OverviewINFO411 Lecture 3 - Online Clustering The ProblemOnline AveragingOnline AveragingSometitive LearningBasic principlesSOMNeural Gas 25/7/2017 25/7/2017 Ecader-Follower The Idea Examples Sometitive Learning Sometitive Learning Sometitive Learning Basic principles Basic principles Sometitive Learning Basic principles Sometitive Learning Basic principles Sometitive Learning Basic principles Basic principles Basic principles Basic principles Basic principles Basic principles Basic pr

	Lecture 3	1 / 31		Lecture 3	2 / 31
				The Problem	
References		'	The k -Means Algor	rithm	

- Alpaydin, Chapter 12 (2nd / 3rd Ed.)
- Hertz, Krogh & Palmer, Introduction to the Theory of Neural Computation, Chapter 9
- Duda, Hart & Stork, Pattern Classification, Section 10.11

- Given k, the k-means algorithm is implemented in the following steps:
 - **(**) Partition data items into k non-empty subsets
 - **②** Obtain the centroids as the centers (mean points) of the partitions.
 - Obtain new partitions: assign each data item to the cluster of the nearest centroid.
 - Stop when no more new assignment is found; otherwise go back to Step 2.
- The algorithm may need to go through many iterations before it terminates or converges.

The Problem

Online Learning: Challenges

Online averaging: a bit of DIY

Online Averaging

- In traditional clustering,
 - Cluster structure can be sensitive to small changes or noises in data.
 - Clustering is mostly done in batch mode.
- In online learning:
 - ► Data may arrive incrementally but constantly
 - Limited memory: data need to go through single-pass
 - ► Limited processing time
 - Evolving data: concept drifts may exist
- What's required:
 - ▶ **Incremental learning** ability: learning data piece-by-piece.
 - ► **Stability**: cluster structure not easily drifted
 - ▶ **Plasticity**: being adaptive and possibly allowing new clusters

$$n \leftarrow 0, \operatorname{avg}_{0} \leftarrow 0$$

while true do

$$x_{n} \leftarrow \operatorname{random}()$$

$$\operatorname{avg}_{n} \leftarrow \operatorname{avg}_{n-1} + \gamma_{n}(x_{n} - \operatorname{avg}_{n-1})$$

$$n \leftarrow n+1$$

end while

Good experimental results can be obtained with very small γ values, or $\gamma_n = 1/n^p, \ p > 1.$

In real-world scenarios with dynamic data environments does this work? Let's find out...

7 / 31

۲

Optimization in Online k-Means

Another take on the reconstruction error for k-means clustering:

$$E(\{\mathbf{m}_i\}_{i=1}^k | \mathcal{X}) = \sum_t \sum_i b_i^t \|\mathbf{x}^t - \mathbf{m}_i\|^2$$

where

$$b_i^t = \begin{cases} 1 & \text{if } \|\mathbf{x}^t - \mathbf{m}_i\|^2 = \min_j \|\mathbf{x}^t - \mathbf{m}_j\| \\ 0 & \text{otherwise} \end{cases}$$

In online learning, we approximate gradient descent with stochastic gradient descent (SGD), doing a small update on clusters at each step. The criterion function at step t is

$$E^{t}(\{\mathbf{m}_{i}\}_{i=1}^{k}|\mathbf{x}^{t}) = \sum_{t}\sum_{i}b_{i}^{t}\|\mathbf{x}^{t} - \mathbf{m}_{i}\|^{2}$$

By SGD (see e.g. (Bottou & Benjio, 1995)), we have

$$\Delta \mathbf{m}_i = -\eta \frac{\delta \mathbf{E}^t}{\delta \mathbf{m}_i} = \eta b_i^t (\mathbf{x}^t - \mathbf{m}_i)$$

Competitive Learning

- Competitive learning is a methodology based on neuroscience research.
- CL schemes
 - ▶ Basic competitive learning
 - Fixed number of clusters
 - "Winner-takes-all"
 - ► Soft competitive learning
 - Allows multiple winning neurons
 - ► Leader-Follower clustering
 - Allows a variable number of neurons

Competitive Learning	Basic principles	Competitive Learning	Basic principles
Basic C.L. algorithm		CL Characteristics	

• a.k.a. 'local k-means'

Pseudocode

- **1** Initialize weights $\{\mathbf{w}_i\}, i = 1, 2, ..., k$
- **2** Randomly select a pattern \mathbf{x}
- Find the winner neuron:
 - $b = \operatorname{argmin} \|\mathbf{x} \mathbf{w}_i\|$
- Update the winner neuron $\Delta \mathbf{w}_b = \gamma(\mathbf{x} - \mathbf{w}_b)$
- Goto step 2 until no significant change in weights.

