

# Optimization and Data Analytics

Henrik Karstoft Carl Peter Leslie Schultz Alexandros Iosifidis

@

Aarhus University, Department of Engineering



# **Optimization and Data Analytics**

#### Course outline, formal stuff

- > Prerequisite
- Lectures
- > Homework
- > Textbook
- > Blackboard



#### What is this and Goals?

#### Some definitions

- > Data Analytics is the discovery, interpretation and communication of meaningful patterns in data
- > **Optimization** is the process for *selecting the best element* (with regard to some criterion) from some set of available alternatives

#### Some course goals

- Increased mathematical skills
- A broad knowledge of the classical and modern mathematical optimization techniques
- Understand the basic Machine Learning models
- › Get experience with Data Analysis through project work
- > Understand data visualization concepts
- › Get introduced to Scientific Writing



machine learning

optimization

communication, visualization

# today: data overload



# data useless without interpretation

TIAA-C	KEF-Stoc	к гера	irea.x	usm - N	licrosc	oft Excel																						- 0
н	0	P	0	R	8	1	U		w	×	Y	- 2	AA	AB	AC	AD	AE	AF	AG	АН	A	AJ	AK	M.	AM	AN	AO	4
	4week		19 week			Grauth				Aurek		19 week			Equity Index				4week		99 week			Internations	Lequity			
in Valvo		durck	Curant			-						Cumulative								4urrk	Cumulating					Unit	Unit Takes	
Ohango	Unit Value	posted	Unit Ve			Deta 3/39/2006	Amount	Valvo 63.17	14 Change	Unit Falue	posted	Unit Telus			Date 3/31/2006	Ameunt	Value 86,9423	X Change	Unit Value	pooled	Unit Valve			Date	Amount	Telus	× Change	Unit
#REF!	#REF!	101.0434	×Ohen-	101,0939		12/24/2006		64.002	#REF!	#REF!	64.102	#REF1	64.002		12/24/2006	962,870.45	45.2116	#REF!	#REF1	45.2116	#REF!	45.21%						×:0
aner:	BREF!	191.27		111.27		12/29/2007		75.64	anar:	BREF!	75.64	14,321390	75.64		12/31/2007	392,32	99.63	anar:	BREF!	99.61	4.4127448	99.64						+
-42.2060				74,70495		12/25/2000			-29,7762000		53,0750	#05E)	\$3,0750		12/21/2000	1,002.93		-27.494412		70.90405	BREE!	70.90405						-
3.24244				27.1465		12/25/2009			B 111111111	0.000647497		******	64.656439		12/31/2009	67.313.09			0.000251054		#REF!	24.005414						+
	5,515714244	41110011		87.1465		4/1/2010	201411111	4611344	101111-4010	6.237499590		41147	64.606604		4/M2010	910,554.43	17/17/46	*********	0.379945772	+4.11241076	ancr:	04.005410						+
19.36155				17.1465		4/16/2010		66,1762		6.346025596			64.606600		4796/2010	243,539.01	16,3643		0.40210404		0.0206427	14.005419						-
8.63434				27,1465		4/23/2010				8,396733014					4/23/2010	249,415,61	22,3405		10,45413236	81.75790833								+
0.92510				07,1465		4/20/2010				5.542397505						242,950.69	86,0994		0.070170349	01.75790033	-2.592021	04.005419						т
7.06546				27,1465		5/1/2010		61,2212	-19,0624000	-1.5704615	65,31066	-6.797014	64,606609		5/7/2010	226,651,60	20.2649	-19.47742:0	0.754243031	85,41218	-6.77647	84,005419						
5.42455	-2,77953994	84.47534	2.251	84,8270857		5/14/2010	105,244,07	62,8662	-96,0076256	1,074319071	44.64030	2,6069777	64.5%143		5/94/2010	202,971.05	12,5021	-17.232343	3.56346656	84,72478	2.7111421	84.202557						т
0.06595	-7.24504002	94.36628	-4.61	10.391275		\$421/2010	177,410.72	60,1429	-20.4179693	-0.36492553	63,46172	-4.331199	62,509003		\$421/2010	222,786.10	79.0390	-20.911196	+1.02961629	93,21969	-4.666640	01.757900						
45050	4.44379247	81.8424	-1.53	13.351275		6/4/2010	175,543.19	59,4606	-21,3794219	-4.31124399	64,07102	-1.121163	62,509013		674/2010	210,727.16	77,3424	-22,40931	-2.99423981	31,00906	-1.094317	81,757900						
