

Computational Learning

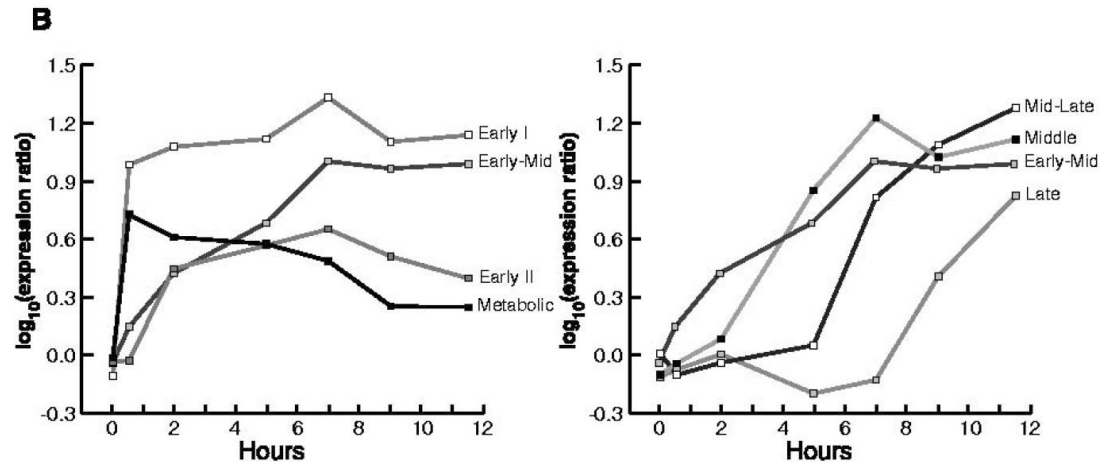
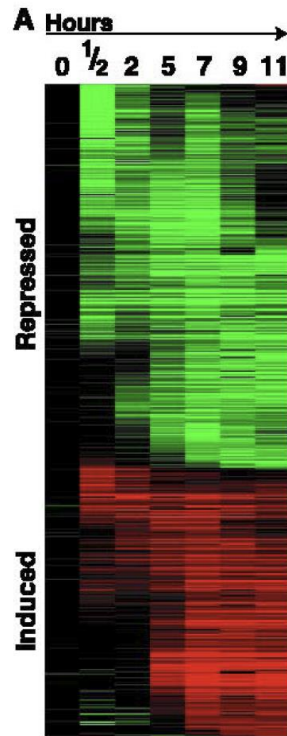
Pract 1

APPLICATION OF CLUSTERING METHODS TO
SPORULATION YEAST MICROARRAY DATA

AGENDA

- Read introductory papers [Ref. 1] and [Ref. 2] in the virtual campus, and some information about Yeasts [Ref. 3].
- By using the code provided, understand what is the aim of clustering and how the silhouette index quantifies the results obtained.
- Use k-Means algorithm for clustering using available datasets.
- In [Ref. 4] and [Ref. 5] there are more advanced clustering analysis on the sporulation data set.