Localized learning - good for online implementation

- Local minimum problem
- \mathbb{N} Fixed number of neurons
- \mathbb{N} Slow adaptability to novelty
 - ► Can you tell why?

	Lecture 3	9 / 31		Lecture 3	10 / 31
Competitive Learning Basic principles			Com		
CL: How to impro	ove?		Self-Organizing Ma	aps	

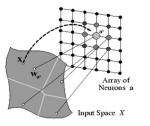
- Instead of tuning the winning neuron alone, other neurons also involved in adapting?
 - ▶ More robustness in the 'codebook'.
 - ▶ Can introduce relationship between prototypes.
 - ► However more time-consuming
- Dealing with uneven winning frequencies: frequency-sensitive FSCL, rival penalty RPCL
- More adaptability? E.g.,
 - ▶ growing and pruning,
 - ▶ merging and splitting etc.
- Can the learnt prototypes be useful for classification?
- Parallel implementation?

- - Kohonen (1982)
 - aka Self-organizing feature map (SOFM) or Kohonen map
 - Found thousands of applications, including:
 - ► Speech recognition
 - ▶ Image compression
 - Bankruptcy prediction
 - ► Telecommunication traffic monitoring
 - Process control
 - ▶ Web document indexing

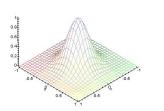
Competitive Learning SOM

The SOM Model

- Introduces a topology for prototype nodes (ordering, neighbourhood)
- Define a neighbourhood function $\Omega(y_i, y_b)$ for prototype indeces $\{y_i\}$:
 - Bubble: $\Omega(y_i, y_b) = 1$ or 0
 - ▶ Gaussian: centered at the winner
 - ▶ "Mexican hat": lateral inhibition
- Nodes within the neighbourhood of the winner also get updated.



The nodes arranged in 2xD grids



The 'Mexican hat' neighbourhood

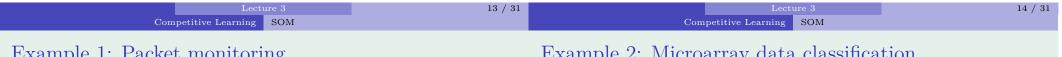
• Each adaptation tunes the winner (or "best matching unit" / BMU) and its neighbours:

 $\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \gamma(t)\Omega(y_i, y_b)(\mathbf{x} - \mathbf{w}_i)$

• During the iterations

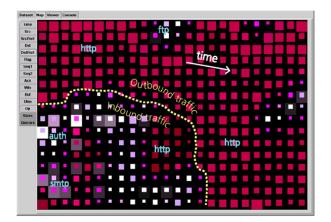
The SOM algorithm

- Neighbourhood $\Omega(y_i, y_b)$ shrinks over time
- Learning rate $\gamma(t)$ reduces over time
- Can operate either incrementally, or in batch mode



Example 1: Packet monitoring

Mapping multi-dimensional packet data, one can use SOM to analyze network traffic, monitor online traffic, or even visualize intrusions.

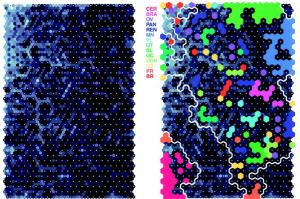


Luc Girardin, USENIX'99 workshop

Example 2: Microarray data classification

Covet et al., Molecular Classification of Cancer: Unsupervised Self-Organizing Map Analysis of Gene

Expression Microarray Data, Mol Cancer Ther, March 2003 2; 317



Example 3: EMU macroeconomics analysis

• Kasabov et al. (2001)

- Macroeconomic data collected for then EMU countries and non-EU countries
- Attributes include: GDP, debt, deficit, inflation rate, interest rate, unemployment, balance of payment, production gap etc.
- Trained SOMs can be coloured by components

