# Supplementary Material: Entropy Weighted Power $k$-means Clustering


**Saptarshi Chakraborty*** Indian Statistical Institute, Kolkata, India

**Debolina Paul*** Indian Statistical Institute, Kolkata, India

**Swagatam Das** Electronics and Communication Sciences Unit, Indian Statistical Institute

**Jason Xu** Department of Statistical Science, Duke University


## 1 Derivation of Closed Form Updates

Consider the minimization problem

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\phi_{ij}\|\boldsymbol{x}_i-\boldsymbol{\theta}_j\|_{\boldsymbol{w}}^2+\lambda\sum_{l=1}^{p}w_l\log w_l \tag{1}$$

with respect to optimization variables $\boldsymbol{\Theta}$ and $\boldsymbol{w}$. The minimization over $\boldsymbol{\Theta}$ is straightforward, and the optimal solutions are given by

$$\boldsymbol{\theta}_j^*=\frac{\sum_{i=1}^{n}\phi_{ij}\boldsymbol{x}_i}{\sum_{i=1}^{n}\phi_{ij}}.$$

Now to minimize equation 1 in $\boldsymbol{w}$, we consider the Lagrangian

$$\mathcal{L}=\sum_{i=1}^{n}\sum_{j=1}^{k}\phi_{ij}\|\boldsymbol{x}_i-\boldsymbol{\theta}_j\|_{\boldsymbol{w}}^2+\lambda\sum_{l=1}^{p}w_l\log w_l-\alpha(\sum_{l=1}^{p}w_l-1).$$

The optimality condition $\frac{\partial\mathcal{L}}{\partial w_l}=0$ implies $\sum_{i=1}^{n}\sum_{j=1}^{k}\phi_{ij}(x_{il}-\theta_{jl})^2+\lambda(1+\log w_l)-\alpha=0$. This further implies that

$$w_l^*\propto\exp\left\{-\frac{\sum_{i=1}^{n}\sum_{j=1}^{k}\phi_{ij}(x_{il}-\theta_{jl})^2}{\lambda}\right\}.$$

Now enforcing the constraint $\sum_{l=1}^{p}w_l=1$, we get

$$w_l^*=\frac{\exp\left\{-\frac{\sum_{i=1}^{n}\sum_{j=1}^{k}\phi_{ij}(x_{il}-\theta_{jl})^2}{\lambda}\right\}}{\sum_{t=1}^{p}\exp\left\{-\frac{\sum_{i=1}^{n}\sum_{j=1}^{k}\phi_{ij}(x_{it}-\theta_{jt})^2}{\lambda}\right\}}.$$

## 2 Proof of Theorem 1

**Theorem 1** Let $s\le 1$ also let $(\boldsymbol{\Theta}_{n,s},\boldsymbol{w}_{n,s})$ be minimizer of $f_s(\boldsymbol{\Theta},\boldsymbol{w})$. Then we have $\boldsymbol{\Theta}_{n,s}\in C^k$.

*Proof.* Let $P_C^{\boldsymbol{w}}(\boldsymbol{\theta})$ denote the projection of $\boldsymbol{\theta}$ onto $C$ w.r.t. the $\|\cdot\|_{\boldsymbol{w}}$ norm. Now for any $\boldsymbol{v}\in C$, using the obtise angle condition, we obtain, $\langle\boldsymbol{\theta}-P_C^{\boldsymbol{w}}(\boldsymbol{\theta}),\boldsymbol{v}-P_C^{\boldsymbol{w}}(\boldsymbol{\theta})\rangle_{\boldsymbol{w}}\le 0$. Since $\boldsymbol{x}_i\in C$, we obtain,

