K-means 聚类及其变分表达、熵正则与空间正则

2025 年 6 月 16 日

目录

期末考纲

在本讲义中，重点围绕以下内容进行整理，并配合完整的推导和证明过程：

1. **K-means 聚类**

- 优化问题
- 主要迭代步骤
- 能量不增证明

2. **熵正则 K-means**

3. **空间正则**

- 两个常用的正则项能量式
- 空间正则 K-means 聚类 (优化问题，迭代步骤)

4. (补充) **TV 空间正则**的思路与简要求解。

1 分类的表达

给定一个样本 f，将其划分为 K 个类 $\Omega_1, \Omega_2, \Omega_3, \ldots, \Omega_K$，其类中心分别记为

$$c_1,\ c_2,\ \ldots,\ c_K,$$

并要求

$$\Omega_i \cap \Omega_j = \varnothing, \quad \bigcup_{k=1}^{K} \Omega_k = \Omega.$$

目标是将 f 分配给离 c_k 最近的类，亦即"就近"原则。

为表示第 k 类的示性函数，常用 $\mathbb{R}^K$ 的标准正交基

$$\boldsymbol{e} = \big(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_K\big) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \cdots \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

我们定义

$$u_k^* = \begin{cases} 1, & 若 f \in \Omega_k, \\ 0, & \text{else}, \end{cases} \qquad 并令 \quad \boldsymbol{u}^* = (\,u_1^*,\, u_2^*, \ldots, u_K^*)^T.$$

这样，如果 $f \in \Omega_k$，则 $\boldsymbol{u}^*$ 是第 k 个标准基向量 (也称 one-hot 向量)。若用 $|f - c_k|^2$ 代表"f 与类中心 c_k 的相似度的负数" (即用距离平方衡量差异)，则可得到如下变分形式来确定 $\boldsymbol{u}^*$:

Theorem 1.1 (Theorem 1.1)**.** 给定 f 以及 $c_1, c_2, \dots, c_K$。假设 $\min\big\{(f-c_1)^2, (f-c_2)^2, \dots, (f-c_K)^2\big\}$ 在某一点严格唯一取得。定义

$$\boldsymbol{u}^* = \arg\min_{\boldsymbol{u}\in\mathbb{U}} \sum_{k=1}^{K} u_k\,(f-c_k)^2,$$

其中 $\mathbb{U} = \Big\{\boldsymbol{u} = (u_1, \dots, u_K)\ :\ \sum_{k=1}^{K} u_k = 1,\ 0 \le u_k \le 1\Big\}$. 则其解为

$$u_k^* = \begin{cases} 1, & k = \arg\min\limits_{k'}\big\{(f-c_{k'})^2\big\}, \\ 0, & else. \end{cases}$$

证明. 设 $m = \min\limits_{k=1,\dots,K}\{(f-c_k)^2\}$. 容易看到，对于任何 $\boldsymbol{u} \in \mathbb{U}$,

$$\sum_{k=1}^{K} u_k\,(f-c_k)^2 \ \ge\ \sum_{k=1}^{K} u_k \cdot m \ =\ m\sum_{k=1}^{K} u_k \ =\ m.$$

当取本定理给出的 $\boldsymbol{u}^*$, 则 $\sum_{k=1}^{K} u_k^*(f-c_k)^2 = m$. 故此时 $\boldsymbol{u}^*$ 达到最小值。再利用 m 的唯一性可知解也唯一。如果 $\min\{\cdots\}$ 处有多个点, 则 $\boldsymbol{u}^*$ 只不过是多解中的任意一个极小元。 □

若把此问题从一维推广到二维，即在平面上给定样本 $f:\Omega\to\mathbb{R}$，分类问题一样成立，只需改写为

$$\min_{\boldsymbol{u}\in\mathbb{U}} \sum_{x\in\Omega}\sum_{k=1}^{K} u_k(x)\,\big(f(x)-c_k\big)^2,$$

并可得 $\boldsymbol{u}^*(x)$ 的解形同 Theorem 1.1 的结论。这样我们便将分类问题归结为一个变分问题:

$$\min_{\boldsymbol{u}\in\mathbb{U},\,\boldsymbol{c}} \mathcal{E}(\boldsymbol{u},\boldsymbol{c}) = \sum_{x\in\Omega}\sum_{k=1}^{K} u_k(x)\,\big[f(x)-c_k\big]^2.$$

