



Article

LJELSR: A Strengthened Version of JELSR for Feature Selection and Clustering

Sha-Sha Wu ¹, Mi-Xiao Hou ¹, Chun-Mei Feng ^{1,2} and Jin-Xing Liu ^{1,*}

¹ School of Information Science and Engineering, Qufu Normal University, Rizhao 276826, China; wushashaqf@gmail.com (S.-S.W.); mixiaohousz@gmail.com (M.-X.H.); strawberry.feng0304@gmail.com (C.-M.F.)

² Bio-Computing Research Center, Harbin Institute of Technology, Shenzhen 518055, China

* Correspondence: sdcavell@qfnu.edu.cn; Tel.: +86-633-3981-241

Received: 4 December 2018; Accepted: 7 February 2019; Published: 18 February 2019



Abstract: Feature selection and sample clustering play an important role in bioinformatics. Traditional feature selection methods separate sparse regression and embedding learning. Later, to effectively identify the significant features of the genomic data, Joint Embedding Learning and Sparse Regression (JELSR) is proposed. However, since there are many redundancy and noise values in genomic data, the sparseness of this method is far from enough. In this paper, we propose a strengthened version of JELSR by adding the L_1 -norm constraint on the regularization term based on a previous model, and call it LJELSR, to further improve the sparseness of the method. Then, we provide a new iterative algorithm to obtain the convergence solution. The experimental results show that our method achieves a state-of-the-art level both in identifying differentially expressed genes and sample clustering on different genomic data compared to previous methods. Additionally, the selected differentially expressed genes may be of great value in medical research.

Keywords: differentially expressed genes; feature selection; L_1 -norm; sample clustering; sparse constraint

1. Introduction

With the emergence of deep sequencing technologies, considerable genomic data have become available. Since genomic data are usually high-dimension small-sample data, that is, the dimension of the gene is large, the dimension of the sample is small, and it is easy to cause interference when performing feature selection and difficult to understand the sample directly [1]. Additionally, a large number of superfluous and extraneous genes are possessed in these genomic data, which severely interfere with the biological processes. As the case stands, only a small minority of genes with biological sense contribute to disease research [2]. Accordingly, how to identify these key genes from the massive high-dimensional genomic data is a hotspot and nodus in research. Furthermore, studies have testified that these key genes are efficaciously extracted by embedding learning [3]. Furthermore, cluster analysis is based on the similarity of each data point to classify the samples or genes, which is helpful for accurate determination of the cancer subtype. Some studies have also demonstrated that embedding learning and sparse regression is good for cluster analysis and feature selection [4,5].

Feature selection is used to pick out k features from m dimensional data ($m > k$) to optimize a specific index of the system [6]. Initially, some information on characteristic genes is extracted to constitute the data after the reduction to achieve dimensionality reduction. Following this, some genes associated with disease are dredged up from the low-dimensional data on medical research. At present, feature selection is extensively and continually studied owing to the usefulness and practicality of this method. However, traditional feature selection approaches have some issues: (1) The manifold

structure is not fully considered, which can reflect the internal geometric structure in data [7]; and (2) they only use the statistical strategy, which can affect the accuracy and reliability of the results. Moreover, procedures of traditional feature selection method are performed independently, such as embedding learning and sparse regression [8,9]. However, a better performance can be achieved by combining two of the above independent procedures. Hou et al. came up with a new method of feature selection via Joint Embedding Learning and Sparse Regression (JELSR), according to the above ideas [5]. This method is a good solution to the above-mentioned issues. JELSR has a good effect on feature selection outstrip of these traditional methods. Nevertheless, there is still a problem thereafter; since the $L_{2,1}$ -norm penalty only sparsely constrains the rows of data, the sparsity of the method is far from satisfactory. However, if the sparseness is not enough, taking too many unrelated genes into account can cause serious errors. Hence, an efficient sparse method should be explored to strengthen the previous method.

LASSO (Least Absolute Shrinkage and Selection Operator) was first proposed by Robert Tibshirani [10]. It constructs an L_1 -norm penalty function to obtain a more refined model and compresses some coefficients to get zero coefficients. Both the L_1 -norm constraint and the $L_{2,1}$ -norm constraint can produce sparse effects, but the sparsity of the L_1 -norm constraint is decentralized and the $L_{2,1}$ -norm constraint can only produce row sparseness, as shown in Figure 1. However, the combination of the L_1 -norm and the $L_{2,1}$ -norm constraints can generate internal row sparsity and enhance the correlation between rows and columns of the matrix [11], as shown in Figure 1, so the occurrence of the redundant and irrelevant genes can be reduced. In this paper, to obtain more sparse effects, we propose a new method by adding an L_1 -norm constraint on the sparse regression matrix based on the JELSR (LJELSR). First, the corresponding graph is constructed to depict the inherent structure of the data. Then, the graph matrix is embedded into the following steps: feature ranking and data regression. Owing to the similarity between data points described by the above constructed graph, the clustering effect is improved to some degree. Finally, to get more sparse effects, we combine embedding learning and sparse regression with L_1 -norm through linear weighting to complete the corresponding feature selection and cluster analysis.

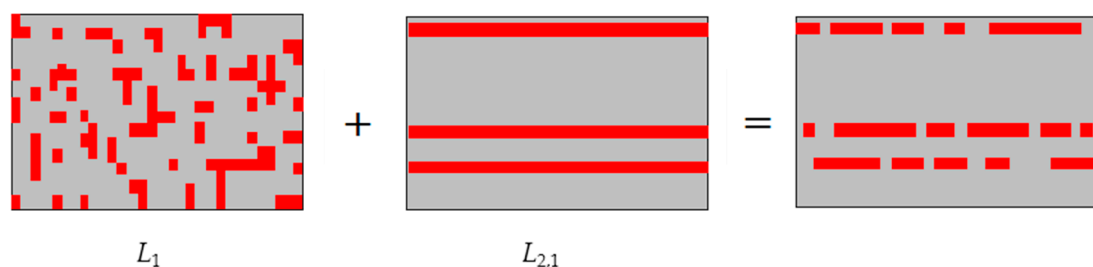


Figure 1. The diagrammatic sketch of different norms.

The major merits of our work are shown below:

1. More zero values are produced by adding an L_1 -norm constraint on the sparse regression matrix, such that we get more sparse results.
2. The internal geometric structure of data is preserved along with dimensionality reduction by embedding learning to reduce the occurrence of inaccurate results.
3. Although an exact solution cannot be obtained by our method, we provide a convergent iterative algorithm to get the optimal results.

The other parts of this paper are arranged as listed below. Section 2 details some comparative experiments and the analysis of experimental results from different datasets. Section 3 describes some related materials and presents the methodology of LJELSR. A conclusion of this paper is given in Section 4.

2. Results

The LJELSR, JELSR [5], ReDac [12], and SMART [11] are used to select differentially expressed genes and test the performance of the proposed method for different genomic data. Among them, the JELSR, ReDac, and SMART are used as comparison methods.