$$\begin{aligned}\|\boldsymbol{x}_i-\boldsymbol{\theta}_j\|_{\boldsymbol{w}}^2&=\|\boldsymbol{x}_i-P_C^{\boldsymbol{w}}(\boldsymbol{\theta}_j)\|_{\boldsymbol{w}}^2+\|P_C^{\boldsymbol{w}}(\boldsymbol{\theta}_j)-\boldsymbol{\theta}_j\|_{\boldsymbol{w}}^2\\&\quad-2\langle\boldsymbol{\theta}-P_C^{\boldsymbol{w}}(\boldsymbol{\theta}_j),\boldsymbol{x}_i-P_C^{\boldsymbol{w}}(\boldsymbol{\theta}_j)\rangle_{\boldsymbol{w}}\\&\ge\|\boldsymbol{x}_i-P_C^{\boldsymbol{w}}(\boldsymbol{\theta}_j)\|_{\boldsymbol{w}}^2+\|P_C^{\boldsymbol{w}}(\boldsymbol{\theta}_j)-\boldsymbol{\theta}_j\|_{\boldsymbol{w}}^2.\end{aligned}$$

Now since, $M_s(\cdot)$ is an increasing function in each of its argument, if we replace $\boldsymbol{\theta}_j$ by $P_C^{\boldsymbol{w}}(\boldsymbol{\theta}_j)$ in $M_s(\|\boldsymbol{x}_i-\boldsymbol{\theta}_1\|_{\boldsymbol{w}}^2,\dots,\|\boldsymbol{x}_i-\boldsymbol{\theta}_k\|_{\boldsymbol{w}}^2)$, the objective function value doesn't go up. Thus we can effectively restrict our attention to $C^k$. Now since the function $f_s(\cdot,\cdot)$ is continuous on the compact set $C^k\times[0,1]^p$, it attains its minimum on $C^k\times[0,1]^p$. Thus, $\boldsymbol{\Theta}^*\in C^k$. □

## 3 Proof of Theorem 2

**Theorem 2** For any decreasing sequence $\{s_m\}_{m=1}^{\infty}$ such that $s_1\le 1$ and $s_m\to-\infty$, $f_{s_m}(\boldsymbol{\Theta},\boldsymbol{w})$ converges uniformly to $f_{-\infty}(\boldsymbol{\Theta},\boldsymbol{w})$ on $C^k\times[0,1]^p$.

For any $(\boldsymbol{\Theta},\boldsymbol{w})\in C^k\times[0,1]^p$, $f_{s_m}(\boldsymbol{\Theta},\boldsymbol{w})$ converges monotonically to $f_{-\infty}(\boldsymbol{\Theta},\boldsymbol{w})$ (this is due to the power mean inequality). Since $C^k\times[0,1]^p$ is compact, the result follows immediately upon applying Dini's theorem from real analysis.

## 4 Proof Details for Uniform Strong Law of Large Numbers

**Theorem 3** (SLLN) Fix $s\le 1$. Let $\mathcal{G}$ denote the family of functions $g_{\boldsymbol{\Theta},\boldsymbol{w}}(\boldsymbol{x})=M_s(\|\boldsymbol{x}-\boldsymbol{\theta}_1\|_{\boldsymbol{w}}^2,\dots,\|\boldsymbol{x}-\boldsymbol{\theta}_k\|_{\boldsymbol{w}}^2)$. Then $\sup_{g\in\mathcal{G}}|\int gdP_n-\int gdP|\to 0$ a.s. $[P]$.

Fix $\epsilon>0$. It is enough to find a finite family of functions $\mathcal{G}_\epsilon$ such that for all $g\in\mathcal{G}$, there exists $\bar{g},\dot{g}\in\mathcal{G}_\epsilon$ such that $\dot{g}\le g\le\bar{g}$ and $\int(\bar{g}-\dot{g})dP<\epsilon$.