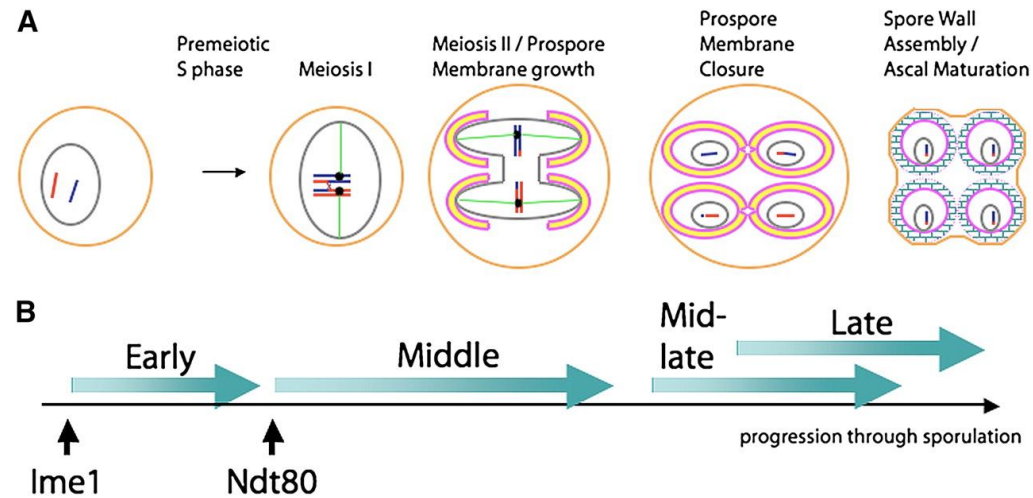
SPORULATION YEAST DATASET



Genes used to create average temporal profiles

Metabolic	Early I	Early II	Early-Mid	Middle	Mid-Late	Late
ACS1	ZIP1	KGD2	YBL078C	YSW1	CDC27	SPS100
PYC1	YDR374C	AGA2	QRI1	SPR28	DIT2	YKL050C
SIP4	DMC1	YPT32	PDS1	SPS2	DIT1	YMR322C
CAT2	HOP1	MRD1	APC4	YLR227C		YOR391C
YOR100C	IME2	SPO16	KNR4	ORC3		
CAR1		NAB4	STU2	YLL005C		
		YPR192W	YNL013C	YLL012W		
			EXO1			

SPORULATION YEAST DATASET



SPORULATION YEAST DATASET

	A	B	C	D	E	F	G	H
1	Genes	t0	t0,5	t2	t5	t7	t9	t11,5
2	YAL025C	1,163781	-1,758143	-0,31495	0,283874	-0,6779	0,470183	0,833156
3	YAL036C	0,955552	-1,48956	-0,499057	-0,099143	-0,674453	1,414381	0,39228
4	YAL040C	1,454598	-0,735149	-0,073602	-0,618896	-0,789697	1,391507	-0,62876
5	YDL037c	1,698578	-1,268638	-0,375511	0,996817	-0,319158	-0,344837	-0,387251
6	YDR184C	1,302831	-1,338633	-0,555163	-0,101402	-0,674167	1,317301	0,049233
7	YDR299W	1,431937	-1,819908	0,056708	0,628218	-0,461532	0,001474	0,163103
8	YDR380W	0,472715	-1,713187	-0,916225	-0,131575	0,433669	0,731209	1,123394
9	YDR398W	1,235163	-1,796627	-0,062241	0,41659	-0,767331	0,539733	0,434713
10	YER006w	1,927985	-1,138106	-0,583176	0,056551	-0,731212	0,225849	0,242109
11	YER064c	1,93726	-1,094734	-0,423042	-0,165499	-0,823443	0,316565	0,252894
12	YER126c	1,579942	-1,77506	-0,136119	0,28206	-0,372171	0,093242	0,328105
13	YGL171W	0,273979	-1,959266	-0,674947	0,189875	0,616735	0,837517	0,716106
14	YIL122W	1,857839	-1,286335	-0,338097	0,062704	-0,199524	0,556297	-0,652884
15	YIL133C	1,298308	-1,614989	0,082606	-0,517775	-0,498323	0,177034	1,073139
16	YKL009W	1,574891	-1,377246	-0,807925	0,426834	-0,654867	0,356929	0,481385
17	YKL143W	0,906054	-2,074963	-0,284458	0,347086	-0,019379	0,424481	0,701179
18	YLR175W	1,421891	-1,723291	-0,522326	0,283152	-0,414976	0,362122	0,593428
19	YLR196W	1,567397	-1,645187	-0,385901	0,239011	-0,459244	0,646757	0,037166
20	YLR222C	1,096441	-1,728575	-0,529348	0,075671	-0,434116	1,060268	0,459658
21	YLR062C	1,46387	-1,311697	-0,836072	-0,48548	-0,238914	0,427468	0,980826
22	YLR129w	1,232222	-1,990873	-0,166368	0,197632	-0,168563	0,344202	0,551748
23	YLL011w	1,28004	-1,650058	-0,731675	0,305035	-0,424647	0,405359	0,815947

Gene expression levels are measured at 7 time intervals: the aim of the lab is to cluster the genes into groups that may be significative in different stages of the process.

TASK

- Analyze the clustering performance of k-Means in the sporulation yeast problem, comparing the obtained results with those from published ones [Ref. 2].
- Write a report containing an introduction to the problem, a description of the methods, most relevant results obtained comparing both methods and the conclusions that can be extracted. The report must be written in Latex by using the LNCS template.

TASK

- You must submit:
 - The report in pdf format.
 - The Latex project (in a compressed format) to generate the report.
 - The code (in ipynb format) you have developed to deal with the problem.
 - A video (in mp4 format) with the execution of the code.
 - A video with a 5 minutes presentation of your work.
 - The slides you used for your presentation.
 - A signed statement assuming the originality of the work submitted.