1194	BG92	BG93	BG94 BG95	BGS	6 BC	697 ^E	G98 T98	GR9	8	JF	98	JP9	5 JF	992	
			0000												93
	GR96	1T96										OE9	7		
	95		s	096	US95	NL9	5 NL	.94		NL96	US	96	DE98	NL	97
PT05	PTOS												198	NLS	18
		intoo			NI 02	DKO					DK	0.0	0K9 7	DK	98
		DK02													98
IDC											K97				10
			U	592							297				10
	93 FN					<u> </u>									17
ES96		UK93	FR93	UKS	4 SV	V94 F	R94	FN9	ES	97 E	598	FN9	7 FI	198	
IT94	BG92	BG93	BG94 BG95	BGS	96 BG	697 E	G98 T98	GR9	B	JF	°98	JP9	5 JF	92	
IT94	BG92	BG93	BG94 BG95	BGS	96 BG	697 F	T98	GR9	B	JF	P98	JP9	5 JF	92	
.94	П	95			IT97		GF	197		JP97	JP	96	IP94	JP	33
	GR96	IT96							Τ.			OE9	7		
SD	95		SI	D96	US95	NL9	5 NL	.94		NL96	US	96	DE98	NLS	97
DTOF	DTOP												198	INLS	10
P195	P190	IK90											NV07	DI	~~
IRS	95		N	L92	NL93	DK9	4	C	E95		DK	196 L	JS97	US	98
		DK92		USS	13 US	694 E	3D94			PT	T97 K97		U	(98	
IRS	94 OE													IR	98
		BD92									097	SD9	7 SI	098	
ENK															7
P NS	93 PN	192 Fh	192			FR9	FR FR	90 1	1480		FR	90		IRE	"
ES94		UK92						ENIC							
	SD PT95 IR IR ES94 IT94 SD PT95 IR IR IR	GROG SDG5 PT95 PT96 IR95 IR94 OI FN93 FF ES94 IT94 BG92 SDG5 PT95 PT96 IR95 PT96 IR95 OI	GR06 TT96 SD05 IR95 PT95 PT96 IR95 DK02 IR94 OE02 PM03 FN02 FN03 FN02 ES86 UK02 ES96 UK03 IT94 BG92 BG93 IG95 IT95 IT96 SD95 IR95 DK92 IR95 DK92 IR95 IR95 DK92 IR92	CR00 TT30 TT30 TT30 TT30 S13 PT95 PT95 IR95 IR95 DK92 DK92 IR95 CR03 PH92 PH92 PH92 PH92 FM03 PH92 PH92 PH92 PH92 PH92 E584 UK83 FR93 BC345 PH93 IT94 BC92 BC93 BC345 SD95 SS SS SS PT95 PT96 IR96 DK92 IR95 ES94 DK92 DK92 IR95 CR95 SS SS P195 P196 CR95 DK92 IR95 DE92 DK92 PU2	CR00 TT30 TT30 SD05 SD05 SD05 PT95 PT95 IR95 CK93 IR95 DK92 SD05 NL92 IR95 CK92 PT99 R09 R09 R09 FN03 FN02 FN03 FN03 FN03 FN03 FN03 FN02 FN03 FN03 FN03 FN03 SD05 SD05 SD05 F05 F05 SD05 SD06 F05 F05 F05 SD05 SD05 SD05 F053 F053 FN93 F194 GR95 IR95 DK92 F053 SD05 SD05 SD05 F053 F053 F053 FR95 F196 CE92 CE92 CE92 CE92 CE92 CE92 CE92 E592 E592 E593 CE92 E593 CE93 CE93 E593 CE93 E593 CE93 F053 F053 F053 <td< td=""><td>GR00 TT30 TT30 SD05 DK33 NL93 IR95 IR96 DK32 NL93 IR95 DK92 DU932 NL93 IR95 OE92 OB93 OE93 OE93 IR95 PH92 FR92 IS93 OE93 OE94 OE93 OE93</td><td>GR00 TT00 SD05 SD06 US93 NL93 DK9 IR95 IR96 DK92 US93 US93 DK9 IR95 IR96 DK92 US93 US93 UK93 IR95 IR96 DK92 US93 UK93 IK96 IR95 PH92 PR92 US93 UK93 IK96 FN03 PH92 PR92 IK96 IK96 PR93 ES94 UK93 