2.1. Datasets

To validate the effectiveness of our method, the LJELSR, JELSR, ReDac, and SMART methods are run on three datasets, including the ALL_AML, the colon cancer, and the ESCA datasets. The ALL_AML dataset includes acute lymphoblastic leukemia (ALL) and acute myelogenous leukemia (AML) [13], and ALL has also been divided into T cell subtypes and B cell subtypes. The colon cancer dataset is obtained by [14], to facilitate clustering, and the samples are organized into two categories, namely, diseased and normal samples. Additionally, the esophageal carcinoma dataset (ESCA) is downloaded from the TCGA (The Cancer Genome Atlas, TCGA). The TCGA is a publicly available dataset, where it is acquired from <https://tcgadata.nci.nih.gov/tcga/>. Some details of the three datasets are listed in Table 1.

Table 1. Details of the three datasets.

Datasets	Genes	Samples	Classes	Description
ALL_AML	5000	38	3	acute lymphoblastic leukemia and acute myelogenous leukemia
colon	2000	62	2	colon cancer
ESCA	20,502	192	2	esophageal carcinoma

2.2. Parameters Selection

In our method, there are mainly four parameters involved, namely three balance parameters α_1 , α_2 , β and the nearest neighbor number q . In ALL_AML, we select parameters α_1 , α_2 and β from range $\{10^5, 10^7, 10^9, 10^{11}, 10^{13}, 10^{15}, 10^{17}, 10^{19}, 10^{21}, 10^{23}, 10^{25}\}$; in colon, α_1 , α_2 and β are found from $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5\}$; and in ESCA, we choose parameters α_1 , α_2 and β in the range of $\{10^0, 10^1, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, 10^9, 10^{10}\}$. The above optimal parameters under different datasets are obtained by five-fold cross-validation. Besides, according to the existing literature [15,16] and a large number of experiments, when q is taken as 5 or 6, the experimental effect is better. In our experiment, we set the value of q to 5.

2.3. Evaluation Metrics

In this study, there are two metrics employed to assess all algorithms: p -value and clustering accuracy (ACC) [17]. Firstly, the size of the p -value is closely related to the relationship between the selected genes and the disease, and there is a negative correlation between them. Additionally, the p -value is obtained by the ToppFun tool, which is a public gene list functional enrichment analysis tool. The p -value cutoff is set as 0.01 in the whole experiment. Secondly, the level of ACC indicates the degree of excellence of the algorithm, and there is a positive correlation between them. The value of the ACC is obtained by the following formula:

$$\text{ACC} = \frac{\sum_{i=1}^n \delta(s_i, \text{map}(c_i))}{n} \quad (1)$$

where c_i means the label of the clustering and s_i is the label of the original data x_i . In addition, the value of the $\delta(x, Y)$ is 1 if $x = y$ and is 0 otherwise, and $\text{map}(\cdot)$ is a mapping function.

2.4. Feature Selection Analysis

2.4.1. Experimental Results and Analysis on ALL_AML Dataset

For the sake of fairness, the LJELSR, JELSR, ReDac, and SMART are respectively used to extract 100 differentially expressed genes from the ALL_AML dataset to analyze their performance. To verify the effectiveness of the algorithm, the selected genes are put into ToppFun to get the p -values and the resulting p -values are arrayed in an ascending sort order. Then, we pick out the first ten terms listed in Table 2. The ten portions in bold font represent the best p -values. It can be seen from Table 2 that the p -values obtained by the LJELSR method are lower than the p -values obtained by the other three methods. Hence, the performance of the LJELSR surpasses the other three methods.

Table 2. The p -values of different methods applied to the ALL_AML dataset.

ID	LJELSR	JELSR	ReDac	SMART
GO:0006955	2.249×10^{-20}	3.157×10^{-18}	8.447×10^{-14}	1.900×10^{-10}
GO:0050776	8.768×10^{-18}	1.312×10^{-15}	7.228×10^{-12}	1.000×10^{-9}
GO:0045321	1.306×10^{-16}	1.808×10^{-14}	1.154×10^{-11}	2.500×10^{-10}
GO:0001775	1.548×10^{-15}	1.604×10^{-13}	6.621×10^{-11}	1.590×10^{-10}
GO:0051251	2.005×10^{-15}	4.157×10^{-14}	9.117×10^{-12}	4.380×10^{-12}
GO:0007159	2.098×10^{-15}	2.754×10^{-15}	3.712×10^{-11}	1.510×10^{-10}
GO:0002682	2.477×10^{-15}	2.725×10^{-14}	7.817×10^{-12}	5.870×10^{-8}
GO:0046649	3.964×10^{-15}	5.426×10^{-14}	3.164×10^{-10}	1.540×10^{-11}
GO:0016337	7.034×10^{-15}	8.993×10^{-14}	5.253×10^{-11}	1.940×10^{-11}
GO:0070486	7.220×10^{-15}	9.350×10^{-15}	1.299×10^{-11}	5.400×10^{-10}

To further illustrate the relationship between the selected genes and ALL_AML, the selected differentially expressed genes are put in GeneCards for testing. GeneCards is a synthesis database of human genes, providing relationships between disease, gene expression, gene function, and so on. The top five differentially expressed genes associated with ALL_AML obtained by the LJELSR method are listed in Table 3 and the official names of these genes and related diseases are also listed. As can be seen from Table 3, the official name of the “CD7” is the “CD7 Molecule”. CD7 is originally found in T cells of acute lymphoblastic leukemia [18]. It is a membrane glycoprotein of human T lymphocytes and thymocytes. Additionally, it plays an essential role in detecting the interactions of T-cells or B-cells during early lymphoid development. Thence, the loss of CD7 can affect the expression of T-cells, which has a great impact on T-cell leukemia [19]. The “MYB” homologous official name is “MYB Proto-Oncogene, Transcription Factor”. From the official name of this gene, MYB is not only a pro-oncogene, but also a factor that affects the transcription of genes. Furthermore, its duplication can cause leukemia [20]. Other genetic analyses are similar to the above analysis. In all, these genes are directly or indirectly related to leukemia. This table only simply displays partial functions of some of these genes.

Table 3. The top five genes selected by LJELSR for the ALL_AML dataset.

Gene	Gene Official Name	Related Diseases
CD34	CD34 Molecule	Dermatofibrosarcoma Protuberans and Hypercalcemic Type Ovarian Small Cell Carcinoma
CD7	CD7 Molecule	Pityriasis Lichenoides Et Varioliformis Acuta and T-Cell Leukemia
MYB	MYB Proto-Oncogene, Transcription Factor	Acute Basophilic Leukemia and Angiocentric Glioma
CXCR4	C-X-C Motif Chemokine Receptor 4	Whim Syndrome and Human Immunodeficiency Virus Infectious Disease
CTSG	Cathepsin G	Papillon-Lefevre Syndrome and Cutaneous Mastocytosis

2.4.2. Experimental Results and Analysis on Colon Cancer Dataset

In this section, the analytical approach and procedure of this dataset are the same as the previous dataset. The 100 genes extracted by each method are tested for the Gene Ontology (GO) detection tool—ToppFun. The p -values gained by four methods are arranged in ascending order. We single out the p -values of the first ten items and list them in Table 4. The best p -values are indicated in bold typeface. From Table 4, it is obvious that the p -values obtained by the LJELSR method are smaller than the p -values obtained by the other three methods. Therefore, the performance of our method is a big plus over the other three methods.