它也就是 K-means 聚类算法的变分表达。

2 K-means 算法及其变分表达

2.1 K-means 算法

我们考虑

$$\min_{\boldsymbol{u},\,\boldsymbol{c}} \mathcal{E}(\boldsymbol{u},\boldsymbol{c}) = \sum_{x\in\Omega}\sum_{k=1}^{K} u_k(x)\,\big[f(x)-c_k\big]^2,$$

其中 $\boldsymbol{u}(x) \in [0,1]^K$ 并满足 $\sum_{k=1}^K u_k(x) = 1$. K-means 算法是通过在 $\boldsymbol{u}$ 与 $\boldsymbol{c}$ 上交替极小来实现的:

$$\begin{cases} \boldsymbol{u}^{t+1} = \arg\min\limits_{\boldsymbol{u}\in\mathbb{U}} \mathcal{E}\big(\boldsymbol{u},\, \boldsymbol{c}^t\big), \\ \boldsymbol{c}^{t+1} = \arg\min\limits_{\boldsymbol{c}} \mathcal{E}\big(\boldsymbol{u}^{t+1},\, \boldsymbol{c}\big). \end{cases}$$

(1) 更新 $\boldsymbol{u}$ 子问题由 Theorem 1.1 的思路可知，对于每个 x,

$$u_k^{t+1}(x) = \begin{cases} 1, & k = \arg\min\limits_{k'} \big\{ \big(f(x) - c_{k'}^t\big)^2 \big\}, \\ 0, & \text{else.} \end{cases} \tag{1}$$

(2) 更新 $\boldsymbol{c}$ 子问题由于 $\mathcal{E}\big(\boldsymbol{u}^{t+1},\, \boldsymbol{c}\big)$ 关于 c_k 呈严格凸，令其一阶导数为 0, 得到

$$c_k^{t+1} = \frac{\sum\limits_{x\in\Omega} u_k^{t+1}(x)\, f(x)}{\sum\limits_{x\in\Omega} u_k^{t+1}(x)}. \tag{2}$$

算法总结 (K-means 聚类):

1. **初始化**: 任意猜测类中心 $\boldsymbol{c}^0$.
2. **迭代** (for $t = 0, 1, 2, \ldots$):
   (a) 根据 (1) 式，更新示性函数 $\boldsymbol{u}^{t+1}$.
   (b) 根据 (2) 式，更新类中心 $\boldsymbol{c}^{t+1}$.
   (c) 若达到终止条件，结束；否则继续。

2.2 能量不增证明

Property 2.1 (K-means 能量单调不增). *由 K-means 更新产生的序列* $\{\boldsymbol{u}^t, \boldsymbol{c}^t\}$ *满足*

$$\mathcal{E}(\boldsymbol{u}^{t+1},\, \boldsymbol{c}^{t+1}) \;\le\; \mathcal{E}(\boldsymbol{u}^t, \boldsymbol{c}^t).$$

证明. 第一步:

$$\mathcal{E}\big(\boldsymbol{u}^{t+1}, \boldsymbol{c}^t\big) \;\le\; \mathcal{E}\big(\boldsymbol{u}^t, \boldsymbol{c}^t\big),$$

是由 $\boldsymbol{u}^{t+1}$ 相对 $\boldsymbol{u}$ 的最优性得到的。第二步:

$$\mathcal{E}\big(\boldsymbol{u}^{t+1}, \boldsymbol{c}^{t+1}\big) \;\le\; \mathcal{E}\big(\boldsymbol{u}^{t+1}, \boldsymbol{c}^t\big),$$

由 $\boldsymbol{c}^{t+1}$ 相对 $\boldsymbol{c}$ 的最优性得到。两者结合即得

$$\mathcal{E}\big(\boldsymbol{u}^{t+1}, \boldsymbol{c}^{t+1}\big) \;\le\; \mathcal{E}\big(\boldsymbol{u}^t, \boldsymbol{c}^t\big).$$

又因 $\mathcal{E} \ge 0$，故能量序列必有极限值。若 $\boldsymbol{u}, \boldsymbol{c}$ 的定义域是紧集，还可进一步证其收敛到某个临界点 (不保证全局最优)。 $\square$