FR93 IK96 IK96 PR94 IT94 BG92 BG93 BG94 BG94 PG94 SD95 IT97 TT97 IT97 IT97 IT97 GR96 IT95 SD96 US95 IL94 IL92 IL94 IL94 SD95 IS92 US93 US93 IS94 IL94 IR95 DB92 ID93 IL93 IL94 IL94</td><td>GR06 TT36 <th< td=""><td>GR00 TT30 <th< td=""><td>GR00 TT00 <th< td=""><td>GR06 TT06 SD96 US95 NL96 NL94 NL96 PT95 IR96 DK95 NL93 DK93 NL93 DK94 OE95 IR95 DK92 NL93 US93 US94 DD94 OE95 IR95 DK92 US93 US93 UK95 DK96 D096 IR95 PT95 IR95 PT95 IR95 DK96 D096 IR95 PT92 IR95 IR96 FR96 FR97 FR97 FN03 FN02 FR02 FR95 FR96 FR97 FR97 FN03 FN04 FR95 TR96 FR96 FR97 FR97 FN03 FR93 IR96 GR97 FR98 GR97 JP97 GR96 T197 GR97 IR97 GR97 JP97 GR96 T197 GR97 IR98 GR98 JP97 JR96 IR98 DK95 NL93 DK94 OE97</td><td>GR66 1766 SD05 SD96 US95 NL95 NL95 NL96 US95 PT95 PT96 IR96 DK93 NL93 DK94 OE94 OE95 D IR95 PT96 IR96 DK92 NL93 DK94 OE95 D IR95 DK92 US93 US94 BD96 BD97 BD97 BD97 BD97 BD97 BD97 BD97 IP97 JP97 JP97</td><td>GR06 TT06 SD09 US95 NL95 NL95 NL96 US96 OE90 PT05 FR06 FR05 NL93 DK93 DK93 DK94 OE95 DK96 PT07 FR05 FR07 BD02 BD03 DE93 UK95 DK95 DK95<</td><td>GR00 TT00 SD00 US00 US00 NL00 NL00 QS00 PT00 PT00 PT00 R00 DK93 NL03 DK94 CE04 CE05 DK95 DK95 DK95 DK95 DK95 DK95 CE04 CE05 DK95 DK9</td><td>GR00 TT06 SD00 US95 NL95 NL94 NL90 US95 NL95 PT05 PT06 R80 DK95 NL95 NL94 OE95 OE95 NL95 PT05 PT06 R80 DK95 OE94 OE95 DK96 DK97 DK96 DK92 US93 US93 US93 US95 DE96 OE95 DK98 D097 DK96 DK92 US93 US93 US93 US95 BD96 OF97 DK96 B092 DB03 OE93 UK92 PK97 PK98 DF97 DK94 B092 DB03 OE93 UK93 PK98 PK98 F89 F89<</td></th<></td></th<></td></th<></td></td<>	GR00 TT30 TT30 SD05 DK33 NL93 IR95 IR96 DK32 NL93 IR95 DK92 DU932 NL93 IR95 OE92 OB93 OE93 OE93 IR95 PH92 FR92 IS93 OE93 OE94 OE93 OE93	GR00 TT00 SD05 SD06 US93 NL93 DK9 IR95 IR96 DK92 US93 US93 DK9 IR95 IR96 DK92 US93 US93 UK93 IR95 IR96 DK92 US93 UK93 IK96 IR95 PH92 PR92 US93 UK93 IK96 FN03 PH92 PR92 IK96 IK96 PR93 ES94 UK93 FR93 IK96 IK96 PR94 IT94 BG92 BG93 BG94 BG94 PG94 SD95 IT97 TT97 IT97 IT97 IT97 GR96 IT95 SD96 US95 IL94 IL92 IL94 IL94 SD95 IS92 US93 US93 IS94 IL94 IR95 DB92 ID93 IL93 IL94 IL94	GR06 TT36 TT36 <th< td=""><td>GR00 TT30 <th< td=""><td>GR00 TT00 <th< td=""><td>GR06 TT06 SD96 US95 NL96 NL94 NL96 PT95 IR96 DK95 NL93 DK93 NL93 DK94 OE95 IR95 DK92 NL93 US93 US94 DD94 OE95 IR95 DK92 US93 US93 UK95 DK96 D096 IR95 PT95 IR95 PT95 IR95 DK96 D096 IR95 PT92 IR95 IR96 FR96 FR97 FR97 FN03 FN02 FR02 FR95 FR96 FR97 FR97 