Table 4. The p -values of different methods for the colon dataset.

ID	LJELSR	JELSR	ReDac	SMART
GO:0006614	1.612×10^{-17}	9.836×10^{-14}	2.677×10^{-14}	1.016×10^{-12}
GO:0006613	3.970×10^{-17}	2.060×10^{-13}	5.617×10^{-14}	1.970×10^{-12}
GO:0045047	5.074×10^{-17}	2.519×10^{-13}	6.872×10^{-14}	2.360×10^{-12}
GO:0072599	8.161×10^{-17}	3.720×10^{-13}	1.016×10^{-13}	3.348×10^{-12}
GO:0022626	3.104×10^{-16}	8.999×10^{-13}	2.465×10^{-13}	1.125×10^{-11}
GO:0000184	3.751×10^{-16}	1.301×10^{-12}	3.568×10^{-13}	1.029×10^{-11}
GO:0003735	5.428×10^{-16}	6.787×10^{-13}	1.412×10^{-13}	3.310×10^{-12}
GO:0070972	6.180×10^{-16}	1.960×10^{-12}	5.384×10^{-13}	1.488×10^{-11}
GO:0019083	9.571×10^{-16}	1.932×10^{-12}	1.547×10^{-11}	3.005×10^{-10}
GO:0044445	1.372×10^{-15}	1.487×10^{-12}	3.106×10^{-13}	8.772×10^{-12}

In addition, the selected differentially expressed genes by the LJELSR method are put in GeneCards for testing. We choose the top five genes associated with colon cancer obtained by the LJELSR method to list in Table 5. Additionally, Table 5 also shows how these genes correspond to the official names and related diseases. From Table 5, the official name of the “*ACTB*” is “*Actin Beta*”. This gene can encode one of six different actin proteins, and change in the protein brought about by changes in the gene. It can affect certain biological processes. Furthermore, *ACTB* goes hand in hand with many cancers and plays a major role in lung and colorectal cancer, and so on [21]. Andersen et al. found that colon cancer is affected by *ACTB* [22]. The “*WW Domain Containing Oxidoreductase*” is abbreviated as the “*WWOX*”. Different from the traditional tumor suppressor genes, its effect is more complicated and extensive on cellular function. Besides, the expression level of *WWOX* is different in two different colon cancer cell lines, which are the *HT29* and *SW480* cell lines [23]. Moreover, the expression of *WWOX* can lead to apoptosis, while defects in this gene are associated with multiple types of cancer. Analysis of the remaining genes is similar to the above-mentioned genes. Table 5 only shows some descriptions of partial genes associated with colon cancer.

Table 5. The top five genes selected by LJELSR for the colon dataset.

Gene	Gene Official Name	Related Diseases
<i>MUC3A</i>	<i>Mucin 3A, Cell Surface Associated</i>	Cap Polyposis and Hypertrichotic Osteochondrodysplasia
<i>ACTB</i>	<i>Actin Beta</i>	Dystonia, Juvenile-Onset and Baraitser-Winter Syndrome 1
<i>WWOX</i>	<i>WW Domain Containing Oxidoreductase</i>	Spinocerebellar Ataxia, Autosomal Recessive 12 and Epileptic Encephalopathy, Early Infantile, 28
<i>SPI1</i>	<i>Spi-1 Proto-Oncogene</i>	Inflammatory Diarrhea and Interdigitating Dendritic Cell Sarcoma
<i>RPS24</i>	<i>Ribosomal Protein S24</i>	Diamond-Blackfan Anemia 3 and Diamond-Blackfan Anemia

2.4.3. Experimental Results and Analysis on ESCA Dataset

In this subsection, the dataset we use is ESCA, which is different from the above two datasets. To further confirm the effectiveness of the algorithm, the experiment is run with this data and the selected differentially expressed genes are placed in ToppFun for GO analysis. We rank all the p -values in ascending order and choose first ten p -values to list in the Table 6, where the best results are highlighted. From Table 6, we can conclude that the p -values obtained by the LJELSR method are mostly smaller than the p -values obtained by the other three methods. Therefore, on the whole, the performance of the LJELSR outperforms the other three methods.

Table 6. The p -values of different methods for the ESCA dataset.

ID	LJELSR	JELSR	ReDac	SMART
GO:0005198	2.772×10^{-27}	8.941×10^{-18}	1.096×10^{-18}	1.036×10^{-4}
GO:0070161	7.504×10^{-21}	3.347×10^{-14}	2.733×10^{-23}	6.527×10^{-14}
GO:0030055	1.400×10^{-19}	1.111×10^{-10}	3.576×10^{-17}	1.619×10^{-12}
GO:0005912	6.655×10^{-19}	1.946×10^{-11}	4.401×10^{-20}	3.498×10^{-12}
GO:0005925	1.319×10^{-18}	7.671×10^{-11}	2.103×10^{-17}	1.061×10^{-12}
GO:0005924	1.746×10^{-18}	9.246×10^{-11}	2.747×10^{-17}	1.313×10^{-12}
GO:0005615	5.538×10^{-16}	8.395×10^{-20}	3.985×10^{-15}	2.117×10^{-17}
GO:0030054	5.243×10^{-14}	4.156×10^{-10}	5.243×10^{-14}	4.510×10^{-9}
GO:0005200	3.076×10^{-13}	1.301×10^{-10}	2.062×10^{-13}	2.358×10^{-4}
GO:0042060	7.442×10^{-13}	2.650×10^{-9}	4.536×10^{-12}	8.790×10^{-11}

Additionally, the selected differentially expressed genes are put in GeneCards for testing. The top five genes associated with ESCA obtained by the LJELSR method, their name, and related diseases are all displayed in Table 7. From Table 7, the “ERBB2” corresponds to the official name “Erb-B2 Receptor Tyrosine Kinase 2”. ERBB2 is the membrane receptor of 185kDa encoded by proto-oncogene ERBB-2, and is one of the members of the epidermal growth factor receptor family. Additionally, it has been verified that its amplification is closely related to the occurrence of esophageal cancer [24]. “KRT5” is the abbreviation of “Keratin 5”, and it is a member of the keratin family of momentous gene families that encodes the corresponding protein. Additionally, its changes affect the expression of this gene’s families, causing some complex diseases. For example, mutations in KRT5 and KRT14 can cause epidermolysis bullosa to a large extent [25]. The analysis of other genes is similar to the above analysis. Table 7 only shows some functions of partial genes, and detailed information on the remaining genes can be obtained from GeneCards.

Table 7. The top five genes selected by LJELSR for the ESCA dataset.