2.3 简要记号与内积

为方便起见，常令

$$O_k(x) = \big[f(x) - c_k\big]^2, \quad \boldsymbol{O}(x) = \big(O_1(x), \ldots, O_K(x)\big).$$

定义内积

$$\langle \boldsymbol{O},\, \boldsymbol{u} \rangle = \sum_{x\in\Omega} \sum_{k=1}^{K} O_k(x)\, u_k(x).$$

于是 K-means 原问题便写成 $\min_{\boldsymbol{u},\, \boldsymbol{c}} \langle \boldsymbol{O},\, \boldsymbol{u} \rangle$.

3 熵正则化 (Entropy-regularized) K-means

3.1 熵正则与光滑化

当要对 K-means 做数值优化时，我们往往需要可微性更好的“软”分类，而不再是 0/1 示性函数。为此，我们可以在目标中添加一个熵正则项 $\varepsilon \langle \boldsymbol{u}, \ln \boldsymbol{u} \rangle$. 考虑

$$\min_{\boldsymbol{u}\in\mathbb{U},\, \boldsymbol{c}} \Big\{ \mathcal{E}(\boldsymbol{u}, \boldsymbol{c}) \;+\; \varepsilon\, \langle \boldsymbol{u},\, \ln \boldsymbol{u} \rangle \Big\},$$

其中 $\varepsilon > 0$ 较小时可以看作对 K-means 的 “小扰动”。记 $\varepsilon \langle \boldsymbol{u}, \ln \boldsymbol{u} \rangle = -\varepsilon\, \mathcal{H}(\boldsymbol{u})$, 其中 $\mathcal{H}$ 表示熵 (neg-entropy)。

3.1.1 熵的凸性与不等式

Property 3.1. *设 $u = (u_k)$ 满足 $\sum_{k=1}^{K} u_k = 1,\ 0 \le u_k \le 1$. 则*

$$\ln K \;\ge\; -\langle u,\, \ln u \rangle \;\ge\; -\sum_{k=1}^{K} u_k \ln u_k \;\ge\; 0.$$

左边等号当且仅当 $u_k = \frac{1}{K}$. 并且用约定 $0 \ln 0 = 0$.

证明. 令 $f(u) = -\ln u$. 则 $f''(u) = \frac{1}{u^2} > 0$, 故 f 为凸函数。当 $\sum_k \alpha_k = 1,\ \alpha_k \ge 0$ 时,

$$f\Big(\sum_k \alpha_k\, x_k\Big) \;\le\; \sum_k \alpha_k\, f(x_k).$$

选 $\alpha_k = u_k,\ x_k = \frac{1}{u_k}$. 则

$$-\ln\Big(\sum_k u_k \cdot \frac{1}{u_k}\Big) \;=\; -\ln K \;\;\le\; \sum_k u_k\, \ln u_k.$$

取两边加负号便得

$$\ln K \;\ge\; -\sum_k u_k \ln u_k.$$

另一方面，由于 $0 \le u_k \le 1$，显然 $\sum_k u_k \ln u_k \le 0$. 故证毕。 □

3.1.2 交替迭代求解

仿照前面的交替框架:

$$\begin{cases} \boldsymbol{u}^{t+1} = \arg\min\limits_{\boldsymbol{u}\in\mathbb{U}} \Big\{ \langle \boldsymbol{O},\, \boldsymbol{u} \rangle \;-\; \varepsilon\, \mathcal{H}(\boldsymbol{u}) \Big\}, \\ \boldsymbol{c}^{t+1} = \arg\min\limits_{\boldsymbol{c}} \Big\{ \langle \boldsymbol{O},\, \boldsymbol{u}^{t+1} \rangle \Big\}. \end{cases}$$