Let us define $\phi(\cdot):\mathbb{R}\to\mathbb{R}$ such that $\phi(x)=\max\{0,x\}$. Since $C$ is compact, for every $\delta_1>0$, we can always construct a finite set $C_{\delta_1}\subset C$ such that if $\boldsymbol{\theta}\in C$, there exist $\boldsymbol{\theta}'\in C_{\delta_1}$ such that $\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|<\delta_1$. Similarly, resorting to the compactness of $[0,1]^p$, for every $\delta_2 > 0$, we can always construct a finite set $W_{\delta_2} \subset [0,1]^p$ such that if $\boldsymbol{w} \in [0,1]^p$, there exist $\boldsymbol{w'} \in W_{\delta_2}$ such that $\|\boldsymbol{w} - \boldsymbol{w'}\| < \delta_2$. Consider the function $h(\boldsymbol{x}, \boldsymbol{\Theta}, \boldsymbol{w}) = M_s(\|\boldsymbol{x} - \boldsymbol{\theta}_1\|^2_{\boldsymbol{w}}, \ldots, \|\boldsymbol{x} - \boldsymbol{\theta}_k\|^2_{\boldsymbol{w}})$ on $C \times C^k \times [0,1]^p$. $h$, being continuous on the compact set $C \times C^k \times [0,1]^p$, is also uniformly continuous. Thus for all $\mathbf{x} \in C$, if $\|\boldsymbol{w} - \boldsymbol{w'}\| < \delta_2$ and $\|\boldsymbol{\theta}_j - \boldsymbol{\theta}'_j\| < \delta_1$ for all $j = 1, \ldots, k$ implies that

$$\Big| M_s(\|\boldsymbol{x} - \boldsymbol{\theta}_1\|^2_{\boldsymbol{w}}, \ldots, \|\boldsymbol{x} - \boldsymbol{\theta}_k\|^2_{\boldsymbol{w}}) - M_s(\|\boldsymbol{x} - \boldsymbol{\theta}'_1\|^2_{\boldsymbol{w'}}, \ldots, \|\boldsymbol{x} - \boldsymbol{\theta}'_k\|^2_{\boldsymbol{w'}}) \Big| < \epsilon/2 \qquad (2)$$

We take

$$\mathcal{G}_\epsilon = \{\phi(M_s(\|\boldsymbol{x} - \boldsymbol{\theta}'_1\|^2_{\boldsymbol{w'}}, \ldots, \|\boldsymbol{x} - \boldsymbol{\theta}'_k\|^2_{\boldsymbol{w'}}) \pm \epsilon/2) : \boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_k \in C_{\delta_1} \text{ and } \boldsymbol{w'} \in W_{\delta_2}\}.$$

Now if we take

$$\bar{g}_{\boldsymbol{\theta},\boldsymbol{w}} = \phi(M_s(\|\boldsymbol{x} - \boldsymbol{\theta}'_1\|^2_{\boldsymbol{w'}}, \ldots, \|\boldsymbol{x} - \boldsymbol{\theta}'_k\|^2_{\boldsymbol{w'}}) + \epsilon/2)$$

and

$$\dot{g}_{\boldsymbol{\theta},\boldsymbol{w}} = \phi(M_s(\|\boldsymbol{x} - \boldsymbol{\theta}'_1\|^2_{\boldsymbol{w'}}, \ldots, \|\boldsymbol{x} - \boldsymbol{\theta}'_k\|^2_{\boldsymbol{w'}}) - \epsilon/2),$$

where $\boldsymbol{\theta}'_j \in C_{\delta_1}$ and $\boldsymbol{w} \in W_{\delta_2}$ for $j = 1, \ldots, k$ such that $\|\boldsymbol{\theta}_j - \boldsymbol{\theta}'_j\| < \delta_1$ and $\|\boldsymbol{w} - \boldsymbol{w'}\| < \delta_2$. From equation (2), we get, $\dot{g} \leq g \leq \bar{g}$. Now we need to show $\int(\bar{g} - \dot{g})dP < \epsilon$. This step is straight forward.

$$\begin{aligned}
&\int (\bar{g} - \dot{g})dP \\
&= \Big[\phi(M_s(\|\boldsymbol{x} - \boldsymbol{\theta}'_1\|^2_{\boldsymbol{w'}}, \ldots, \|\boldsymbol{x} - \boldsymbol{\theta}'_k\|^2_{\boldsymbol{w'}}) + \epsilon/2) \\
&- \phi(M_s(\|\boldsymbol{x} - \boldsymbol{\theta}'_1\|^2_{\boldsymbol{w'}}, \ldots, \|\boldsymbol{x} - \boldsymbol{\theta}'_k\|^2_{\boldsymbol{w'}}) - \epsilon/2)\Big] dP \\
&\leq \epsilon \int dP = \epsilon.
\end{aligned}$$

Hence the result.