Gene	Gene Official Name	Related Diseases
ERBB2	Erb-B2 Receptor Tyrosine Kinase 2	Glioma Susceptibility 1 and Ovarian Cancer, Somatic
KRT14	Keratin 14	Epidermolysis Bullosa Simplex, Koebner Type and Epidermolysis Bullosa Simplex, Recessive 1
KRT5	Keratin 5	Epidermolysis Bullosa Simplex, Dowling-Meara Type and Epidermolysis Bullosa Simplex, Weber-Cockayne Type
KRT19	Keratin 19	Anal Canal Adenocarcinoma and Thyroid Cancer
KRT4	Keratin 4	White Sponge Nevus 1 and White Sponge Nevus Of Cannon, Krt4-Related

2.4.4. Differentially Expressed Genes Comparing by Methods

In this subsection, the selected differentially expressed genes are further analyzed. For the three datasets mentioned above, we explore the common differentially expressed genes and the unique differentially expressed genes obtained by different methods for the same dataset. We select 100 genes for each algorithm and pair them with the officially published disease-causing gene pool to obtain the

verified genes. This disease-causing gene pool can be downloaded directly from GeneCards. Table 8 shows the common and unique differentially expressed genes obtained by the four methods for the ESCA dataset. The bold italic indicates the common differentially expressed genes excavated by the four methods. The underlined italic indicates the unique differentially expressed genes excavated by LJELSR, which are not found by other methods. The “Number” means the total number of verified genes. As can be seen from the table, LJELSR has selected more proven genes than other methods; *ANXA1*, *MUC6*, *FN1*, *PKM*, *CD24*, *GLUL*, *PLEC*, *PIGR*, *ACTB*, *PABPC1*, *LYZ*, and *SPRR1B* are the common differentially expressed genes; the unique differentially expressed genes of the LJELSR method are *FSCN1*, *ITGB4*, *LAMC2*, *HLA-B*, *LAMB3*, *HLA-C*, *SLC7A5*, *ENO1*, and so on. These genes play an important role in studying the relationship between genes and diseases, the details of these genes can be obtained from GeneCards, and some of the genes have been explained above. In summary, LJELSR has an advantage over other methods in that it can mine more and more valuable differentially expressed genes.

Table 8. Differentially expressed genes of four methods for the ESCA dataset.

Methods	Differentially Expressed Genes	Number
LJELSR	<i>ERBB2</i> , <i>KRT14</i> , <i>KRT5</i> , <i>KRT19</i> , <i>KRT4</i> , <i>TFF1</i> , <i>FSCN1</i> , <i>KRT13</i> , <i>ITGB4</i> , <i>ANXA1</i> , <i>MUC6</i> , <i>LAMC2</i> , <i>HLA-B</i> , <i>KRT16</i> , <i>JUP</i> , <i>KRT17</i> , <i>LAMB3</i> , <i>ATP4A</i> , <i>DSP</i> , <i>LAMA3</i> , <i>FOS</i> , <i>FN1</i> , <i>CTSB</i> , <i>MYH11</i> , <i>HLA-C</i> , <i>LDHA</i> , <i>PKM</i> , <i>SLC7A5</i> , <i>PSCA</i> , <i>SERPINA1</i> , <i>S100A7</i> , <i>S100A9</i> , <i>CRNN</i> , <i>S100A8</i> , <i>B2M</i> , <i>DMBT1</i> , <i>CD24</i> , <i>ENO1</i> , <i>TNC</i> , <i>KRT15</i> , <i>GLUL</i> , <i>HSPA1A</i> , <i>NDRG1</i> , <i>LCN2</i> , <i>COL17A1</i> , <i>CEACAM6</i> , <i>REG1A</i> , <i>PLEC</i> , <i>GAPDH</i> , <i>PIGR</i> , <i>AGR2</i> , <i>ANPEP</i> , <i>PKP1</i> , <i>ACTB</i> , <i>FLNA</i> , <i>PI3</i> , <i>FTL</i> , <i>CTSE</i> , <i>PABPC1</i> , <i>PGC</i> , <i>ALDOA</i> , <i>EEF2</i> , <i>KRT6B</i> , <i>LYZ</i> , <i>CLDN18</i> , <i>SPRR1B</i> , <i>KRT6C</i> , <i>PPP1R1B</i> , <i>PGA3</i> , <i>COL3A1</i> , <i>C3</i> , <i>REG1B</i> , <i>PERP</i> , <i>KRT6A</i> , <i>PGA5</i> , <i>CES1</i> , <i>PGA4</i> , <i>EIF1</i>	78
JELSR	<i>MUC1</i> , <i>KRT14</i> , <i>KRT5</i> , <i>KRT19</i> , <i>KRT8</i> , <i>KRT4</i> , <i>TFF1</i> , <i>FN1</i> , <i>MUC6</i> , <i>S100A8</i> , <i>CTSD</i> , <i>SPRR3</i> , <i>ATP4A</i> , <i>HSPB1</i> , <i>FOS</i> , <i>CEACAM5</i> , <i>GJB2</i> , <i>H19</i> , <i>EZR</i> , <i>KRT16</i> , <i>KRT13</i> , <i>CTSB</i> , <i>JUP</i> , <i>ANXA1</i> , <i>TFF2</i> , <i>S100A9</i> , <i>SNF8</i> , <i>KRT17</i> , <i>PSCA</i> , <i>S100A2</i> , <i>MUC5B</i> , <i>COL1A1</i> , <i>CD24</i> , <i>PKM</i> , <i>DSC3</i> , <i>GLUL</i> , <i>MALAT1</i> , <i>REG1A</i> , <i>ACTN4</i> , <i>DSG3</i> , <i>LCN2</i> , <i>DSP</i> , <i>S100A7</i> , <i>B2M</i> , <i>MYH9</i> , <i>PKP1</i> , <i>S100A11</i> , <i>PIGR</i> , <i>HSPA8</i> , <i>PLEC</i> , <i>TRIM29</i> , <i>EEF1A1</i> , <i>ATP1B1</i> , <i>AGR2</i> , <i>LYZ</i> , <i>ACTB</i> , <i>PGC</i> , <i>PABPC1</i> , <i>SPRR2A</i> , <i>SPRR1B</i> , <i>SPRR1A</i> , <i>CA2</i> , <i>REG4</i> , <i>P4HB</i> , <i>CLDN18</i> , <i>CTSE</i> , <i>EEF2</i> , <i>CREB3L1</i> , <i>KRT6A</i> , <i>A2M</i> , <i>PGA3</i>	71
ReDac	<i>KRT14</i> , <i>KRT5</i> , <i>KRT4</i> , <i>FN1</i> , <i>CRNN</i> , <i>TGM3</i> , <i>MUC6</i> , <i>S100A8</i> , <i>IL1RN</i> , <i>SPRR3</i> , <i>ATP4A</i> , <i>HSPB1</i> , <i>MAL</i> , <i>KRT16</i> , <i>KRT13</i> , <i>JUP</i> , <i>THBS1</i> , <i>ANXA1</i> , <i>S100A9</i> , <i>PSCA</i> , <i>ECM1</i> , <i>CD24</i> , <i>PKM</i> , <i>FTL</i> , <i>HSPG2</i> , <i>DES</i> , <i>GLUL</i> , <i>MALAT1</i> , <i>PPL</i> , <i>EMPI1</i> , <i>ACTN4</i> , <i>MYH11</i> , <i>CSTA</i> , <i>GAPDH</i> , <i>TAGLN</i> , <i>DSP</i> , <i>B2M</i> , <i>MYH9</i> , <i>GSN</i> , <i>PKP1</i> , <i>S100A11</i> , <i>PIGR</i> , <i>HSPA8</i> , <i>FLNA</i> , <i>PLEC</i> , <i>MYLK</i> , <i>CSTB</i> , <i>TRIM29</i> , <i>EEF1A1</i> , <i>RPL3</i> , <i>LYZ</i> , <i>PSAP</i> , <i>ACTB</i> , <i>ALDOA</i> , <i>PGC</i> , <i>PABPC1</i> , <i>SYNM</i> , <i>SPRR2A</i> , <i>SPRR1B</i> , <i>SPRR1A</i> , <i>REG4</i> , <i>P4HB</i> , <i>EEF2</i> , <i>KRT6A</i> , <i>ACTG2</i> , <i>PGA3</i>	66
SMART	<i>ERBB2</i> , <i>CCND1</i> , <i>GSTP1</i> , <i>CD44</i> , <i>KRT19</i> , <i>MUC4</i> , <i>MUC2</i> , <i>GRB7</i> , <i>TFF1</i> , <i>HLA-A</i> , <i>FN1</i> , <i>SOD2</i> , <i>ITGA6</i> , <i>NDRG1</i> , <i>SPP1</i> , <i>SERPINA1</i> , <i>MUC6</i> , <i>CTSD</i> , <i>HSPB1</i> , <i>CEACAM5</i> , <i>H19</i> , <i>CTSB</i> , <i>F5</i> , <i>ITGB1</i> , <i>ANXA1</i> , <i>SDC1</i> , <i>DMBT1</i> , <i>CLU</i> , <i>LDHA</i> , <i>CD24</i> , <i>APP</i> , <i>PKM</i> , <i>FTL</i> , <i>HSPG2</i> , <i>TNC</i> , <i>GLUL</i> , <i>MALAT1</i> , <i>NTS</i> , <i>LCN2</i> , <i>MYH9</i> , <i>PIGR</i> , <i>FLNA</i> , <i>CD55</i> , <i>PLEC</i> , <i>TSPAN8</i> , <i>EEF1A1</i> , <i>AGR2</i> , <i>LYZ</i> , <i>GPX2</i> , <i>ACTB</i> , <i>DSG2</i> , <i>PABPC1</i> , <i>SPRR1B</i> , <i>REG4</i> , <i>SCD</i> , <i>CLDN18</i> , <i>FAT1</i>	57