记 $\boldsymbol{O}$ 关于 $\boldsymbol{c}$ 的更新同 K-means，一阶取零即可得 (2) 式:

$$c_k^{t+1} = \frac{\sum_x u_k^{t+1}(x)\, f(x)}{\sum_x u_k^{t+1}(x)}.$$

关键在于如何解 $\boldsymbol{u}$ 子问题. 我们写出拉格朗日形式 (注意约束 $\sum_k u_k(x) = 1$):

$$\mathcal{L}(\boldsymbol{u}, \boldsymbol{\lambda}) = \langle \boldsymbol{O}, \boldsymbol{u} \rangle \;-\; \varepsilon\, \langle \boldsymbol{u}, \ln \boldsymbol{u} \rangle \;+\; \sum_{x\in\Omega} \lambda(x) \Big(\sum_k u_k(x) - 1 \Big).$$

对 $u_k(x)$ 做变分分析，令其一阶导数为 0，可得

$$O_k(x) \;+\; \varepsilon\,(\ln u_k(x) + 1) \;+\; \lambda(x) = 0,$$

进而

$$\ln u_k(x) = \frac{-O_k(x) - \lambda(x)}{\varepsilon} - 1, \quad \Longrightarrow \quad u_k(x) = \exp\Big[\frac{-\,O_k(x) - \lambda(x)}{\varepsilon} - 1 \Big].$$

因为 $\sum_{k=1}^K u_k(x) = 1$，将上式对 k 求和得到

$$1 = \sum_k \exp\Big[\frac{-\,O_k(x) - \lambda(x)}{\varepsilon} - 1 \Big] = \exp\Big[\frac{-\lambda(x)}{\varepsilon} - 1 \Big] \sum_k \exp\Big[\frac{-\,O_k(x)}{\varepsilon} \Big].$$

$\therefore$

$$\exp\Big[\frac{-\lambda(x)}{\varepsilon} - 1 \Big] = \frac{1}{\sum_k \exp\big(\frac{-O_k(x)}{\varepsilon} \big)},$$

于是

$$u_k(x) = \frac{\exp\Big[\frac{-\,O_k(x)}{\varepsilon} \Big]}{\sum_{k'=1}^K \exp\Big[\frac{-\,O_{k'}(x)}{\varepsilon} \Big]} = \mathrm{Softmin}_\varepsilon\big(\boldsymbol{O}(x) \big)_k.$$

这即对应于 Softmax 的变体。且当 $\varepsilon \to 0^+$, 则 $\boldsymbol{u}^{t+1}(x)$ 会退化成 K-means 中的 0/1 分类。小结: ε 越大，表示对“二值约束”松弛程度越强，$\boldsymbol{u}$ 越“软”; ε 越小，$\boldsymbol{u}$ 越贴近 0/1，逐渐逼近硬分类。

Remark 3.2. 在计算机中实现时，为避免 exp(大数) 溢出，也常采用

$$\mathrm{Softmax}(\boldsymbol{O}) = \frac{\exp(O_k - M)}{\sum \exp(O_k - M)}, \quad 其中\, M = \max\{O_1, \ldots, O_K\}.$$

4 空间正则化 (Spatial Regularization)

4.1 模型的构造

K-means 及上述 Softmin/max 主要依赖“点与点的差异”，即 $\boldsymbol{O}(x)$ 中每个分量都只跟 $f(x)$ 及 c_k 有关，而 **没有**考虑到图像/数据 邻域中是否期望“同类”。在真实情形中，常常要求邻域性质：若 x 附近的 $\mathbb{N}_x$ “看起来”很像 x 本身，那么它们应当具有“相同”或“相似”的分类标签。为增强对局部平滑一致性的偏好，可再加一个正则项 $\mathcal{R}(\boldsymbol{u})$.

例如，设 $\mathbb{N}_x$ 为 x 的某个邻域，如果 $\boldsymbol{u}(x)$ 在该邻域里变动太大，就要给出额外的惩罚。最简单的做法是：若 $u_k(x) = 1$，则希望 $\forall y \in \mathbb{N}_x$, $u_k(y) = 1$. 用罚函数来逼迫这一点，可考虑

$$\mathcal{R}(\boldsymbol{u}) = \sum_x \sum_{k=1}^{K} u_k(x) \sum_{y \in \mathbb{N}_x} \big[\, 1 - u_k(y) \big].$$

该和式在“同类”时加 0 (无罚)，在“不同类”时加正值。

4.1.1 卷积形式 (邻域核函数)