FN03 FN04 FR95 TR96 FR96 FR97 FR97 FN03 FR93 IR96 GR97 FR98 GR97 JP97 GR96 T197 GR97 IR97 GR97 JP97 GR96 T197 GR97 IR98 GR98 JP97 JR96 IR98 DK95 NL93 DK94 OE97</td><td>GR66 1766 SD05 SD96 US95 NL95 NL95 NL96 US95 PT95 PT96 IR96 DK93 NL93 DK94 OE94 OE95 D IR95 PT96 IR96 DK92 NL93 DK94 OE95 D IR95 DK92 US93 US94 BD96 BD97 BD97 BD97 BD97 BD97 BD97 BD97 IP97 JP97 JP97</td><td>GR06 TT06 SD09 US95 NL95 NL95 NL96 US96 OE90 PT05 FR06 FR05 NL93 DK93 DK93 DK94 OE95 DK96 PT07 FR05 FR07 BD02 BD03 DE93 UK95 DK95 DK95<</td><td>GR00 TT00 SD00 US00 US00 NL00 NL00 QS00 PT00 PT00 PT00 R00 DK93 NL03 DK94 CE04 CE05 DK95 DK95 DK95 DK95 DK95 DK95 CE04 CE05 DK95 DK9</td><td>GR00 TT06 SD00 US95 NL95 NL94 NL90 US95 NL95 PT05 PT06 R80 DK95 NL95 NL94 OE95 OE95 NL95 PT05 PT06 R80 DK95 OE94 OE95 DK96 DK97 DK96 DK92 US93 US93 US93 US95 DE96 OE95 DK98 D097 DK96 DK92 US93 US93 US93 US95 BD96 OF97 DK96 B092 DB03 OE93 UK92 PK97 PK98 DF97 DK94 B092 DB03 OE93 UK93 PK98 PK98 F89 F89<</td></th<></td></th<></td></th<>	GR00 TT30 TT30 <th< td=""><td>GR00 TT00 <th< td=""><td>GR06 TT06 SD96 US95 NL96 NL94 NL96 PT95 IR96 DK95 NL93 DK93 NL93 DK94 OE95 IR95 DK92 NL93 US93 US94 DD94 OE95 IR95 DK92 US93 US93 UK95 DK96 D096 IR95 PT95 IR95 PT95 IR95 DK96 D096 IR95 PT92 IR95 IR96 FR96 FR97 FR97 FN03 FN02 FR02 FR95 FR96 FR97 FR97 FN03 FN04 FR95 TR96 FR96 FR97 FR97 FN03 FR93 IR96 GR97 FR98 GR97 JP97 GR96 T197 GR97 IR97 GR97 JP97 GR96 T197 GR97 IR98 GR98 JP97 JR96 IR98 DK95 NL93 DK94 OE97</td><td>GR66 1766 SD05 SD96 US95 NL95 NL95 NL96 US95 PT95 PT96 IR96 DK93 NL93 DK94 OE94 OE95 D IR95 PT96 IR96 DK92 NL93 DK94 OE95 D IR95 DK92 US93 US94 BD96 BD97 BD97 BD97 BD97 BD97 BD97 BD97 IP97 JP97 JP97</td><td>GR06 TT06 SD09 US95 NL95 NL95 NL96 US96 OE90 PT05 FR06 FR05 NL93 DK93 DK93 DK94 OE95 DK96 PT07 FR05 FR07 BD02 BD03 DE93 UK95 DK95 DK95<</td><td>GR00 TT00 SD00 US00 US00 NL00 NL00 QS00 PT00 PT00 PT00 R00 DK93 NL03 DK94 CE04 CE05 DK95 DK95 DK95 DK95 DK95 DK95 CE04 CE05 DK95 DK9</td><td>GR00 TT06 SD00 US95 NL95 NL94 NL90 US95 NL95 PT05 PT06 R80 DK95 NL95 NL94 OE95 OE95 NL95 PT05 PT06 R80 DK95 OE94 OE95 DK96 DK97 DK96 DK92 US93 US93 US93 US95 DE96 OE95 DK98 D097 DK96 DK92 US93 US93 US93 US95 BD96 OF97 DK96 B092 DB03 OE93 UK92 PK97 PK98 DF97 DK94 B092 DB03 OE93 UK93 PK98 PK98 F89 F89<</td></th<></td></th<>	GR00 TT00 TT00 <th< td=""><td>GR06 TT06 SD96 US95 NL96 NL94 NL96 PT95 IR96 