The Classification of Simulated Gene Expression Data Using SVM

September 8, 2019

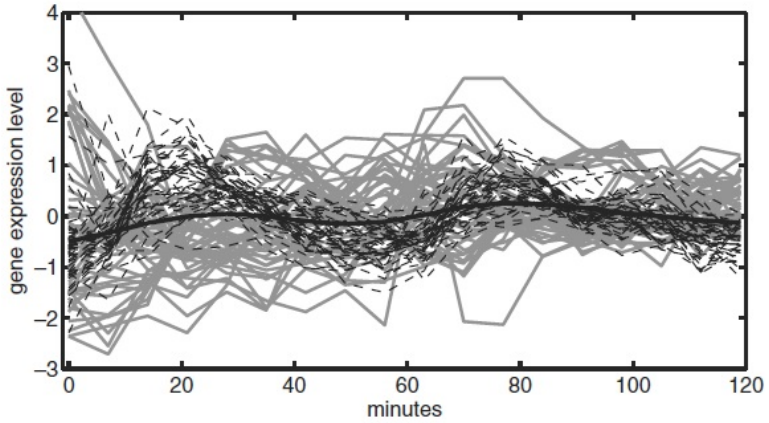

Fig. 1. Temporal gene expression profiles of yeast cell cycle. Dashed lines: G_1 phase; Gray solid lines: non-G_1 phases; Black solid line: overall mean curve.

- Previously simulated temporal gene expression data were used for SVM classification

- With these simulation data sets, five FPC scores were calculated

$$\hat{\epsilon}'_{im} = \sum_{k=1}^{S}((\hat{X}_i(k) - \hat{\mu}'(k))\hat{\rho}'_m(k),\ m = 1, ..., 5,\ S = 18$$

```
> head(set.e$train)
          e1          e2          e3          e4          e5 group
1 -0.8692159 -0.2356220  0.43377707  0.55379482  0.01063533    g1
2  0.5567556  0.9838879  0.41187097 -0.28080909 -0.23185472    g1
3 -2.8207010  0.6693475  0.12276236 -0.54525786  0.22456225    g1
4 -1.0252821  0.1991172  0.82753306  0.08093052 -0.24228697    g1
5 -0.5937367 -1.4236541  0.64962053  0.16398799 -0.51363319    g1
6 -1.0928235  0.4482800 -0.02650774 -0.55378990 -0.16838594    g1
```

```
glm(group ~ ., data=set.e$train)
```

```
svm(group ~ ., data=set.e$train)
```

- Using FPC scores in test data, the group of each case in test data was predicted by logistic regression and SVM models
- The kernels used in SVM are linear, polynomial, radial(Gaussian), and sigmoid
- And compared the classification error rates of the two models

- Classification Error rates

Table: FPCA and B-spline methods using the logistic regression

No. of FPCs or base functions	Group 1 FPCA	B-S	Group 2 FPCA	B-S	overall FPCA	B-S
1	32.72 (8.41)	27.32 (7.86)	32.70 (8.31)	26.90 (7.86)	32.71 (5.26)	27.11 (4.58)
2	22.16 (6.65)	24.08 (6.37)	22.06 (6.15)	24.80 (6.55)	22.11 (4.33)	24.44 (3.97)
3	7.58 (4.58)	7.92 (3.96)	8.26 (5.34)	8.76 (4.71)	7.92 (3.35)	8.34 (2.70)
4	7.14 (4.14)	8.18 (4.18)	7.62 (5.10)	8.98 (5.00)	7.38 (3.11)	8.58 (2.96)
5	7.40 (4.07)	7.68 (4.29)	7.86 (5.26)	8.58 (5.01)	7.63 (3.06)	8.13 (3.15)

Table: SVM with FPC scores

No. of FPCs	Group 1 linear	(degree=3) polynomial	Group 2 linear	polynomial	overall linear	polynomial
1	63.20 (16.53)	62.02 (25.20)	47.38 (4.97)	42.32 (27.92)	55.29 (6.86)	52.17 (4.85)
2	21.90 (7.15)	25.54 (15.76)	22.80 (6.81)	24.28 (14.04)	22.35 (4.22)	24.91 (5.01)
3	7.60 (4.60)	10.02 (7.57)	8.32 (5.02)	9.84(7.28)	7.96 (3.27)	9.93 (4.16)
4	6.86 (4.19)	8.48 (5.82)	7.82 (4.92)	9.64 (6.47)	7.34 (2.98)	9.06 (3.48)
5	7.14 (4.02)	9.66 (6.26)	7.88 (5.22)	9.34 (6.05)	7.51 (3.10)	9.50 (3.72)