当 $\mathbb{N}_x$ 表示距离 $\|y - x\| < r$ 的所有像素集合，可用指示函数 $\mathbf{1}_{\mathbb{N}}(z)$ 表示 $\mathbb{N} = \{\, z : \|z\| < r\}$. 故

$$\mathcal{R}(\boldsymbol{u}) = \sum_x \sum_{k=1}^{K} u_k(x) \sum_y \mathbf{1}_{\mathbb{N}}(x-y) \left[1 - u_k(y)\right] = \left\langle \boldsymbol{u},\, k * (1 - \boldsymbol{u}) \right\rangle, \mathbf{1}_{\mathbb{N}}(x-y) = \begin{cases} 1, & y \in \mathbb{N} \\ 0, & y \in \Omega \backslash \mathbb{N} \end{cases}$$

其中 $k(z)$ 为核函数 (邻域权)，$k*$ 表示卷积。

4.1.2 空间正则化 K-means 的变分问题

综上，我们可得带空间正则项的 K-means 形式 (只写离散型):

$$\min_{\boldsymbol{u} \in \mathbb{U}} \Big\langle \boldsymbol{O}, \boldsymbol{u} \Big\rangle \; + \; \lambda \left\langle \boldsymbol{u},\, k * (1 - \boldsymbol{u}) \right\rangle,$$

若类中心 $\boldsymbol{c}$ 也未知，则完整地写为

$$\min_{\boldsymbol{u} \in \mathbb{U}, \boldsymbol{c}} \sum_{k=1}^{K} \sum_{x \in \Omega} u_k(x) \left[f(x) - c_k\right]^2 \; + \; \lambda \sum_{k=1}^{K} \sum_{x \in \Omega} u_k(x) \left[k * (1 - \boldsymbol{u})\right]_k (x).$$

其中 $\lambda \geq 0$ 为正则化系数。这样我们在分类时，也兼顾到若干邻域平滑性/一致性。

4.2 求解模型：DC 算法 (Difference of Convex)

我们发现 $\mathcal{R}(\boldsymbol{u}) = \langle \boldsymbol{u}, k * (1-\boldsymbol{u})\rangle$ 里包含了 $\boldsymbol{u}^T k\, \boldsymbol{u}$ 这样的二次项，很可能会让问题变得非凸 (或者更复杂)。假如 k 是正定核函数，则可证明 $\langle \boldsymbol{u}, k * \boldsymbol{u}\rangle$ 为正项，从而 $\boldsymbol{u}^T k \boldsymbol{u}$ 是凸；$\mathcal{R}(\boldsymbol{u})$ 里却是 $\boldsymbol{u} \cdot (1-\boldsymbol{u})$，通常被视为凹的一部分。因此可将 $\langle \boldsymbol{u}, k * (1-\boldsymbol{u})\rangle$ 拆分为凸 - 凸的形式。这类问题可用 DC 迭代 (difference of convex) 来做：

4.2.1 DC 基本思路与证明

设我们想最小化 $H(x) = f(x) - g(x)$，其中 f 与 g 皆为凸函数。在第 t 步时，令

$$\widetilde{H}(x) = f(x) \;-\; \big[g(x^t) + g'(x^t)\,(x - x^t)\big].$$

其中 $g'(x^t)$ 是 g 在 x^t 处的梯度或次梯度。于是

$$x^{t+1} = \arg\min_x \widetilde{H}(x), \qquad \widetilde{H}(x) = f(x) - \big[g(x^t) + \langle g'(x^t),\, x - x^t\rangle\big].$$

可被看作：

$$x^{t+1} = \arg\min_x \Big\{ f(x) \;+\; \langle g'(x^t),\, x\rangle \Big\},$$

省略常数项不影响最小点。

Property 4.1. 若 $\{x^t\}$ 由上式产出，则序列能量不增*:*

$$H(x^{t+1}) \;\le\; H(x^t) =\; f(x^t) - g(x^t).$$

证明. 仿照凸分析的思想，对 g 做 Taylor 展开 (或者一阶下界)：

$$g(x^{t+1}) \ge g(x^t) \;+\; \big\langle g'(x^t), (x^{t+1} - x^t)\big\rangle.$$

故

$$\begin{aligned} H(x^{t+1}) - H(x^t) &= [\, f(x^{t+1}) - f(x^t)\,] \;-\; [\, g(x^{t+1}) - g(x^t)\,] \\ &\le\; f(x^{t+1}) - f(x^t) \;-\; \langle g'(x^t),\, x^{t+1} - x^t\rangle. \end{aligned}$$