## 5 Proof Details of Main Consistency Result

**Theorem 4** Under the condition A1, $\boldsymbol{\Theta}_{n,s} \xrightarrow{a.s.} \boldsymbol{\Theta}^*$ and $\boldsymbol{w}_{n,s} \xrightarrow{a.s.} \boldsymbol{w}^*$ as $n \to \infty$ and $s \to -\infty$.

*Proof.* It is enough to show that given any neighbourhood $N$ of $(\boldsymbol{\Theta}^*, \boldsymbol{w}^*)$, there exists $M_1 < 0$ and $M_2 > 0$ such that if $s < M_1$ and $n > M_2$ such that $(\boldsymbol{\Theta}, \boldsymbol{w}) \in N$ almost surely. By assumption A1, it is enough to show that for all $\eta > 0$, there exists $M_1 < 0$ and $M_2 > 0$ such that if $s < M_1$ and $n > M_2$ such that $\Phi(\boldsymbol{\Theta}, \boldsymbol{w}) \leq \Phi(\boldsymbol{\Theta}^*, \boldsymbol{w}^*) + \eta$ almost surely. For notational convenience, we write $\mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}, \boldsymbol{w})$ for $M_s(\|\boldsymbol{x} - \boldsymbol{\theta}_1\|^2_{\boldsymbol{w}}, \ldots, \|\boldsymbol{x} - \boldsymbol{\theta}_k\|^2_{\boldsymbol{w}})$ and $\alpha(\boldsymbol{w}) = \lambda \sum_{l=1}^p w_l \log w_l$. Now since $(\boldsymbol{\Theta}_{n,s}, \boldsymbol{w}_{n,s})$ is the minimizer for $\int \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}, \boldsymbol{w})dP_n + \lambda \sum_{l=1}^p w_l \log w_l$, we get,

$$\begin{aligned}
&\int \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}_{n,s}, \boldsymbol{w}_{n,s})dP_n + \lambda\alpha(\boldsymbol{w}_{n,s}) \\
&\leq \int \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}^*, \boldsymbol{w}^*)dP_n + \lambda\alpha(\boldsymbol{w}^*). \qquad (3)
\end{aligned}$$

Now observe that $\Phi(\boldsymbol{\Theta}_{n,s}, \boldsymbol{w}_{n,s}) - \Phi(\boldsymbol{\Theta}^*, \boldsymbol{w}^*) = \xi_1 + \xi_2 + \xi_3$, where,

$$\xi_1 = \Phi(\boldsymbol{\Theta}_{n,s}, \boldsymbol{w}_{n,s}) - \int \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}_{n,s}, \boldsymbol{w}_{n,s})dP - \lambda\alpha(\boldsymbol{w}_{n,s}),$$

$$\xi_2 = \int \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}_{n,s}, \boldsymbol{w}_{n,s})dP - \int \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}_{n,s}, \boldsymbol{w}_{n,s})dP_n,$$

$$\xi_3 = \int \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}_{n,s}, \boldsymbol{w}_{n,s})dP_n + \lambda\alpha(\boldsymbol{w}_{n,s}) - \Phi(\boldsymbol{\Theta}^*, \boldsymbol{w}^*).$$

We first choose $M_1 < 0$ such that if $s < M_1$ then

$$\Big| \min_{1 \leq j \leq k} \|\boldsymbol{x} - \boldsymbol{\theta}_j\|_{\boldsymbol{w}} - \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}, \boldsymbol{w}) \Big| < \eta/6 \qquad (4)$$

for all $\boldsymbol{x} \in C$, $\boldsymbol{\Theta} \in C^k$ and $\boldsymbol{w} \in [0,1]^p$. Thus for $s < M_1$, $\min_{1\leq j\leq k} \|\boldsymbol{x} - \boldsymbol{\theta}_j\|_{\boldsymbol{w}} \leq \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}, \boldsymbol{w}) + \eta/6$ which in turn implies that $\int \min_{1\leq j\leq k} \|\boldsymbol{x} - \boldsymbol{\theta}_j\|_{\boldsymbol{w}} dP_n \leq \int \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}, \boldsymbol{w})dP_n + \eta/3$. Substituting $\boldsymbol{\Theta}_{n,s}$ for $\Theta$ and $\boldsymbol{w}_{n,s}$ for $\boldsymbol{w}$ in the above expression and adding $\lambda\alpha(\boldsymbol{w}_{n,s})$ to both sides, we get $\xi_1 < \eta/6$. We also observe that the quantity $\xi_2$ can also be made smaller that $\eta/3$ by appealing to the uniform SLLN (Theorem 3). Now to bound $\xi_3$, we observe that