.45442	7 -6.00464096	00.2764	2.046	00.350275		6/11/2010	155,405.00	60,6914	-19.7628239	-2.42224400	40.07004	2.0542112	62,509003		6/19/2010	955,640.29	79,3004	-20.436440	-0.44637477	79,65006	2.5499404	01.757900				1		
.95902	7 -6.0046-4096	99.07790	. 0	10.391275		4/10/2010	195,405.11	60,6914	-19,7621239	-2.42224401	60,7721		62,509013		679975940	155,640.29	79,3004	-20.434440	-0.44637477	79,49996	0	11.757900						
.44452	4 -6,0060331	79.5277	0.004	13.351275		6/25/2010	153,374.13	59,0476	-20,8743712	-3.77009964	60,16030	-1.390312	62,509013		6/25/2010	153,534,54	74,9641	-21.510459	#REFT	70,69075	-1.437105	81,757900						
.04452	4 -6.0040330	79.74224	. 0	00.350275		7/2/2010	153,374.13	59,8476	-20.0703712	-3.77009964	60.10132		62,509003		7/2/2010	150,504.54	70.9601	-21.510454	-1.0777616	70.531025	0	01.757900						
.33445	7 -5.26416911	99.324	0.709	13.1640531		7/9/2010	152,700.76	54,5333	-21,2930921	-4.21422134	60,12226	-0.525167	62.210177		T/M2010	153,475.61	72,0666	-21.510454	-1.0777616	70.73025	-0.129141	11.757900						
14527	1.76342731	81.06362	3.445	43.2147071		7/24/2010	154,220.73	62,0171	-18.0101798	-9.21094534	69.3874	4.1721100	42.241304		TV24/2010	151,341.21	69,9526	-21.442745	-2.00517172	76,4276	3.7352722	81,473942						
.21003	0.104302394	92.12446	1.904	10.34321		9/9/2019	160,000.00	62.2915	-17.4301544	0.161500750	40.70112	0.4537450	62.26396		97975910	199,999.03	91,244	·10.757424	1.655201040	70.04635	0.3227054	01.420064						
.52197	4.21010716			13.2430125		9/13/2019			-21.1160694	-4,06094112	60.67274	-4,223697	62,101969		875342010	103.21	77,9060	-11,495115	1,913329916	79.615325	-4,809150	11.425173						
.19664				83.9500176		8/20/2010			-21,0749603			0.053120			8/20/2010	102.67			-2.10534244		-0.451154							
.369640	-5.30412411	92,64024	-0.64	13.0201333		9/20/2019	152,003.64	\$9,0511	-21,4313155	-5.05940744	60.54656	-1.019112	61.790617		9/20/2010	264.37	77.2545	-22.121200	-2.55345117	79,460725	-9.413044	11.000254						
24,3144				13.9137		9/3/2010			1007179111	10001111001		3,4204721			9/3/2010	276.6			-3.02457023		3.4499934							$\perp$
	9.46300506			02.7924241		1/10/2010			-10.6561276			0.2633367			9/90/2010	454.69			0.746006925		0.2767244							┸
	9.732257403			12.5453105		4V1TV2010				1.102764578		2.2031712	*******		9/17/2010	420.09			1,029602060		1.6426217	441161545						+
	7 3.074467027			12.4610316		9/24/2010				3.714904543		2.5121551			9/24/2010	544.79	10.9534		2.615152641	79,949425		79,761417						╄
	3.590152055			92,4991195		10/1/2010	,			3.402740641					10/5/2010	\$67.17	13.6000		4.002255473		0.0663049							₽
	5.504034343	7112777	-			10/10/2010	1111111111111		1015-5-65-6	5.335219702	401000	1,7901256	***********		1070/2010	711,61	77.74.17	-14.122743	41101111111		1.6900135	1110000						+
	6.882750945	*******		13.3060361		10/15/2010	***************************************			7.548435004		2.9913059	********		10/15/2010	798.87	85,8895		6.725497000		1.0205410	111111111111		90/15/2010	101,312.32			1
.04355				14.3544667		10/24/2010				0.466510177		-	42.103770		10/24/2010	191.23		BREFT	#REF1	14.04003333	#REF!	80.913747		10/24/2010				
1,24244	10.464000035			85.1343411 85.4242554		19/30/2019				12.43335799		3.4572090	62,010400		10/30/2010	1,024.45		-13.2551%	12,55074674	07.74206667		10,934159		10/34/2010	105,779.10		4.32441414	
.054733				16,5946444		19/12/2010				10.30435042					11/12/2010	1,024.45			12.99074674			01.544214		19/2/2010	103,631.52		4.32441414	
.05473				17,2450333		1915/2010	116,157.44			19.30435042			64,100791		11/19/2010	1,09.97	88,0239		10,49394959	88,4572		12,100554		19/15/2010	103,631,62		2,20009964	
40640				87.9699009		11/26/2010			-6.07043092			2.4762536			19/26/2010	1,302.00	90.2277		10,49394959		2.5034302			11/26/2010			5.05401922	_