DK95 NL93 DK93 NL93 DK94 OE95 IR95 DK92 NL93 US93 US94 DD94 OE95 IR95 DK92 US93 US93 UK95 DK96 D096 IR95 PT95 IR95 PT95 IR95 DK96 D096 IR95 PT92 IR95 IR96 FR96 FR97 FR97 FN03 FN02 FR02 FR95 FR96 FR97 FR97 FN03 FN04 FR95 TR96 FR96 FR97 FR97 FN03 FR93 IR96 GR97 FR98 GR97 JP97 GR96 T197 GR97 IR97 GR97 JP97 GR96 T197 GR97 IR98 GR98 JP97 JR96 IR98 DK95 NL93 DK94 OE97</td><td>GR66 1766 SD05 SD96 US95 NL95 NL95 NL96 US95 PT95 PT96 IR96 DK93 NL93 DK94 OE94 OE95 D IR95 PT96 IR96 DK92 NL93 DK94 OE95 D IR95 DK92 US93 US94 BD96 BD97 BD97 BD97 BD97 BD97 BD97 BD97 IP97 JP97 JP97</td><td>GR06 TT06 SD09 US95 NL95 NL95 NL96 US96 OE90 PT05 FR06 FR05 NL93 DK93 DK93 DK94 OE95 DK96 PT07 FR05 FR07 BD02 BD03 DE93 UK95 DK95 DK95<</td><td>GR00 TT00 SD00 US00 US00 NL00 NL00 QS00 PT00 PT00 PT00 R00 DK93 NL03 DK94 CE04 CE05 DK95 DK95 DK95 DK95 DK95 DK95 CE04 CE05 DK95 DK9</td><td>GR00 TT06 SD00 US95 NL95 NL94 NL90 US95 NL95 PT05 PT06 R80 DK95 NL95 NL94 OE95 OE95 NL95 PT05 PT06 R80 DK95 OE94 OE95 DK96 DK97 DK96 DK92 US93 US93 US93 US95 DE96 OE95 DK98 D097 DK96 DK92 US93 US93 US93 US95 BD96 OF97 DK96 B092 DB03 OE93 UK92 PK97 PK98 DF97 DK94 B092 DB03 OE93 UK93 PK98 PK98 F89 F89<</td></th<>	GR06 TT06 SD96 US95 NL96 NL94 NL96 PT95 IR96 DK95 NL93 DK93 NL93 DK94 OE95 IR95 DK92 NL93 US93 US94 DD94 OE95 IR95 DK92 US93 US93 UK95 DK96 D096 IR95 PT95 IR95 PT95 IR95 DK96 D096 IR95 PT92 IR95 IR96 FR96 FR97 FR97 FN03 FN02 FR02 FR95 FR96 FR97 FR97 FN03 FN04 FR95 TR96 FR96 FR97 FR97 FN03 FR93 IR96 GR97 FR98 GR97 JP97 GR96 T197 GR97 IR97 GR97 JP97 GR96 T197 GR97 IR98 GR98 JP97 JR96 IR98 DK95 NL93 DK94 OE97	GR66 1766 SD05 SD96 US95 NL95 NL95 NL96 US95 PT95 PT96 IR96 DK93 NL93 DK94 OE94 OE95 D IR95 PT96 IR96 DK92 NL93 DK94 OE95 D IR95 DK92 US93 US94 BD96 BD97 BD97 BD97 BD97 BD97 BD97 BD97 IP97 JP97 JP97	GR06 TT06 SD09 US95 NL95 NL95 NL96 US96 OE90 PT05 FR06 FR05 NL93 DK93 DK93 DK94 OE95 DK96 PT07 FR05 FR07 BD02 BD03 DE93 UK95 DK95 DK95<	GR00 TT00 SD00 US00 US00 NL00 NL00 QS00 PT00 PT00 PT00 R00 DK93 NL03 DK94 CE04 CE05 DK95 DK95 DK95 DK95 DK95 DK95 CE04 CE05 DK95 DK9	GR00 TT06 SD00 US95 NL95 NL94 NL90 US95 NL95 PT05 PT06 R80 DK95 NL95 NL94 OE95 OE95 NL95 PT05 PT06 R80 DK95 OE94 OE95 DK96 DK97 DK96 DK92 US93 US93 US93 US95 DE96 OE95 DK98 D097 DK96 DK92 US93 US93 US93 US95 BD96 OF97 DK96 B092 DB03 OE93 UK92 PK97 PK98 DF97 DK94 B092 DB03 OE93 UK93 PK98 PK98 F89 F89<

