

List of Figures

1.1	Graphical Abstract	15
2.1	Illustration of the behavior of sequence $\{\alpha_n\}$ vs. iterations.	23
3.1	Performance of algorithms on the basis of $F(x_n) - F(x^*)$ on datasets. $F(x_n)$ is the function value achieved after n^{th} iteration and $F(x^*)$ is the function value at the optimal point x^* . Shown graphs are log-log plots. We consider the convergence is reached if $F(x_n) - F(x_{n-1})$ becomes $10e-5$. Maximum number of iterations is set as $10e3$. The value of parameter θ is set 0.05 in all the experiments.	43
3.2	Performance of algorithms on the basis of $F(x_n) - F(x^*)$ on datasets. $F(x_n)$ is the function value achieved after n^{th} iteration and $F(x^*)$ is the function value at the optimal point x^* . Shown graphs are log-log plots. We consider the convergence is reached if $F(x_n) - F(x_{n-1})$ becomes $10e-5$. Maximum number of iterations is set as $10e3$. The value of parameter θ is set 0.05 in all the experiments.	44
3.3	Performance of NAGA on the basis of Reduction in Function Values in each iteration for the datasets. We consider the convergence is reached if $F(x_n) - F(x_{n-1})$ becomes $10e-5$. Maximum number of iterations is set as $10e3$. The value of parameter θ is set 0.05 in all the experiments.	46
3.4	Performance of NAGA on the basis of Reduction in Function Values in each iteration for the datasets. We consider the convergence is reached if $F(x_n) - F(x_{n-1})$ becomes $10e-5$. Maximum number of iterations is set as $10e3$. The value of parameter θ is set 0.05 in all the experiments.	47
3.5	Performance of NAGA on the basis of root Mean Square Error rate on the datasets. We consider the convergence is reached if $F(x_n) - F(x_{n-1})$ becomes $10e-5$. Maximum number of iterations is set as $10e3$. The value of parameter θ is set 0.05 in all the experiments.	48
3.6	Performance of NAGA on the basis of root Mean Square Error rate on the datasets. We consider the convergence is reached if $F(x_n) - F(x_{n-1})$ becomes $10e-5$. Maximum number of iterations is set as $10e3$. The value of parameter θ is set 0.05 in all the experiments.	49
3.7	Performance of NAGA-MURL in comparison to MURL on both the datasets.	57
3.8	Precision-recall curves on both the datasets for the two cross-modal retrieval tasks.	58

3.9	Performance of algorithms for the overlapping group lasso problem with the Breast cancer dataset. 3.9(a) shows the variation in iterations at different values of θ for ten randomly selected training samples. For all the other experiments we set $\theta = 0.005$. 3.9(b) is the log-log plot showing the convergence rate of the algorithms, which is considered to achieve when the $F(x_n) - F(x_{n-1}) < \varepsilon$ with $\varepsilon = 10E - 6$. 3.9(d) shows the convergence rates of NAGEL for different values of parameter β_n	63
3.10	Performance of algorithms for the Prostate Dataset. 3.10(a) shows the variation in iterations at different values of θ for ten randomly selected training samples. For all the other experiments we set $\theta = 5E - 5$. 3.10(b) is the log-log plot showing the convergence rate of the algorithms, which is considered to achieve when the $F(x_n) - F(x_{n-1}) < \varepsilon$ with $\varepsilon = 10E - 6$. 3.10(d) shows the convergence rates of NAGEL for different values of parameter β_n	65
3.11	Performance of algorithms for the Leukemia Dataset. 3.11(a) shows the variation in iterations at different values of θ for ten randomly selected training samples. For all the other experiments we set $\theta = 7.5E - 5$. 3.11(b) is the log-log plot showing the convergence rate of the algorithms, which is considered to achieve when the $F(x_n) - F(x_{n-1}) < \varepsilon$ with $\varepsilon = 10E - 6$. 3.11(d) shows the convergence rates of NAGEL for different values of parameter β_n	66
4.1	Single-task vs. Multitask Learning Frameworks	71
4.2	DNA to Protein Conversion. At 5' end (Left end), <i>cap</i> represents the added 7-methylguanosine, which we get after capping process. At the other end (3' end; right end), a poly(A) tail is formed with about 250 adenine residues.	76
4.3	Performance of VAGA on the basis of $F(W_n) - F(W^*)$ on the Sarcos, School and Parkinsons Datasets, respectively. $F(W_n)$ is the function value achieved after n^{th} iteration and $F(W^*)$ is the optimal function value. Shown graphs are log-log plots.	84
4.4	Performance of VAGA on the basis of Reduction in Function Value in each Iteration for the Sarcos, School and Parkinsons Datasets.	85
4.5	Performance of VAGA on the basis of rMSE Error rate for the Sarcos, School and Parkinsons Datasets.	86
4.6	Sparsity of Predictive Model when Changing Regularization Parameter for the School dataset.	87
4.7	Stability Results in terms of rMSE Values for all the three datasets.	88
4.8	Results for the joint Splice-site Recognition task	90
4.9	Results for the splice-site data sets from 6 eukaryotic genomes: Shown are auPRC performances of the few methods for each organism. We can observe that VAGA achieves the best results among all methods.	91

5.1	Performance of VAGA on the basis of $F(x_n) - F(x^*)$ on the four Datasets, Colon, Duke, Leukemia and Prostate, respectively. $F(x_n)$ is the function value achieved after n^{th} iteration and $F(x^*)$ is the optimal function value. Shown graphs are log-log plots. We consider the convergence is reached if $F(x_n) - F(x_{n-1})$ becomes $10e-4$. Maximum iterations is considered as $10e4$	107
5.2	Performance of the newly proposed operator splitting algorithm along with the accelerated variant on the basis of number of iterations required by each algorithm for the four Datasets, Colon, Duke, Leukemia and Prostate, respectively.	108
5.3	Performance of the newly proposed operator splitting algorithm along with the accelerated variant on the basis of Reduction in Function Value in each Iteration for the four Datasets, Colon, Duke, Leukemia and Prostate, respectively.	109
5.4	Performance of the newly proposed operator splitting algorithm along with the accelerated variant on the basis of classification accuracy by each algorithm for the four Datasets, Colon, Duke, Leukemia and Prostate, respectively.	110
5.5	Performance of the newly proposed operator splitting algorithm along with the accelerated variant on the basis of CPU Time required by each algorithm to reach the convergence for the four Datasets, Colon, Duke, Leukemia and Prostate, respectively. We consider the convergence is reached if $F(x_n) - F(x_{n-1})$ becomes $10e-4$	111
A.1	Illustration of convex functions.	119
A.2	Illustration of Subdifferential [197]	120