5.50417				22.53506.11		12/3/2010				54.50077494		0.7495652			12/3/2010	1,320,36			13.26031134		13499312			12/2/2010			3.496.01413	
35040		7916000	1010	89.0209667		12/10/2010			-5.91300222		43111	0.279561	40.00		12/10/2010	1,450.99	1114575	Provide a season	14.70020105		0.3432408			52/18/2010			3.04014635	-
6.42167				19.7619667		12/17/2010			-5.30206621			0.5642001			12/17/2010	106,070.40		-7.4530490		90.360625		15.334524		12/17/2010	157.00		4.49007431	
1.4216.7				99.5401272		12/24/2010				15,86543619			67,020309		12/24/2010	105,870,48			16.50394155	91,957575	0			12/24/2010	157.03		4.45007435	
.90725				91,3710333		12/21/2010			-5.44936542		71,35002	-0.074212	67.70		12/21/2010	904,951,53	92,7440	10,1014501		92,203575	-0.049239			52/21/2010	150.0	23,0132	5.6210091	
.62557	0.315544341	45.34574	0.3%	92.0219667		1/2/2011	145,241.15	72,421	-4.24642046	1272245343	71.64994	1,2722653	60.327534		1/7/2011	100,013.50	92,7479	42.0014345	1.050469690	42,53545	1.0504647	17.112425		1/7/2011	157.02	22,7572	4.44640261	2
.93567	1.132+35-402	15.84314	0.155	12,7101270		1/14/2011		72,5431	-4.04782126			0.1045301	61,940505		1/14/2011	101,433.21	94.0412	12,6690000	1,374141571	93,034325	0.31241403	44,663435		1/14/2011		23,1734	6.35751134	4 1
73200	1.345094043	96.21990	0.23	10.3342167		1/21/2011			-4.06702126	1,461160571	72,12014		69,470339		1/21/2011	100,633.20		92.9640404	1.374641571	10.3417	0	09.303724		1/21/2011	159.9		6.35751134	
.93107	1.132435402	44.99153	-0.23	93,8307722		1/21/2011	145,729.90	72,9431	-4.06702126	1,461161571	72.32700		69,425139		1/24/2011	191,633.21	94.0412	12.1140434	1,374641571	93,649075		40.000453		1/21/2011	159.9	23,1734	6.35751134	4 1
.62494	3.012500341	97.20194	2.45	94,4473996		2/4/2011	150,461.73	74.9227	-0.94838777	4.710473223	73.000	3.2517905	70.512422		2/4/2011	111,646.07	96.4510	10.1400315	4.100019179	10.967675	2.7760173	10.614324		2/4/2011	162.29	23,5214	7.95377357	1 2
.96054	4.574950714	99.09900	0.724	95.0061722		2/11/2011	153,447.70	74.3523	0.94169751	6.759407750	73.79216	1,9000999	71,116417		2/11/2011	113,519.33	90.2203	91.0445049	5.099036205	94,69315	1.6414993	91,291441		2/11/2011	961.17	23.3543	7.2052901	1
1.61750	7 6.144240746	99,05014	1.501	45.5444411		2/10/2011	154,950.10	77,0999	1,93006346	7.404737542	74.70022	0.9791454	71,605461		2/10/2011	114,878.92	99,3536	91.4295459	7.901300669	95,743125	1.1353404	92,024141		2710/2011	163,74	23,7240	8,90795182	2
0.61750	7 6.564240746	99.97370	0	96.1727309		2/25/2011	154,950.90	77.0999	1.93006346	7.004737542	75.60750		72.296709		2/25/2011	914,070.92	99.3536	12.5414464	7.991300669	97.071225	0	12,712151		2/25/2011	163.74	23.7243	0.90795982	2
	5.054522451			96,5295270		3/4/2011				5.005223461		-1,700547			3/4/2011	113,257.72			5.400007394	90.399325		43.352553		3/4/2011				
	5.854522491			96,8783167		341145011				5.005223461			72,860767		3/11/2011	113,257.72			5,400007394	90,690975		43.100124		34445644				
	5.054522451			97.3410033		3411/2011				5.005223464			73.254322		3/10/2011	913,257.72			5.400007394	90,602		94.411053		3/10/2011	161.96			
11.11136				97,70005		345545011	-			4.129121291	75.4507				3/25/2011	113,072.71	97,4960		5.054594011	90.2262		94,919112		3/25/2011	194.19		3.00730962	
	5.300150112			98.9157444		4/1/2011	153,743.49			6.491313108		1,7757225	73.439522		4/1/2011	914,977.40	99,0984	99.5215411	6.026207070		1.6144016			4/1/2011	160.03		6.97481221	
433799	5.640014611	100.45214	0.340	90.5157270		4/11/2011	152,992.07	75,0649	0.29732946	6.077909950	75.719	-0.57494	74.20799		4/0/2011	154,705.14	90,7494	92,950%	6.44560053	90,064	-0.356242	45.165706		4/0/2011	162.90	23,5624	9.56424111	# 7