2.5. Clustering Analysis

Kmeans is one of simple algorithms that solve the clustering problem, and its procedure is a relatively easy and simple way to classify the samples by setting a certain number of clusters in advance [26]. For all methods, different datasets have different numbers of sample clustering. The Kmeans method is used to cluster samples on the ALL_AML, colon cancer, and ESCA datasets, where their number of sample clustering is three, two, and two, respectively. Different methods are run for different data to get the values of ACC, and the details are shown in Table 8, where the largest values are marked in bold typeface. From Table 9, we can conclude that the values of ACC obtained

by the LJELSR method for different datasets are almost all higher than the other methods. In this section, based on the above analysis, we can summarize that the efficiency of our method is higher than other methods.

Table 9. ACC of different methods for different datasets.

Methods	ALL_AML	Colon	ESCA
LJELSR	81.579	64.520	96.354
JELSR	81.579	61.290	95.833
ReDac	68.421	61.290	95.313
SMART	44.740	63.980	94.790
Kmeans	78.530	53.420	96.350

3. Materials and Methods

3.1. Related Notations and Definitions

In this paper, m^i and m_j are defined as the i -th row and j -th column of matrix M , respectively. Denote $X \in \mathbb{R}^{m \times n}$ as the original data matrix, where its row represents a gene (feature) and its column represents a sample. For an arbitrary matrix M , $L_{r,s}$ norm is represented as below:

$$\|M\|_{r,s} = \left(\sum_{i=1}^n \left(\sum_{j=1}^m |m_{ij}|^r \right)^{\frac{s}{r}} \right)^{\frac{1}{s}} \tag{2}$$

where r and s represent a positive number. When $r = s = 1$, the above formula becomes the expression of the L_1 norm or LASSO, i.e.,

$$\|M\|_1 = \sum_{i=1}^n \sum_{j=1}^m |m_{ij}| \tag{3}$$

Similarly, when $r = 2, s = 1$, and (2) becomes the expression of the $L_{2,1}$ norm, i.e.,

$$\|M\|_{2,1} = \sum_{i=1}^n \left(\sum_{j=1}^m |m_{ij}|^2 \right)^{\frac{1}{2}} \tag{4}$$

Additionally, the inherent geometric structure of data is fully taken into account to deal with data for lessening the appearance of the inexact results in actual applications. Considering its benefits, it is also added to our work. Therefore, we firstly construct a q -nearest-neighbor graph G with n vertices in the data space, where a vertex corresponds to a data point [15], and q is the nearest neighbor number. w_{ij} represents the correlation between two data point x_j and x_j . All w_{ij} make up the weight matrix W . There are many strategies to compute W . However, three strategies are frequently employed [15].

(1) 0-1 weighting:

$$w_{ij} = \begin{cases} 1, & \text{if } x_i \in N_q(x_j) \text{ or } x_j \in N_q(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $N_q(x_i)$ indicates the set gained by the way of q -nearest neighbors of the data point x_i .

(2) Heat kernel weighting:

$$w_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{\sigma}}, & \text{if } x_i \in N_q(x_j) \text{ or } x_j \in N_q(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where σ is a proper constant, and its value is obtained by the previous experience.

(3) Dot-product weighting:

$$w_{ij} = \begin{cases} x_i^T x_j, & \text{if } x_i \in N_q(x_j) \text{ or } x_j \in N_q(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

These three strategies are applied to different occasions. Since the operation of the 0-1 weighting is relatively uncomplicated, it is commonly used for computing the weight matrix. The second is widely applied to image data, and the third is frequently used in the IR community for processing documents [15].

Next, we define a diagonal matrix D , where its diagonal values are given as $d_{ii} = \sum_j w_{ij}$. The graph Laplacian matrix L is defined as $L = D - W$ [27].