IDBT/GDP1 - SOMann1bi son

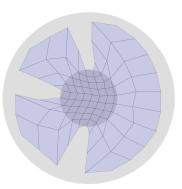
SOM: Characteristics

Positives:

- ▶ Multi-dimension scaling (often onto 2-D)
- Probability density approximation: more 'prototypes' generated for regions of higher probability densities.
- Topology preserving: any two close input patterns should match to the same neuron, or two neurons in a neighbourhood on the map.

Negatives:

- Rigid map topology
- ► Fixed number of units
- Limited online learning ability



DemoGNG results on "Fovea"



Neural Gas

Leader-Follower

- Martinetz (1993)
- Topology constraint in SOM removed ©
- Prototypes organised in the original space
- Weight updating rule: $\Delta w_i = \gamma h(k_i)(\mathbf{x} \mathbf{w}_i)$
 - k_i : neighbour rank of the prototypes
 - E.g. for winner, $k_i = 1$; second winner $k_i = 2$ etc.
 - $h(\vec{k}_i(\mathbf{x};\mathbf{w})) = e^{-k_i(\mathbf{x};\mathbf{w})/\lambda}$
- $\bullet\,$ Neighbour ranking is time-consuming $\odot\,$
- $\bullet\,$ Fixed number of neurons $\odot\,$

- Model itself is incremental; allows adaptive clustering without a known number of clusters
- \bullet Needs a similarity threshold (vigilance) or a distance threshold T
- This threshold implicitly controls the number of prototypes generated
- Procedure:
 - **①** Take initial inputs as prototypes (leaders)
 - Modify existing prototypes with new input if they are similar (followers)
 - Otherwise add the new input as a new prototype
 - Repeat Steps 2-3 on new arriving data

Leader-Follower The Idea	Leader-Follower Examples
Leader-Follower Algorithms	ART algorithms
Pseudocode # Assign first input to node 1 $\mathbf{w}_{1} \leftarrow \mathbf{x}$ # Number of nodes set as 1 K = 1 while more data are available accept new \mathbf{x} $b \leftarrow \arg \min_{i} \mathbf{x} - \mathbf{w}_{i} $ # find best match unit if $ \mathbf{x} - \mathbf{w}_{b} < T$ # if close enough, update BMU modify \mathbf{w}_{i} else # otherwise insert as new $K \leftarrow K + 1$	 An implementation of L.F. algorithm Carpenter & Grossberg (1987). Adaptive resonance theory is to model how biological neural networks coping with novel patterns. Uses a vigilance parameter A family ART1 for binary patterns ART2/ART3 for analog patterns ARTMAP as a supervised model Fuzzy ARTMAP as a fuzzy variation

 Lecture 3
 21 / 31
 Lecture 3
 22 / 31

 Leader-Follower
 Examples
 Examples
 Examples

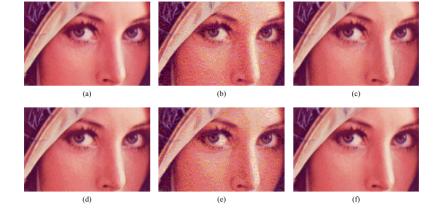
Evolving SOM

 $\mathbf{w}_K \leftarrow \mathbf{x}$

endif

- Remove rigid 'grid' topology for the nodes (borrowing GNG's idea).
- Starts from a null network.
- Nodes establish connections by searching out two nearest neighbours.
- If input is far away from any nodes (using a distance threshold, similar to GNG ☺), add a node;
- Otherwise, update nodes within the neighbourhood according to their activation (a_i) stimulated by the input:
 - $\blacktriangleright \Delta \mathbf{w}_i = h_i(\mathbf{x})(\mathbf{x} \mathbf{w}_i)$
 - Weighting, not ranking as in NG: $h_i(x) = \frac{a_i}{\sum_k a_k}$
 - The closer \mathbf{x} is to \mathbf{w}_i , the bigger a_i

Example: Image colour quantization



True-colour images quantized into $256\ {\rm colours.}$

(a) Original, (b) Median cut, (c) Octree, (d) Wu's method, (e) Local K-means, (f) ESOM.