$$\begin{aligned}
\xi_3 &\leq \int \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}^*, \boldsymbol{w}^*)dP_n + \lambda\alpha(\boldsymbol{w}^*) - \Phi(\boldsymbol{\Theta}^*, \boldsymbol{w}^*) \\
&= \int \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}^*, \boldsymbol{w}^*)dP_n - \int \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}^*} \|\boldsymbol{x} - \boldsymbol{\theta}\|_{\boldsymbol{w}^*} dP
\end{aligned}$$

This inequality is obtained by appealing to equation (3). Again appealing to the uniform SLLN, we get that for large enough $n$,

$$\begin{aligned}
\xi_3 &\leq \int \mathcal{M}_s(\boldsymbol{x}, \boldsymbol{\Theta}^*, \boldsymbol{w}^*)dP - \int \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}^*} \|\boldsymbol{x} - \boldsymbol{\theta}\|_{\boldsymbol{w}^*} dP + \eta/6 \\
&\leq \int [\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}^*} \|\boldsymbol{x} - \boldsymbol{\theta}\|_{\boldsymbol{w}^*} + \eta/6]dP - \int \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}^*} \|\boldsymbol{x} - \boldsymbol{\theta}\|_{\boldsymbol{w}^*} dP \\
&+ \eta/6 = \eta/3.
\end{aligned}$$

The second inequality follows from equation (4). Thus we get, $\Phi(\boldsymbol{\Theta}_{n,s}, \boldsymbol{w}_{n,s}) - \Phi(\boldsymbol{\Theta}^*, \boldsymbol{w}^*) = \xi_1 + \xi_2 + \xi_3 \leq \eta/6 + \eta/3 + \eta/3 < \eta$ almost surely. Hence the result. □

Table S1: NMI values for Simulation 1, showing the effect of increasing number of clusters.

| | $d = 5$ | $d = 10$ | $d = 20$ | $d = 50$ | $d = 100$ |
|---|---|---|---|---|---|
| $k$-means | 0.3913 (0.002) | 0.3701 (0.002) | 0.3674 (0.003) | 0.3629(0.002) | 0.3517 (0.003) |
| $WK$-means | 0.5144(0.002) | 0.50446(0.003) | 0.5050(0.003) | 0.5026(0.005) | 0.5029(0.003) |
| Power $k$-means | 0.3924(0.001) | 0.3873(0.002) | 0.3722 (0.001) | 0.3967 (0.003) | 0.3871 (0.004) |
| Sparse $k$-means | 0.3679 (0.002) | 0.3677 (0.002) | 0.3668 (0.001) | 0.3675 (0.002) | 0.3637 (0.002) |
| EWP-$k$-means | **0.9641** (0.001) | **0.9217** (0.001) | **0.9139** (0.001) | **0.9465** (0.001) | **0.9082** (0.003) |

Table S2: NMI values for Simulation 2, showing the effect of the number of unimportant features.

| | $k = 20$ | $k = 100$ | $k = 200$ | $k = 500$ |
|---|---|---|---|---|
| $k$-means | 0.0674(0.001) | 0.2502(0.021) | 0.3399 (0.031) | 0.3559 (0.014) |
| $WK$-means | 0.0587(0.001) | 0.2247(0.002) | 0.3584(0.018) | 0.3678(0.009) |
| Power $k$-means | 0.0681(0.001) | 0.2785(0.001) | 0.3578 (0.002) | 0.3867(0.001) |
| Sparse $k$-means | 0.0679(0.001) | 0.2490(0.058) | 0.6705(0.007) | 0.3537 (0.002) |
| EWP-$k$-means | **0.9887**(0.001) | **0.9844** (0.002) | **0.9756**(0.001) | **0.9908** (0.001) |