3.2. Joint Embedding Learning and Sparse Regression (JELSR)

Traditional feature selection methods are performed independently. To further heighten the performance of the previous algorithm, the JELSR method was proposed by Hou et al. [5]. Firstly, it uses linear approximation weights and the $L_{2,1}$ -norm regularization to combine embedding learning and sparse regression to establish a new objective function. Then, the sparse regression matrix is used to finish the corresponding feature selection [5]. The objective function of the JELSR algorithm is written as follows:

$$\begin{aligned} \min_{P,Y} \left(YLY^T \right) + \beta \left(\|P^T X - Y\|_2^2 + \alpha \|P\|_{2,1} \right) \\ \text{s.t. } YY^T = I_{k \times k} \end{aligned} \quad (8)$$

where $Y \in \mathbb{R}^{k \times n}$ is a low dimensional embedding matrix; $P \in \mathbb{R}^{m \times k}$ is the sparse regression matrix; α and β are two balance parameters; and k is the dimension of low-dimensional space.

3.3. The Proposed Method

Assuming that the expression level of the gene is within the normal range, the higher the sparseness of the sparse regression matrix, the easier it is to find the differentially expressed genes. What is more, the results of identifying the differentially expressed genes are much more accurate.

Compared with the previous methods, the JELSR method proposed by Hou et al. [5] has a good effect on the feature selection. However, it is inevitable that some redundancy values and artificial noise values in the sparse regression matrix have to be taken into account, such that the sparseness of this method is far from satisfactory. Consequently, it is necessary to discover an efficacious sparse method to improve the performance. Therefore, we propose the LJELSR method, which may improve the sparseness of the algorithm and the accuracy of the results.

The objective function of LJELSR is:

$$\begin{aligned} \min_{P,Y} \left(YLY^T \right) + \beta \left(\|P^T X - Y\|_2^2 + \alpha_1 \|P\|_1 + \alpha_2 \|P\|_{2,1} \right) \\ \text{s.t. } YY^T = I_{k \times k} \end{aligned} \quad (9)$$

where $Y \in \mathbb{R}^{k \times n}$ is a low-dimensional embedding matrix; $P \in \mathbb{R}^{m \times k}$ is the sparse regression matrix; and α_1 , α_2 , and β are three balance parameters.

3.4. Optimization

Since there are the L_1 -norm and the $L_{2,1}$ -norm constraints on the objective formula, it is difficult to optimize and solve the optimal solution directly. With the elicitation of Wang et al. [11] and Nie et al. [28], the iterative strategy is introduced to solve the above problem. Now, we will explain in detail the specific optimization process of our method.

Before solving the optimization problem, we introduce two diagonal matrices, $U \in \mathbb{R}^{m \times m}$ and $\tilde{U} \in \mathbb{R}^{m \times m}$, whose i -th diagonal values are defined as follows:

$$U = \sum_{i=1}^k U_i, (U_i)_{jj} = \frac{1}{2|p_{ji}|} (j = 1 \dots m) \tag{10}$$

$$\tilde{U}_{jj} = \frac{1}{2\|p^j\|_2} \tag{11}$$

where U_i represents i -th diagonal matrix U ; $(U_i)_{jj}$ represents the j -th diagonal value of the i -th matrix U ; and \tilde{U}_{jj} indicates the j -th diagonal value of matrix \tilde{U} .

To prevent the emergence of spillover, a small constant ϵ is added to the diagonal matrices U and \tilde{U} , respectively, that is:

$$U = \sum_{i=1}^k U_i, (U_i)_{jj} = \frac{1}{2\max(|p_{ji}|, \epsilon)} (j = 1 \dots m) \tag{12}$$

$$\tilde{U}_{jj} = \frac{1}{2\max(\|p^j\|_2, \epsilon)} \tag{13}$$

Owing to the fact that the partial derivatives of $\|P\|_1$ and $\text{tr}(P^TUP)$ on P are identical, $\|P\|_1$ can be replaced by $\text{tr}(P^TUP)$. Analogously, we carry out an exchange for $\|P\|_{2,1}$ and $\text{tr}(P^T\tilde{U}P)$. Therefore, (9) can be rewritten as:

$$\begin{aligned} \min_{P,U,\tilde{U},Y} \text{tr}(YLY^T) + \beta(\|P^TX - Y\|_2^2 + \alpha_1\text{tr}(P^TUP) + \alpha_2\text{tr}(P^T\tilde{U}P)) \\ \text{s.t. } YY^T = I_{k \times k} \end{aligned} \tag{14}$$

We firstly optimize the matrix P . We denote

$$L(P, U, \tilde{U}) = \|P^TX - Y\|_2^2 + \alpha_1\text{tr}(P^TUP) + \alpha_2\text{tr}(P^T\tilde{U}P) \tag{15}$$

When two diagonal matrices U and \tilde{U} are fixed, we compute the partial derivative of $L(P, U, \tilde{U})$ on P and make it equal to zero. Therefore, we can get the following equation:

$$\frac{\partial L(P)}{\partial P} = 2XX^TP - 2XY^T + 2\alpha_1UP + 2\alpha_2\tilde{U}P = 0 \tag{16}$$

namely,

$$P = (XX^T + \alpha_1U + \alpha_2\tilde{U})^{-1}XY^T \tag{17}$$

To facilitate optimization, we introduce an auxiliary variable A into the objective formula, and denote $A = XX^T + \alpha_1U + \alpha_2\tilde{U}$. According to the above analysis, (17) is brought into (14), such that we will get

$$L(P, U, \tilde{U}, Y) = \text{tr}(Y(L + \beta I_{n \times n} - \beta X^T A^{-1} X)Y^T) \tag{18}$$

Since Y is subject to the orthogonal constraint $YY^T = I_{k \times k}$, the optimization problem of Y becomes

$$\begin{aligned} \arg \min_Y \text{tr}(Y(L + \beta I_{n \times n} - \beta X^T A^{-1} X)Y^T) \\ \text{s.t. } YY^T = I_{k \times k} \end{aligned} \tag{19}$$

When A and L are fixed, we use the strategy of the eigen-decomposition of the matrix $G = (L + \beta I_{n \times n} - \beta X^T A^{-1} X)Y^T$ to update Y in (19). For $Y_i (i = 1, 2, \dots, k)$, we firstly choose the k smallest

eigenvalues of the matrix G , and then seek out the corresponding eigenvectors to constitute a new matrix $Y = (Y_1, Y_2, \dots, Y_k)$ [29]. In addition, the diagonal matrices U and \tilde{U} are updated by (12) and (13) when P is fixed, respectively. Additionally, we initialize U and \tilde{U} as an identity matrix, respectively.