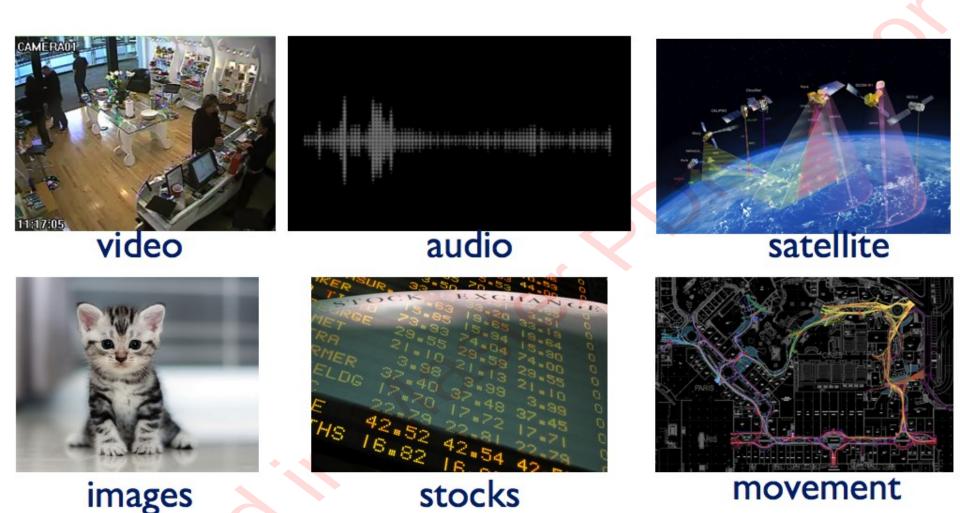
anything interesting happening here?