Table S3: Source and Description of the Datasets

| **Datasets** | **Source** | **k** | **n** | **p** |
|---|---|---|---|---|
| Iris | Keel Repository | 3 | 150 | 4 |
| Automobile | Keel Repository | 6 | 150 | 25 |
| Mammographic | Keel Repository | 2 | 830 | 5 |
| Newthyroid | Keel Repository | 3 | 215 | 5 |
| Wine | Keel Repository | 3 | 178 | 13 |
| WDBC | Keel Repository | 2 | 569 | 30 |
| Movement Libras | Keel Repository | 15 | 360 | 90 |
| Wall Robot 4 | UCI Repository | 4 | 5456 | 4 |
| WarpAR10P | ASU Repository | 10 | 130 | 2400 |
| WarpPIE10P | ASU Repository | 10 | 210 | 2420 |

Table S4: NMI of Real-Life Datasets

| Datasets | $k$-means | Power $k$-means | $WK$-means | Sparse $k$-means | EWP-$k$-means |
|---|---|---|---|---|---|
| Newthyroid | $0.4031^{+}$(0.002) | $0.2625^{+}$(0.002) | $0.4072^{+}$(0.004) | $0.1022^{+}$(0.002) | **0.5321**(0.003) |
| Automobile | $0.1655^{+}$(0.009) | $0.2034^{+}$(0.010) | $0.1687^{+}$(0.005) | $0.1684^{+}$(0.007) | **0.3111**(0.003) |
| WarpAR10P | $0.1708^{+}$(0.042) | $0.2334^{+}$(0.031) | $0.2016^{+}$(0.019) | $0.1853^{+}$(0.008) | **0.3502**(0.047) |
| WarpPIE10P | $0.2406^{\simeq}$(0.031) | $0.2407^{\simeq}$(0.028) | $0.1804^{+}$(0.022) | $0.1799^{+}$(0.002) | **0.2761**(0.041) |
| Iris | $0.7581^{+}$(0.003) | $0.7885^{+}$(0.005) | $0.7419^{+}$(0.005) | $0.8138^{\simeq}$(0.002) | **0.8498**(0.005) |
| Wine | $0.4287^{+}$(0.001) | $0.6427^{+}$(0.005) | $0.4167^{+}$(0.002) | $0.4287^{+}$(0.001) | **0.7476**(0.003) |
| Mammographic | $0.1074^{+}$(0.001) | $0.0194^{+}$(0.003) | $0.1158^{+}$(0.001) | $0.1102^{+}$(0.002) | **0.4051**(0.002) |
| WDBC | $0.4636^{+}$(0.002) | $0.0056^{+}$(0.005) | $0.4648^{+}$(0.002) | $0.4674^{+}$(0.003) | **0.6564**(0.001) |
| LIBRAS | $0.5532^{\simeq}$(0.017) | $0.3390^{+}$(0.020) | $0.4615^{+}$(0.021) | $0.2543^{+}$(0.014) | **0.5751**(0.009) |
| Wall Robot 4 | $0.1677^{+}$(0.027) | $0.1836^{+}$(0.013) | $0.1716^{+}$(0.030) | $0.1861^{+}$(0.012) | **0.2344**(0.003) |

Table S5: ARI values for Simulation 1, showing the effect of the number of unimportant features.

| | $d = 5$ | $d = 10$ | $d = 20$ | $d = 50$ | $d = 100$ |
|---|---|---|---|---|---|
| $k$-means | 0.0120 | 0.0173 | 0.0314 | 0.0114 | 0.0154 |
| $WK$-means | 0.0746 | 0.0846 | 0.0145 | 0.0121 | 0.0012 |
| Power | 0.0125 | 0.0249 | 0.0462 | 0.0164 | 0.0097 |
| Sparse | 0.0245 | 0.0148 | 0.0551 | 0.0137 | 0.0178 |
| EWP | **0.8963** | **0.9016** | **0.8961** | **0.8831** | **0.8615** |

Table S6: ARI values for Simulation 2, showing the effect of increasing number of clusters.