3.5. Feature Selection

According to the above update rules, the sparse regression matrix P is acquired after repeated iterations of LJELSR. Then, to acquire some differentially expressed genes, we conduct a detailed analysis of the matrix P . In the first place, each element of the matrix P is subjected to absolute processing. Secondly, we sum the absolute values by each row of the sparse regression matrix P and get a new vector, as follows.

$$\bar{P} = (\bar{P}_1, \bar{P}_2, \dots, \bar{P}_m)^T \quad (20)$$

$$\bar{P}_i = \sum_{j=1}^k |P_{ij}| \quad (21)$$

Then, we rank the vector \bar{P}_i in descending order to get a new vector, as follows:

$$\tilde{P} = (\tilde{P}_1, \tilde{P}_2, \dots, \tilde{P}_m)^T \quad (22)$$

Finally, the genes corresponding to the first l values are selected as differentially expressed genes to analyze their property ($l < m$). By and large, the value of the element is directly proportional to the importance of the corresponding gene. Next, we put the selected genes into TopFun and GeneCards to analyze them, where they are publicly accessible from <https://toppgene.cchmc.org/enrichment.jsp> and <http://www.genecards.org/>, respectively.

In summary, the procedure of LJELSR is shown in Algorithm 1.

Algorithm 1. Procedure of LJELSR.

Input: Data matrix X ; Neighborhood size q ; Balance parameters $\alpha_1, \alpha_2, \beta$; Dimensionality of embedding k ; Feature selection number l .

Output: Selected feature index set $\{P_1, P_2, \dots, P_m\}$

Stage one: Graph construction

Construct the weight matrix W ; Compute the diagonal matrix D , graph Laplacian matrix L ;

Stage two: Alternative optimization,

Initialize $U = \tilde{U} = I_{m \times m}$;

Loop

Update Y and fix A, L by (19),

Update P and fix U, \tilde{U} by (17),

Update U, \tilde{U} and fix P by (12) and (13).

until convergence

Stage three: Feature selection

3.6. Convergence Analysis

In this study, an alternative algorithm is utilized to finish iteratively updating work of the proposed method. Now, let us analyze its convergence behavior of LJELSR. The lemma given below was proposed by Nie et al. [28].

Lemma 1. For any non-zero vectors $P, P_t \in R^k$, the following inequality holds:

$$\|P\|_2 - \frac{\|P\|_2^2}{2\|P_t\|_2} \leq \|P_t\|_2 - \frac{\|P_t\|_2^2}{2\|P_t\|_2} \quad (23)$$

The convergence result of the proposed method is explained by the following theorem:

Theorem 1. *The value of the target function for each iteration is monotonically decreasing in the Algorithm 1. Detailed proof of Theorem 1 is given in the Appendix A.*

4. Conclusions

In this paper, we discuss a new feature selection method by adding an L_1 -norm constraint on the sparse regression matrix based on the JESLR method. Firstly, the four methods are executed to select differentially expressed genes and cluster the samples from ALL_AML, colon cancer, and ESCA datasets, respectively. Secondly, some of materials related to this paper are presented, and our methods and the corresponding optimization strategy are given. Finally, the conclusion is drawn that the performance of the proposed method is better than other methods through the experimental results.

Author Contributions: The study was jointly designed by S.-S.W. and J.-X.L. Both the experimental execution of the LJESLR method and the writing of the manuscript were completed by S.-S.W. The data analysis was contributed by M.-X.H and C.-M.F. The final manuscript was read and approved by all authors.

Acknowledgments: This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61872220 and 61572284.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof. We can come to the following conclusion in accordance with the objective function of the proposed method.

$$\begin{aligned} L(P^{t+1}, Y^{t+1}) &= \text{tr}(Y^t L(Y^t)^T) + \beta (\| (P^t)^T X - Y^t \|_2^2 + \alpha_1 \| P^t \|_1 + \alpha_2 \| P^t \|_{2,1}) \\ &= \text{tr}(Y^t L(Y^t)^T) + \beta \left(\text{tr} \left((P^t)^T X - Y^t \right)^T \left((P^t)^T X - Y^t \right) \right) + \beta \left(\alpha_1 \sum_{i=1}^k (p_i^t)^T U_i^t p_i^t + \alpha_2 \text{tr} \left((P^t)^T \right) \tilde{U}^t (P^t) \right) \end{aligned}$$

Consequently, we obtain

$$\begin{aligned} &\text{tr}(Y^{t+1} L(Y^{t+1})^T) + \beta \left(\text{tr} \left((P^{t+1})^T X - Y^{t+1} \right)^T \left((P^{t+1})^T X - Y^{t+1} \right) \right) + \beta \left(\alpha_1 \sum_{i=1}^k (p_i^{t+1})^T U_i^t p_i^{t+1} + \alpha_2 \text{tr} \left((P^{t+1})^T \right) \tilde{U}^t (P^{t+1}) \right) \\ &\leq \text{tr}(Y^t L(Y^t)^T) + \beta \left(\text{tr} \left((P^t)^T X - Y^t \right)^T \left((P^t)^T X - Y^t \right) \right) + \beta \left(\alpha_1 \sum_{i=1}^k (p_i^t)^T U_i^t p_i^t + \alpha_2 \text{tr} \left((P^t)^T \right) \tilde{U}^t (P^t) \right) \\ &\rightarrow \\ &\text{tr}(Y^{t+1} L(Y^{t+1})^T) + \beta \left(\text{tr} \left((P^{t+1})^T X - Y^{t+1} \right)^T \left((P^{t+1})^T X - Y^{t+1} \right) \right) \\ &+ \alpha_1 \beta \sum_{i=1}^m \sum_{j=1}^k \left(\frac{(p_{ij}^{t+1})^2}{2 \| p_{ij}^{t+1} \|} - \| p_{ij}^{t+1} \| + \| p_{ij}^{t+1} \| \right) + \alpha_2 \beta \sum_{c=1}^m \left(\frac{\| (P^{t+1})^c \|_2^2}{2 \| (P^{t+1})^c \|_2} - \| (P^{t+1})^c \|_2 + \| (P^{t+1})^c \|_2 \right) \\ &\leq \text{tr}(Y^t L(Y^t)^T) + \beta \left(\text{tr} \left((P^t)^T X - Y^t \right)^T \left((P^t)^T X - Y^t \right) \right) \\ &+ \alpha_1 \beta \sum_{i=1}^m \sum_{j=1}^k \left(\| p_{ij}^t \| + \frac{(p_{ij}^t)^2}{2 \| p_{ij}^t \|} - \| p_{ij}^t \| \right) + \alpha_2 \beta \sum_{c=1}^m \left(\| (P^t)^c \|_2 + \frac{\| (P^t)^c \|_2^2}{2 \| (P^t)^c \|_2} - \| (P^t)^c \|_2 \right) \\ &\rightarrow \\ &\text{tr}(Y^{t+1} L(Y^{t+1})^T) + \beta \left(\text{tr} \left((P^{t+1})^T X - Y^{t+1} \right)^T \left((P^{t+1})^T X - Y^{t+1} \right) \right) + \beta \left(\alpha_1 \sum_{i=1}^m \sum_{j=1}^k (\| p_{ij}^{t+1} \|) + \alpha_2 \sum_{c=1}^m (\| (P^{t+1})^c \|_2) \right) \\ &\leq \text{tr}(Y^t L(Y^t)^T) + \beta \left(\text{tr} \left((P^t)^T X - Y^t \right)^T \left((P^t)^T X - Y^t \right) \right) + \beta \left(\alpha_1 \sum_{i=1}^m \sum_{j=1}^k (\| p_{ij}^t \|) + \alpha_2 \sum_{c=1}^m (\| (P^t)^c \|_2) \right) \end{aligned}$$

