}.$$

更新步骤

对每个簇 $j = 1, \ldots, k$ 重新计算中心

$$\mu_j \leftarrow \frac{1}{|S_j|}\sum_{x_n \in S_j} x_n.$$

until 簇划分不再变化;

收敛性　由于目标函数

$$E_{\text{in}} = \sum_{j=1}^{k}\sum_{x_n \in S_j} \|x_n - \mu_j\|^2$$

在每一步交替优化中单调不增，算法必在有限步内收敛。

注 k-Means 是最流行的聚类算法之一；注意与 k-最近邻（k-NN）区分，二者同名但目的不同。

算法 30.3.2: RBF 网络结合 k-Means 的完整训练流程

输入: 训练集 $\{(x_n, y_n)\}_{n=1}^N$；原型数 M；RBF 类型（如高斯）

输出: 预测函数 $g_{\text{RBF}}(x)$

1. 无监督原型获取

运行 k-Means 算法簇数设为 $k = M$，得到原型中心

$$\mu_1, \mu_2, \ldots, \mu_M.$$

2. 构造特征变换

构建径向基特征映射

$$\Phi(x) = \Big[\text{RBF}(x, \mu_1),\ \text{RBF}(x, \mu_2),\ \ldots,\ \text{RBF}(x, \mu_M)\Big]^\top \in \mathbb{R}^M.$$

3. 监督线性模型训练

对变换后的数据

$$\big\{\big(\Phi(x_n),\, y_n\big)\big\}_{n=1}^N$$

执行线性模型（例如线性回归或逻辑回归），得到权重

$$\beta \in \mathbb{R}^M.$$

4. 返回模型

$$g_{\text{RBF}}(x) = \text{LinearHypothesis}(\beta, \Phi(x)).$$

注

- 步骤 1 利用无监督聚类（k-Means）完成特征提取，与自编码器思想类似；
- 典型超参数：原型数 M、RBF 带宽 γ（高斯核 γ）。

30.4 k-均值与 RBF 网络实践

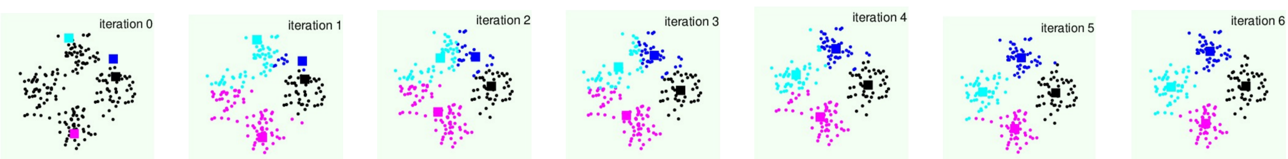


图 30.4.1: k-means 算法示例

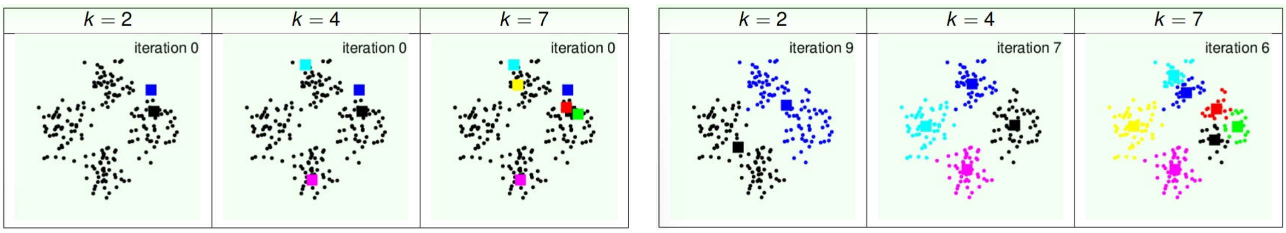
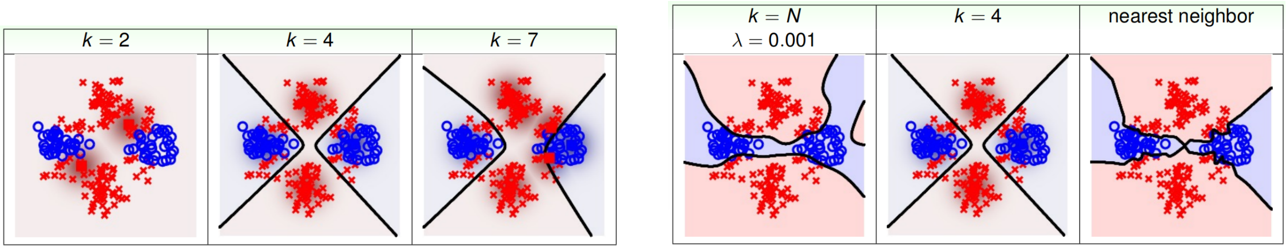


图 30.4.2: k-means 种类不同时的示例



(a) 使用 k 均值算法的径向基函数网络　　(b) 全径向基函数网络

图 30.4.3: 径向基函数网络示例

例题 30.3 选择题：RBF 网络与岭回归结合的正则化程度

当与岭线性回归结合时，以下哪种 RBF 网络的“正则化程度最高（most regularized）”?

1) 小 M 且小 λ
2) 小 M 且大 λ
3) 大 M 且小 λ
4) 大 M 且大 λ

解答　正确选项为 $\boxed{2}$。正则化程度由模型复杂度和正则化强度共同决定：

- M（RBF 隐藏层神经元数）越小，模型基础复杂度越低；
- λ（岭回归正则化参数）越大，对参数的惩罚越强，正则化效果越显著。

“最正则化”要求低复杂度模型 + 强正则化惩罚，即小 M 且大 λ。 ■

30.5 总结

笔记 [径向基函数网络]

- RBF 网络假设：以一组原型（prototypes）而非神经元作为非线性变换。
- RBF 网络学习：对原型“假设”进行线性组合。
- k-均值算法：通过交替优化实现聚类。
- k-均值与 RBF 网络实践：原型数量的恰当选择至关重要。