"interpretation" = analysis
= finding patterns

what are examples of data?

what are examples of patterns?



computers can help!
with lots of examples computers can build "pattern finder"

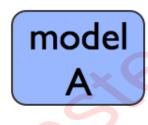


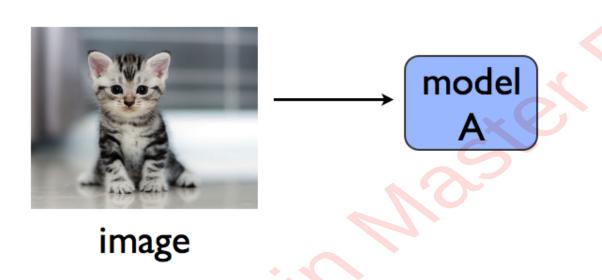
machine learning

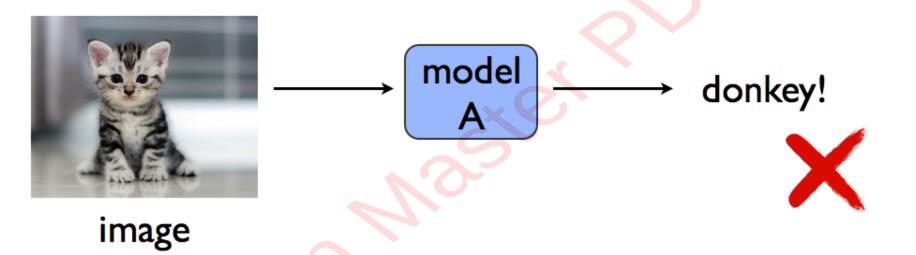
optimization

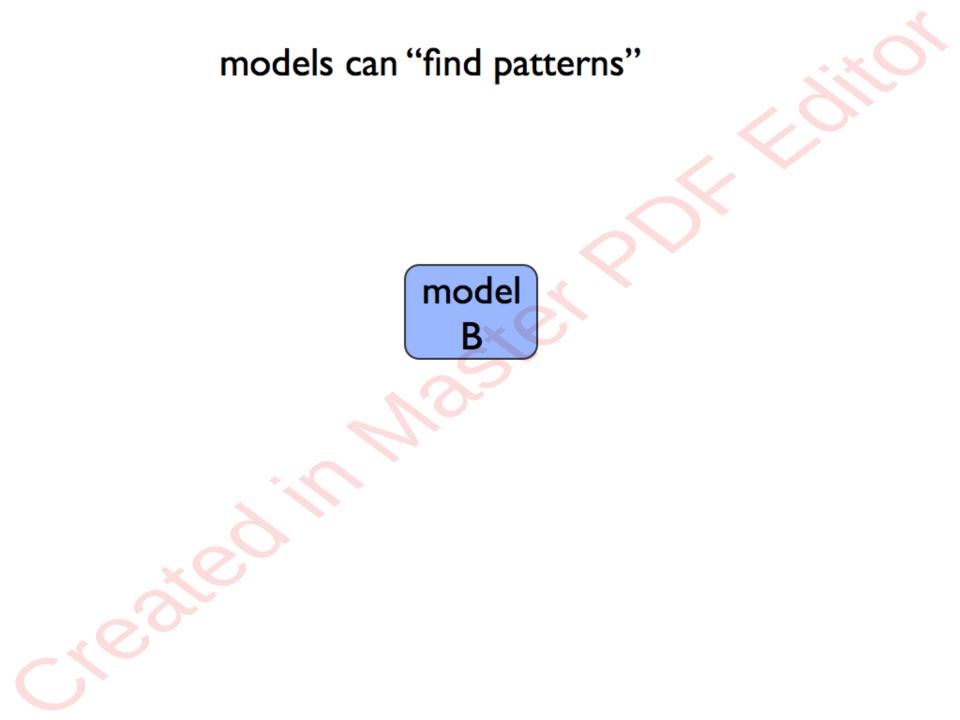
communication, visualization

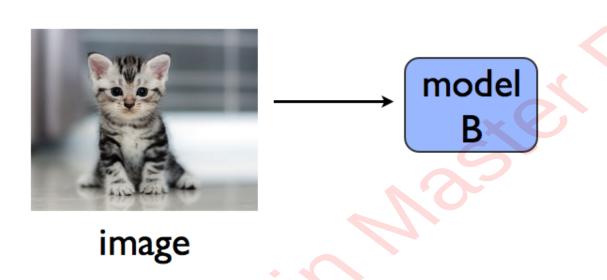
machine learning: building models from lots of examples