According to Lemma 1, the last step is established. Thus, this algorithm monotonically decreases the objective value in each iteration. \square

References

1. Church, G.M.; Gilbert, W. Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 1991–1995. [[CrossRef](#)] [[PubMed](#)]
2. Liao, J.C.; Boscolo, R.; Yang, Y.-L.; Tran, L.M.; Sabatti, C.; Roychowdhury, V.P. Network component analysis: Reconstruction of regulatory signals in biological systems. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15522–15527. [[CrossRef](#)] [[PubMed](#)]
3. Constantinopoulos, C.; Titsias, M.K.; Likas, A. Bayesian feature and model selection for gaussian mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1013–1018. [[CrossRef](#)] [[PubMed](#)]
4. Nie, F.; Zeng, Z.; Tsang, I.W.; Xu, D.; Zhang, C. Spectral embedded clustering: A framework for in-sample and out-of-sample spectral clustering. *IEEE Trans. Neural Netw.* **2011**, *22*, 1796–1808. [[PubMed](#)]
5. Hou, C.; Nie, F.; Yi, D.; Wu, Y. Feature selection via joint embedding learning and sparse regression. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2011), Barcelona, Spain, 16–22 July 2011; pp. 1324–1329.
6. D’Addabbo, A.; Papale, M.; Di Paolo, S.; Magaldi, S.; Colella, R.; d’Onofrio, V.; Di Palma, A.; Ranieri, E.; Gesualdo, L.; Ancona, N. Svd based feature selection and sample classification of proteomic data. In *Knowledge-Based Intelligent Information and Engineering Systems*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 556–563.
7. Cai, D.; He, X.; Han, J. Spectral regression for efficient regularized subspace learning. *Proceedings* **2007**, *149*, 1–8.
8. Zhao, Z.; Wang, L.; Liu, H. Efficient spectral feature selection with minimum redundancy. In Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10), Atlanta, GA, USA, 11–15 July 2010; pp. 673–678.
9. Cai, D.; Zhang, C.; He, X. Unsupervised feature selection for multi-cluster data. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 25–28 July 2010; pp. 333–342.
10. Tibshirani, R.J. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* **1996**, *58*, 267–288. [[CrossRef](#)]
11. Wang, H.; Nie, F.; Huang, H.; Risacher, S.; Ding, C.; Saykin, A.J.; Shen, L. Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 557–562.
12. Zhao, Q.; Meng, D.; Xu, Z. A recursive divide-and-conquer approach for sparse principal component analysis. *arXiv*, 2012; arXiv:1211.7219.
13. Brunet, J.P.; Tamayo, P.; Golub, T.R.; Mesirov, J.P. Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4164–4169. [[CrossRef](#)] [[PubMed](#)]
14. Alon, U.; Barkai, N.; Notterman, D.A.; Gish, K.; Ybarra, S.; Mack, D.; Levine, A.J. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6745–6750. [[CrossRef](#)] [[PubMed](#)]
15. Cai, D.; He, X.; Han, J.; Huang, T.S. Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1548–1560. [[PubMed](#)]
16. Shang, R.; Wang, W.; Stolkin, R.; Jiao, L. Non-negative spectral learning and sparse regression-based dual-graph regularized feature selection. *IEEE Trans. Cybern.* **2017**, *1*–14. [[CrossRef](#)] [[PubMed](#)]
17. Wu, M.; Schölkopf, B. A local learning approach for clustering. In Proceedings of the International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 4–7 December 2006; pp. 1529–1536.
18. Aruffo, A.; Seed, B. Molecular cloning of two cd7 (T-cell leukemia antigen) cdnas by a cos cell expression system. *EMBO J.* **1987**, *6*, 3313–3316. [[CrossRef](#)] [[PubMed](#)]
19. Liu, T.Y.; Chen, C.Y.; Tien, H.F.; Lin, C.W. Loss of cd7, independent of galectin-3 expression, implies a worse prognosis in adult T-cell leukaemia/lymphoma. *Histopathology* **2009**, *54*, 214–220. [[CrossRef](#)] [[PubMed](#)]
20. Lahortiga, I.; De Keersmaecker, K.; Van Vlierberghe, P.; Graux, C.; Cauwelier, B.; Lambert, F.; Mentens, N.; Beverloo, H.B.; Pieters, R.; Speleman, F. Duplication of the myb oncogene in t cell acute lymphoblastic leukemia. *Nat. Genet.* **2007**, *39*, 593–595. [[CrossRef](#)] [[PubMed](#)]
21. Guo, C.; Liu, S.; Wang, J.; Sun, M.-Z.; Greenaway, F.T. Actb in cancer. *Clin. Chim. Acta* **2013**, *417*, 39–44. [[CrossRef](#)] [[PubMed](#)]

22. Andersen, C.L.; Jensen, J.L.; Ørntoft, T.F. Normalization of real-time quantitative reverse transcription-pcr data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **2004**, *64*, 5245–5250. [[CrossRef](#)] [[PubMed](#)]
23. Nowakowska, M.; Pospiech, K.; Lewandowska, U.; Piastowska-Ciesielska, A.W.; Bednarek, A.K. Diverse effect of wwox overexpression in ht29 and sw480 colon cancer cell lines. *Tumor Biol.* **2014**, *35*, 9291–9301. [[CrossRef](#)] [[PubMed](#)]
24. Dahlberg, P.S.; Jacobson, B.A.; Dahal, G.; Fink, J.M.; Kratzke, R.A.; Maddaus, M.A.; Ferrin, L.J. Erbb2 amplifications in esophageal adenocarcinoma. *Ann. Thorac. Surg.* **2004**, *78*, 1790–1800. [[CrossRef](#)] [[PubMed](#)]
25. Bolling, M.; Lemmink, H.; Jansen, G.; Jonkman, M. Mutations in krt5 and krt14 cause epidermolysis bullosa simplex in 75% of the patients. *Br. J. Dermatol.* **2011**, *164*, 637–644. [[CrossRef](#)] [[PubMed](#)]
26. Xu, J.; Liu, H. Web user clustering analysis based on kmeans algorithm. In Proceedings of the International Conference on Information NETWORKING and Automation, Kunming, China, 18–19 October 2010.
27. Zhou, D.; Huang, J.; Schölkopf, B. Learning with hypergraphs: Clustering, classification, and embedding. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; pp. 1601–1608.
28. Nie, F.; Huang, H.; Cai, X.; Ding, C.H. Efficient and robust feature selection via joint $\ell_2, 1$ -norms minimization. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 6–9 December 2010; pp. 1813–1821.
29. Hou, C.; Nie, F.; Li, X.; Yi, D.; Wu, Y. Joint embedding learning and sparse regression: A framework for unsupervised feature selection. *IEEE Trans. Cybern.* **2014**, *44*, 793–804. [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).