另一方面，$\widetilde{H}(x^{t+1}) \le \widetilde{H}(x^t)$ 意味着

$$f(x^{t+1}) \;-\; \big\langle g'(x^t),\, x^{t+1}\big\rangle \;\le\; f(x^t) \;-\; \big\langle g'(x^t),\, x^t\big\rangle,$$

即

$$f(x^{t+1}) - f(x^t) \;-\; \langle g'(x^t), (x^{t+1} - x^t)\rangle \;\le\; 0.$$

合并可得 $H(x^{t+1}) \le H(x^t)$. 证毕。 □

4.2.2 空间正则 K-means 的 DC 形式

回到 $\min_{\boldsymbol{u}}\Big\langle \boldsymbol{O},\,\boldsymbol{u}\Big\rangle \;+\; \lambda\Big\langle \boldsymbol{u},\,k*(1-\boldsymbol{u})\Big\rangle$. 我们记

$$\mathcal{F}(\boldsymbol{u}) = \langle \boldsymbol{O},\,\boldsymbol{u}\rangle, \quad \mathcal{G}(\boldsymbol{u}) = -\lambda\big\langle \boldsymbol{u},\,k*(1-\boldsymbol{u})\big\rangle,$$

而 $H(\boldsymbol{u}) = \mathcal{F}(\boldsymbol{u}) - \mathcal{G}(\boldsymbol{u})$. DC 迭代即:

$$\boldsymbol{u}^{t+1} = \arg\min_{\boldsymbol{u}\in\mathbb{U}}\Big\{\langle \boldsymbol{O},\,\boldsymbol{u}\rangle \;-\; \Big\langle \frac{\delta\mathcal{G}}{\delta\boldsymbol{u}}\big|_{\boldsymbol{u}^t},\;\boldsymbol{u}\Big\rangle\Big\},$$

即

$$\boldsymbol{u}^{t+1} = \arg\min_{\boldsymbol{u}\in\mathbb{U}}\Big\{\big\langle \boldsymbol{O} - \tfrac{\delta\mathcal{G}}{\delta\boldsymbol{u}}\big|_{\boldsymbol{u}^t},\;\boldsymbol{u}\big\rangle\Big\}.$$

只要 $\boldsymbol{O} - \frac{\delta\mathcal{G}}{\delta\boldsymbol{u}}(\boldsymbol{u}^t)$ 可明确表达，则此时对每个像素 x，我们在 K 个分量中 “选最小者为 1，其它为 0”，与 K-means 类似。

计算 $\frac{\delta\mathcal{G}}{\delta\boldsymbol{u}}$ 因为 $\mathcal{G} = -\lambda\,\langle \boldsymbol{u},\,k*(1-\boldsymbol{u})\rangle$, 故

$$\frac{\delta\mathcal{G}}{\delta u_k} = -\lambda\,\frac{\delta}{\delta u_k}\Big\langle u_k,\;k*(1-u_k)\Big\rangle.$$

国外文献常做类似推导，这里省略部分交换积分之类的细节，结论得到

$$\frac{\delta\mathcal{G}}{\delta u_k} = -\lambda\,k*\big[\,1-2\,u_k\big].$$

(其中 k 常假设对称 $k(z)=k(-z)$). 由此，(9) 式的 $\boldsymbol{u}$-子问题成为:

$$u_k^{t+1}(x) = \begin{cases} 1, & k = \arg\min_{k'}\Big\{\,O_{k'}(x) \;+\; \lambda\,[\,k*(1-2\,u_{k'}^t)\,](x)\Big\}, \\ 0, & \text{else.}\end{cases}$$