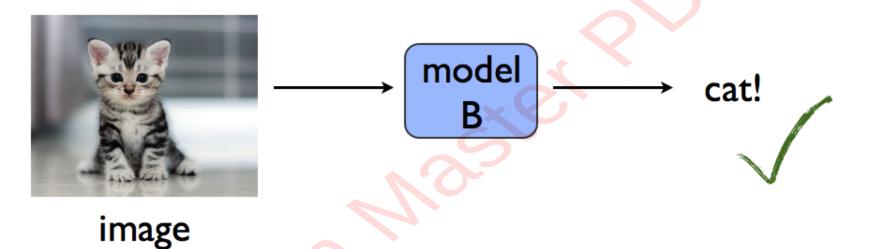














# machine learning

optimization

# communication, visualization



### imagine set of all possible models

model

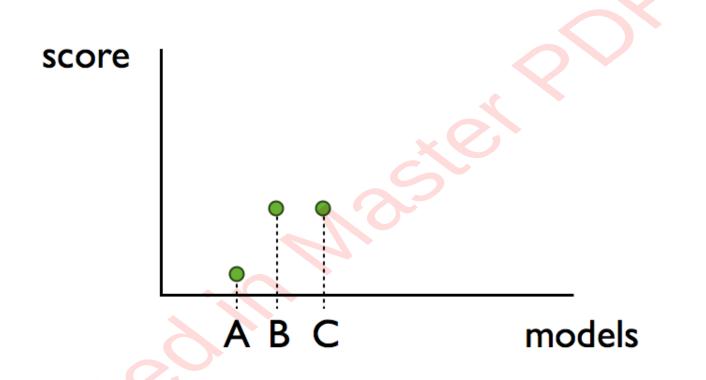
model B

model C

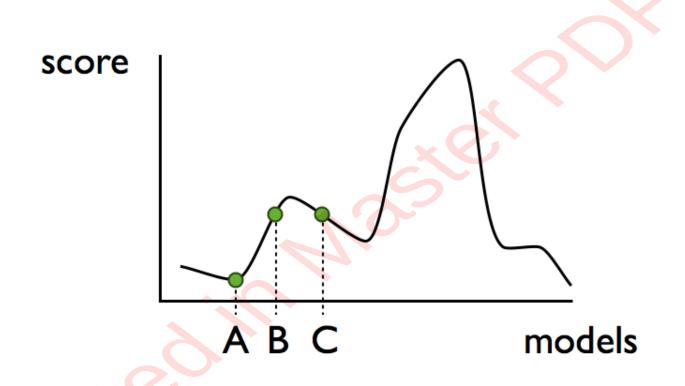
• • •

let's give each model a **score** 

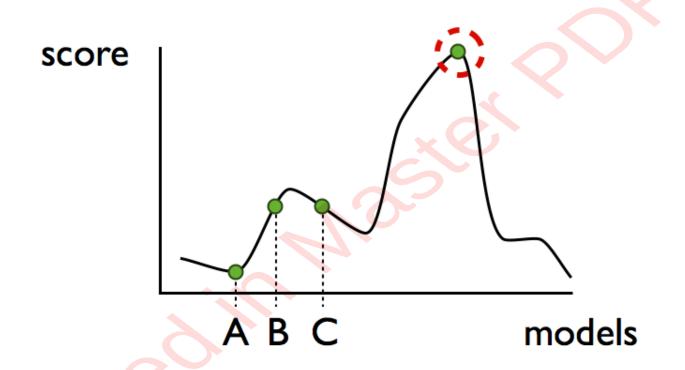
#### we want **best model** we can find



#### we want **best model** we can find

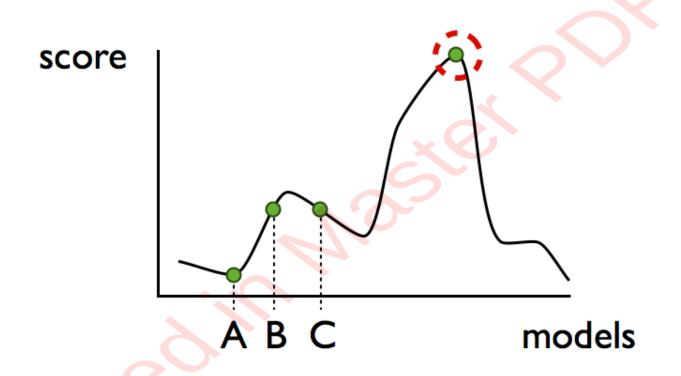