No. of FPCs	Group 1 radial(γ=1)	sigmoid	Group 2 radial	sigmoid	overall radial	sigmoid
1	62.32 (17.94)	53.68 (12.34)	46.48 (10.58)	46.86 (10.92)	54.40 (6.19)	50.27 (3.34)
2	22.34 (8.34)	26.68 (8.28)	23.72 (7.60)	27.74 (8.75)	23.03 (4.18)	27.21 (5.69)
3	7.86 (4.38)	9.56 (5.21)	9.14 (5.30)	9.94 (6.00)	8.50 (3.30)	9.75 (3.36)
4	7.86 (4.39)	8.56 (4.94)	8.32 (4.80)	9.20 (4.79)	8.09 (3.13)	8.88 (2.86)
5	8.66 (4.75)	8.28 (4.85)	8.96 (5.01)	8.72 (5.18)	8.81 (3.30)	8.50 (3.11)

```
> head(set.i$train)
          1          2          3          4          5          6          7          8
1 -0.8618787 -0.6352476 -0.3415388 -0.0531040  0.04309674  0.02047068 -0.01099684 -0.04526450
2  0.4191141  0.1455574 -0.1963211 -0.4509614 -0.52776327 -0.45267674 -0.42304929 -0.05560363
3 -1.5101324 -1.5695820 -1.3400406 -0.9154486 -0.39168051  0.05849471  0.29938958  0.40799901
4 -0.8856889 -0.6730592 -0.4612614 -0.3003018 -0.23947601 -0.38612667 -0.23344855 -0.21814454
5 -1.3146280 -0.4275777  0.2698759  0.5934907  0.53701947  0.21786719 -0.11710036 -0.35663968
6 -0.7651697 -0.7336608 -0.6406515 -0.4694104 -0.39103834 -0.01681058  0.09959576  0.13128335
          9         10         11         12         13         14         15         16
1  0.03973835  0.1762405 0.36304832  0.45021401  0.48652470  0.42424557  0.33174576  0.23100277
2  0.07672491  0.1263219 0.15229632  0.08219731 -0.01198886 -0.05288686 -0.03319578 -0.05733075
3  0.36852375  0.2799748 0.20390030  0.28020091  0.40217020  0.59500617  0.77938190  0.84693489
4 -0.13431123  0.1173825 0.31469464  0.41586603  0.42518691  0.41705869  0.25553538  0.40524798
5 -0.40351146 -0.2491179 0.06205509  0.28848360  0.37162681  0.39401966  0.30478151  0.21419569
6  0.17888619  0.1212061 0.01733962 -0.00283926  0.04391332  0.01369034  0.30113695  0.40186591
         17         18 group
1 0.02828115 -0.1102631    g1
2 0.27576290  0.4480444    g1
3 0.88092514  0.8613920    g1
4 0.33932037  0.3067383    g1
5 0.10807329  0.0451095    g1
6 0.49712274  0.6789717    g1
```

```
svm(group ~ ., data=set.i$train)
```

```
svm(group ~ ., data=set.pca.scores)
```

- Next, SVM was applied to temporal data and five PC scores data to classify the groups
- And the classification error rates were compared with the previous results

- Classification Error rates

Table: FPCA and B-spline methods using the logistic regression

No. of FPCs or base functions	Group 1 FPCA	B-S	Group 2 FPCA	B-S	overall FPCA	B-S
3	7.58 (4.58)	7.92 (3.96)	8.26 (5.34)	8.76 (4.71)	7.92 (3.35)	8.34 (2.70)
4	7.14 (4.14)	8.18 (4.18)	7.62 (5.10)	8.98 (5.00)	7.38 (3.11)	8.58 (2.96)
5	7.40 (4.07)	7.68 (4.29)	7.86 (5.26)	8.58 (5.01)	7.63 (3.06)	8.13 (3.15)

Table: SVM with FPC scores

No. of FPCs	Group 1 linear	(degree=3) polynomial	Group 2 linear	polynomial	overall linear	polynomial
3	7.60 (4.60)	10.02 (7.57)	8.32 (5.02)	9.84(7.28)	7.96 (3.27)	9.93 (4.16)
4	6.86 (4.19)	8.48 (5.82)	7.82 (4.92)	9.64 (6.47)	7.34 (2.98)	9.06 (3.48)
5	7.14 (4.02)	9.66 (6.26)	7.88 (5.22)	9.34 (6.05)	7.51 (3.10)	9.50 (3.72)

Table: SVM with temporal data and PC scores

kernels	Group 1 temporal	(No. of PC=5) PC	Group 2 temporal	PC	overall temporal	PC
linear	7.06 (3.98)	15.98 (15.73)	8.18 (5.22)	15.32 (14.85)	7.62 (2.99)	15.65 (14.38)
polynomial	9.12 (7.14)	18.02 (16.56)	10.90 (7.44)	16.72 (13.81)	10.01 (3.79)	17.37 (14.22)
radial	8.08 (4.72)	16.50 (13.82)	8.90 (5.12)	16.67(13.79)	8.49 (3.30)	16.58 (13.48)
sigmoid	12.42 (6.59)	15.96 (14.03)	13.08 (6.20)	16.72 (14.08)	12.75 (4.05)	16.34 (13.68)