

# Multiple Classification of Simulated Gene Expression Data

November 4, 2019

① Prediction Comparison Between GLM and LDA

② Multiple Classification

# 1. Prediction Comparison Between GLM and LDA

- Checked whether two predictions coincide (GLM vs LDA)

\$M1			\$M2			\$M3			\$M4			\$M5		
0		1	0		1	0		1	0		1	0		1
0	5038	18	0	5020	47	0	4903	74	0	4861	120	0	4845	139
1	24	4920	1	59	4874	1	129	4894	1	141	4878	1	152	4864

**Table:** The number of correct predictions

No. of FPCs	GLM correct	LDA correct	GLM equal to LDA
1	6718	6732	6704
2	7761	7769	7712
3	9221	9220	9119
4	9239	9312	9145
5	9240	9303	9126

- LDA showed slightly better prediction

## 2. Multiple Classification

- Generate the simulation data curves using FPC scores( $\epsilon_m$ ) from normal distribution

$$X_i(t) = \mu(t) + \sum_{m=1}^M \epsilon_{im} \rho_m(t) \quad 0 \leq t \leq T.$$

- With the simulation data curves, five FPC scores were calculated

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

- Multiple(six groups) classification was performed with four mean sets

```
> mean.set1
```

```
$mean1
```

```
[1] 0.6 0.5 0.4 0.3 0.2
```

```
$mean2
```

```
[1] 3.0 2.5 2.0 1.5 1.0
```

```
$mean3
```

```
[1] 100 99 98 97 96
```

```
> mean.set2
```

```
$mean1
```

```
[1] 0.6 0.5 0.4 0.3 0.2
```

```
$mean2
```

```
[1] 5.0 4.5 4.0 3.5 3.0
```

```
$mean3
```

```
[1] 10.0 9.5 9.0 8.5 8.0
```

```
> mean.set3
```

```
$mean1
```

```
[1] 0.6 0.5 0.4 0.3 0.2
```

```
$mean2
```

```
[1] 3.0 2.5 2.0 1.5 1.0
```

```
$mean3
```

```
[1] 1 6 11 16 21
```

```
> mean.set4
```

```
$mean1
```

```
[1] 0.6 0.5 0.4 0.3 0.2
```

```
$mean2
```

```
[1] 3.0 2.5 2.0 1.5 1.0
```

```
$mean3
```

```
[1] 10.0 9.5 9.0 8.5 8.0
```

## 2. Multiple Classification

**Table:** Multiple Classification Error rates of LDA, QDA, SVM

No. of FPCs	LDA			
	mean set 1	mean set 2	mean set 3	mean set 4
1	16.56 (2.27)	11.10 (1.74)	16.56 (2.27)	16.56 (2.27)
2	12.37 (2.26)	8.65 (1.92)	12.98 (2.14)	12.28 (2.25)
3	8.56 (1.68)	5.90 (1.70)	9.02 (1.88)	8.46 (1.69)
4	8.47 (1.77)	5.76 (1.62)	8.90 (1.65)	8.32 (1.78)
5	8.40 (1.74)	5.65 (1.52)	8.77 (1.70)	8.31 (1.73)

  

No. of FPCs	QDA			
	mean set 1	mean set 2	mean set 3	mean set 4
1	16.63 (2.36)	11.08 (1.70)	16.63 (2.36)	16.63 (2.36)
2	11.52 (2.03)	7.82 (1.51)	11.52 (2.04)	11.52 (2.04)
3	4.92 (1.41)	2.65 (0.94)	4.92 (1.42)	4.92 (1.42)
4	4.75 (1.36)	2.40 (0.78)	4.75 (1.36)	4.75 (1.36)
5	4.88 (1.38)	2.51 (0.85)	4.88 (1.38)	4.88 (1.38)

  

No. of FPCs	SVM			
	mean set 1	mean set 2	mean set 3	mean set 4
1	17.68 (2.24)	11.12 (1.69)	16.88 (2.37)	16.68 (2.20)
2	35.14 (4.74)	8.54 (1.50)	13.19 (1.90)	12.34 (1.89)
3	35.90 (2.99)	3.50 (1.10)	8.17 (1.87)	6.96 (1.59)
4	35.63 (3.10)	3.39 (0.99)	8.10 (1.81)	6.84 (1.58)
5	35.54 (3.01)	3.45 (1.02)	8.14 (1.83)	6.96 (1.53)

## 2. Multiple Classification

- Contingency table for mean set 1 and set 4

Figure: mean set 1 - left:LDA, mid:QDA, right:SVM

	1	2	3	4	5	6
1	3829	619	75	5	0	0
2	630	3829	10	89	0	0
3	406	120	4915	0	0	0
4	135	432	0	4906	0	0
5	0	0	0	0	5000	0
6	0	0	0	0	0	5000

	1	2	3	4	5	6
1	4439	401	104	26	0	0
2	331	4349	15	106	0	0
3	183	36	4881	0	0	0
4	47	214	0	4868	0	0
5	0	0	0	0	5000	0
6	0	0	0	0	0	5000

	1	2	3	4	5	6
1	3511	551	2124	1451	0	0
2	498	3397	1426	2099	0	0
3	635	382	1220	240	0	0
4	356	670	230	1210	0	0
5	0	0	0	0	5000	0
6	0	0	0	0	0	5000

Figure: mean set 4 - left:LDA, mid:QDA, right:SVM

	1	2	3	4	5	6
1	3852	609	79	5	0	0
2	602	3841	10	91	0	0
3	407	118	4911	0	0	0
4	139	432	0	4904	0	0
5	0	0	0	0	5000	0
6	0	0	0	0	0	5000

	1	2	3	4	5	6
1	4439	401	104	26	0	0
2	331	4349	15	106	0	0
3	183	36	4881	0	0	0
4	47	214	0	4868	0	0
5	0	0	0	0	5000	0
6	0	0	0	0	0	5000

	1	2	3	4	5	6
1	4057	483	32	9	0	0
2	369	3947	12	40	0	0
3	448	104	4956	0	0	0
4	126	466	0	4951	0	0
5	0	0	0	0	5000	0
6	0	0	0	0	0	5000

- In means set 1, huge difference between mean1 and mean3 led to large classification error for SVM