#### we want best model we can find



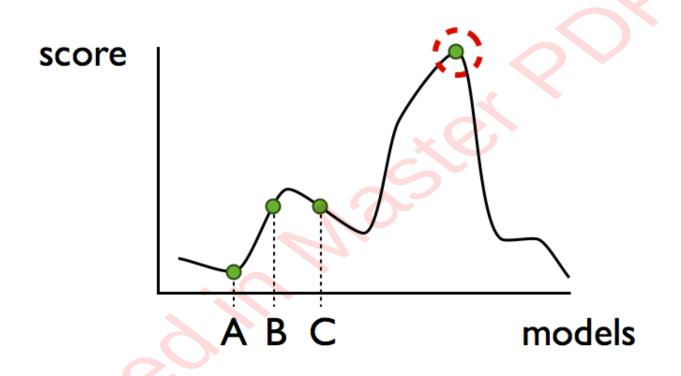
optimization: find model that maximises score

# how did I explain this concept to you?



# how did I explain this concept to you?

I visualized set of models as a chart



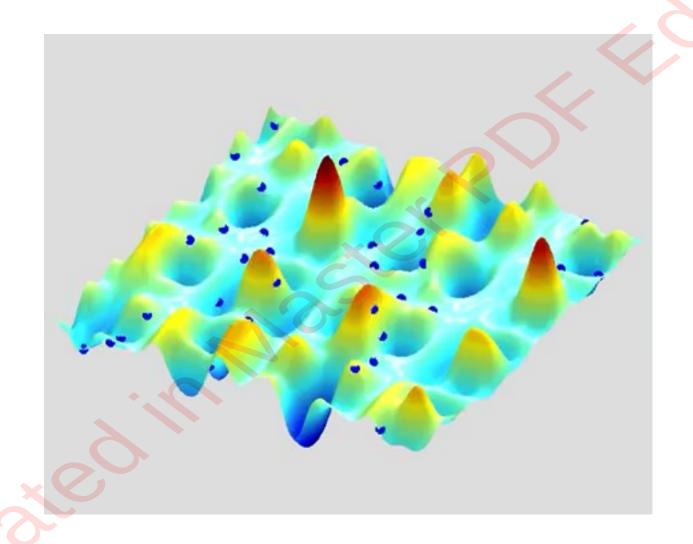


machine learning

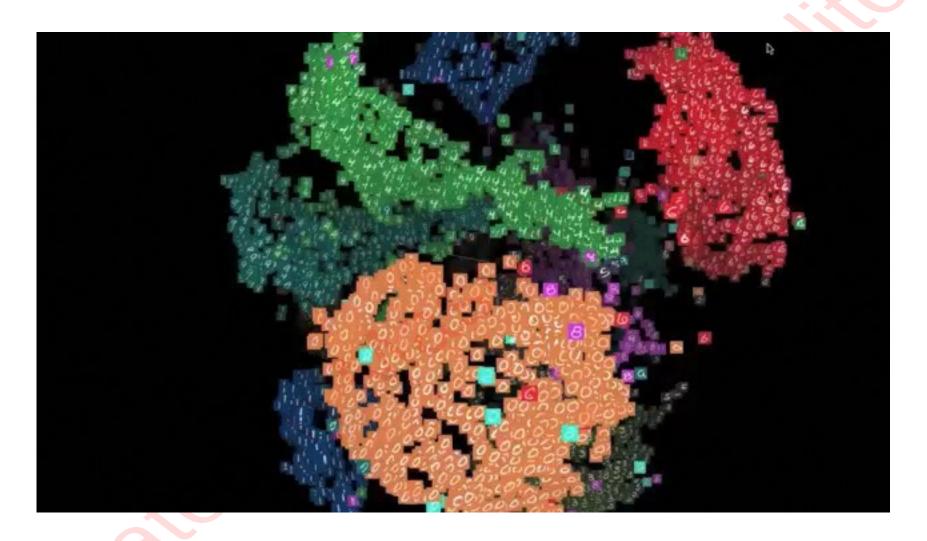
optimization

communication, visualization

# tools for thinking, understanding, communicating



# tools for thinking, understanding, communicating



learnt classification (from Google)