| Algorithm | $k = 20$ | $k = 100$ | $k = 200$ | $k = 500$ |
|---|---|---|---|---|
| $k$-means | 0.0371 | 0.1671 | 0.2486 | 0.2743 |
| $WK$-means | 0.0247 | 0.1293 | 0.2573 | 0.2795 |
| Power $k$-means | 0.0471 | 0.1843 | 0.2462 | 0.2936 |
| Sparse | 0.0148 | 0.1547 | 0.5043 | 0.2847 |
| EWP | **0.9701** | **0.9826** | **0.9612** | **0.9982** |

Table S7: Mean ARI and (standard deviation), GLIOMA data

| $k$-means | $WK$-means | Power | Sparse | EWP |
|---|---|---|---|---|
| 0.281 (0.059) | 0.288 (0.068) | 0.287(0.037) | 0.007(0.006) | **0.448**(0.001) |

Table S8: ARI values on Benchmark Real Data

| Dataset | $k$-means | $WK$-means | Power $k$-means | Sparse $k$-means | EWP $k$-means |
|---|---|---|---|---|---|
| Newthyroid | $0.483^{+}$(0.002) | $0.164^{+}$(0.003) | $0.458^{+}$(0.003) | $0.053^{+}$(0.003) | **0.625**(0.001) |
| Automobile | $0.111^{+}$(0.005) | $0.136^{+}$(0.005) | $0.111^{+}$(0.002) | $0.133^{+}$(0.004) | **0.181**(0.002) |
| WarpAR10P | $0.183^{+}$(0.003) | $0.258^{+}$(0.003) | $0.231^{+}$(0.003) | $0.207^{+}$(0.002) | **0.412**(0.003) |
| WarpPIE10P | $0.253^{\simeq}$(0.005) | $0.267^{\simeq}$(0.003) | $0.214^{+}$(0.002) | $0.205^{+}$(0.004) | **0.323**(0.003) |
| Iris | $0.671^{+}$(0.008) | $0.748^{+}$(0.005) | $0.706^{+}$(0.005) | $0.759^{\simeq}$(0.007) | **0.904**(0.001) |
| Wine | $0.364^{+}$(0.003) | $0.561^{+}$(0.002) | $0.360^{+}$(0.001) | $0.434^{+}$(0.001) | **0.794**(0.002) |
| Mammographic | $0.137^{+}$(0.003) | $0.001^{+}$(0.002) | $0.137^{+}$(0.005) | $0.137^{+}$(0.010) | **0.356**0.004 |
| WDBC | $0.490^{+}$(0.004) | $0.013^{+}$(0.005) | $0.482^{+}$(0.004) | $0.491^{+}$(0.004) | **0.715**(0.005) |
| LIBRAS | $0.302^{\simeq}$(0.004) | $0.112^{+}$(0.006) | $0.481^{+}$(0.001) | $0.589^{+}$(0.003) | **0.592**(0.003) |
| Wall Robot 4 | $0.075^{+}$(0.006) | $0.109^{+}$(0.002) | $0.209^{+}$(0.003) | $0.080^{+}$(0.001) | **0.288**(0.004) |

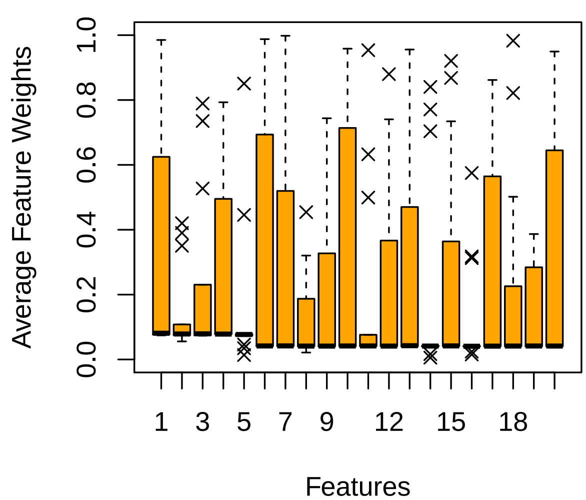


Figure S1: Boxplot shows that $WK$-means fails to identify the correct features.