类似地，若要估计中心 $\boldsymbol{c}$，可与 K-means 同理进行 (2) 式更新。

5 TV 空间正则 (补充)

另一种常见的空间正则方式是使用全变差 (TV):

$$TV(u) = \int_\Omega \big|\nabla u(\boldsymbol{x})\big|\,d\boldsymbol{x},$$

它本质上惩罚函数在邻域内过大的梯度变动，从而鼓励局部分块常值的结构。

5.1 TV 正则的对偶表示

常见结论 (如文献中的 Rudin–Osher–Fatemi 模型等):

$$TV(u) = \max_{\boldsymbol{p}\in\mathbb{P}}\Big\{\int_\Omega u(\boldsymbol{x})\ \operatorname{div}\boldsymbol{p}(\boldsymbol{x})\,d\boldsymbol{x}\Big\},$$

其中 $\mathbb{P} = \{\boldsymbol{p}\in C_0^1(\Omega), \|\boldsymbol{p}\|_\infty \le 1\}$. 对向量值 $\boldsymbol{u}$，则

$$TV(\boldsymbol{u}) = \sum_{k=1}^{K} TV(u_k) = \max_{(\boldsymbol{p}_1,\dots,\boldsymbol{p}_K)\in\mathbb{P}}\Big\{\sum_{k=1}^{K}\int_\Omega u_k(\boldsymbol{x})\operatorname{div}(\boldsymbol{p}_k)(\boldsymbol{x})\,d\boldsymbol{x}\Big\}.$$

由鞍点原理，可以做类似的 p-子问题与 u-子问题的交替，$\boldsymbol{p}$-子问题通常用投影到 $\|\boldsymbol{p}\|\le 1$ 的梯度下降。最终，可以得到 TV 正则分类/分割的迭代算法。在此，我们不再详尽展开 (可参见 ROF 模型等资料)。只需知道：TV 正则与卷积正则一样都强调局部一致性，但 TV 强调更“分段常数块”的结构，适合图像分割等任务。

6 内容总结

从以下四方面对分类与聚类的变分思想做了系统梳理:

1. **K-means 聚类**.

- 优化问题:

$$\min_{\boldsymbol{u},\boldsymbol{c}}\sum_{x\in\Omega}\sum_{k=1}^{K}u_k(x)\left[f(x)-c_k\right]^2,\quad \boldsymbol{u}(x)\in\{0,1\}^K,\ \sum_{k=1}^{K}u_k(x)=1.$$

- 主要迭代步骤: (1) 式 + (2) 式交替。

- 能量不增证明: 见 §2.2.

2. **熵正则 K-means**. 在原目标里加入 $\varepsilon\langle\boldsymbol{u},\ln\boldsymbol{u}\rangle$ 构成可微的“软化”分类，迭代公式类似 Softmin/Softmax。并利用拉格朗日方法推导过溢出处理 (exp(大数))。

3. **空间正则**

- 两个正则项能量式：(1) 卷积 (邻域) 型 $\mathcal{R}(\boldsymbol{u})$, (2) TV 型 $\sum|\nabla u_k|$.

- 空间正则 K-means 聚类:

$$\min_{\boldsymbol{u},\boldsymbol{c}}\sum_{k=1}^{K}\sum_{x}u_k(x)\left[f(x)-c_k\right]^2+\lambda\sum_{k=1}^{K}\sum_{x}u_k(x)\left[k*(1-\boldsymbol{u})\right]_k(x).$$

(或把 $\sum|\nabla u_k|$ 替换上去)

- (优化问题，迭代步骤) 通过 DC 算法或者鞍点法做求解。对 $\boldsymbol{u}$ 子问题采用与 K-means 类似的“选最小 index ”，对 $\boldsymbol{c}$ 子问题采用 (2) 式。

4. **TV 正则**

- 全变差 $TV(\boldsymbol{u})$ 关于 $\boldsymbol{u}$ 是凸的。
- 对偶表示法: $TV(\boldsymbol{u}) = \max_{\boldsymbol{p}\in\mathbb{P}} \int \boldsymbol{u} \operatorname{div} \boldsymbol{p}$.
- 具体在分类/分割中根据鞍点原理拆成 $\boldsymbol{p}$-子问题和 $\boldsymbol{u}$-子问题交替进行。

这些内容为期末考纲中所需掌握的重点公式、推导及算法思路。本讲义自始至终遵照了变分方法 (energy minimization) 的思路，由最简单的硬分类 (K-means) 逐步引入熵正则 (Softmax) 与空间正则 (邻域一致或 TV)，打通了分类和分割中核心的变分分析路径。