Leader-Follower Examples Example: Online EM for background modeling

The Adaptive MoG Model

• Each pixel is modelled by a mixture of K Gaussian distributions:

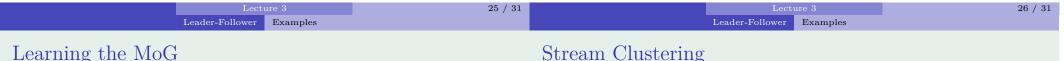
Leader-Follower

$$\eta(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{D/2} |\boldsymbol{\Sigma}_k|^{1/2}} e^{-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)}$$
$$\boldsymbol{\Sigma}_k = \sigma_k \mathbf{I}$$

Examples

- Look for Gaussians winning the most with the least variance; order models by w_i/σ_i
- The first B distributions are used as a model of the background (T is a threshold):

$$B = \operatorname{argmin}_{b} (\sum_{j=1}^{b} w_{j} > T)$$



• If Model k matched to the current pixel value at time t, update its weight (α is a learning rate):

• Problem: monitor pixel changes in a video frame and separate

• Adaptive mixture of multi-modal Gaussians per pixel.

▶ Method for updating the Gaussian parameters.

▶ Heuristic for determining the background.

▶ Probabilistic model for separating the background and foreground.

$$w_{k,t} = (1-\alpha)w_{k,t-1} + \alpha$$

• Updating the matched model:

foreground from background

• Solution (Stauffer & Grimson CVPR'99):

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho X_t$$
$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho (X_t - \mu_t)^T (X_t - \mu_t)$$
where $\rho = \alpha \eta (X_t | \mu_k, \sigma_k)$

- Stream Clustering
- Mining massive, unbounded sequences of data objects of rapid but often changeable rates.
- Example applications: Sensor networks, smart homes, Internet traffic monitoring ATM transactions ...
- Approaches: partition (ClusStream), grid-based (DStream), density-based (DenStream)
- Tools: MOA, RapidMiner etc.
- Challenges: concept drift

Leader-Follower Exam	mples	Recap	
Mini-batch k-means		Recap	
• Sculley, Web-Scale K-Means Cluster	ering, WWW'10		

• Mini-batches tend to have lower stochastic noise than individual examples in SGD

Algorithm 1 Mini-batch k-Means.1: Given: k, mini-batch size b, iterations t, data set X2: Initialize each $\mathbf{c} \in C$ with an \mathbf{x} picked randomly from X3: $\mathbf{v} \leftarrow 0$ 4: for i = 1 to t do5: $M \leftarrow b$ examples picked randomly from X

```
for \mathbf{x} \in M do
 6:
 7:
              \mathbf{d}[\mathbf{x}] \leftarrow f(C, \mathbf{x}) // Cache the center nearest to \mathbf{x}
          end for
 8:
 9:
          for \mathbf{x} \in M do
10:
              \mathbf{c} \leftarrow \mathbf{d}[\mathbf{x}]
                                            // Get cached center for this \mathbf{x}
              \mathbf{v}[\mathbf{c}] \leftarrow \mathbf{v}[\mathbf{c}] + 1 // Update per-center counts
11:
              \eta \leftarrow \frac{1}{\mathbf{v}[\mathbf{c}]}
                                           // Get per-center learning rate
12:
                                                        // Take gradient step
13:
              \mathbf{c} \leftarrow (1-\eta)\mathbf{c} + \eta \mathbf{x}
          end for
14:
15: end for
```

- The online averaging problem
- Competitive learning: online k-means
- Other online algorithms
- Leader-follower
- Density-based
- © Your algorithm?

Lecture 3 29 / 31 Lecture 3 30 / 31 Recap Further Readings

- ER3: Kaur et al., Stream clustering algorithms: a primer, in *Big* Data in Complex Systems, 105-145, 2015.
- ER4: Lühr and Mihai Lazarescu. 2009. Incremental clustering of dynamic data streams using connectivity based representative points. *Data Knowl. Eng.* 68.
- Silva et al., Data stream clustering: A survey, ACM Computing Surveys, 46:1, DOI: 10.1145/2522968.2522981.
- Cao et al., Density-based clustering over an evolving data stream with noise, SDM'06, DOI: 10.1137/1.9781611972764.29.
- DemoGNG, URL http://www.demogng.de