part 3

machine learning

part 1

optimization

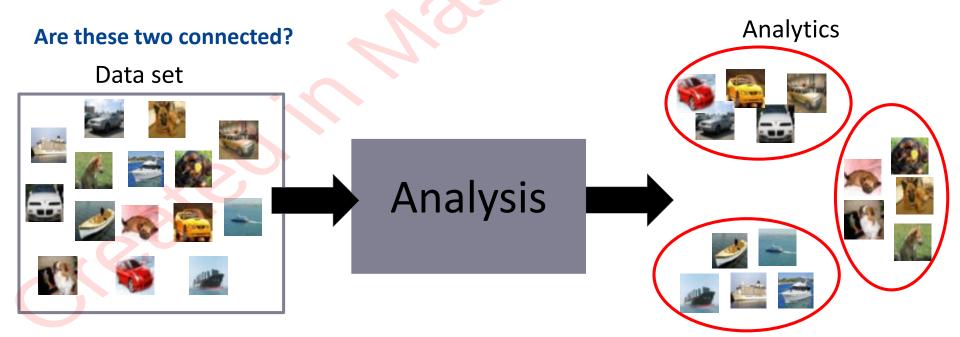
part 2 communication, visualization



#### What is this?

#### Some definitions

- > Data Analytics is the discovery, interpretation and communication of meaningful patterns in data
- > **Optimization** is the process for *selecting the best element* (with regard to some criterion) from some set of available alternatives

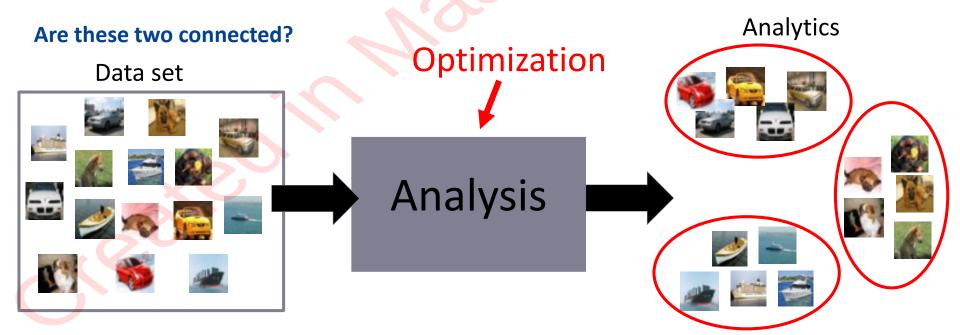




#### What is this?

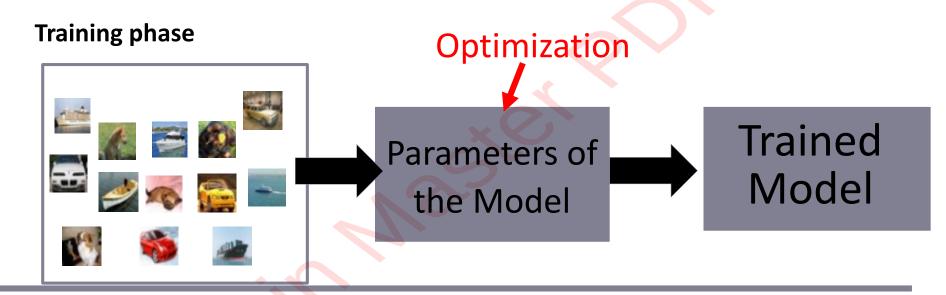
#### Some definitions

- > Data Analytics is the discovery, interpretation and communication of meaningful patterns in data
- > **Optimization** is the process for *selecting the best element* (with regard to some criterion) from some set of available alternatives





# Optimization for parameter estimation

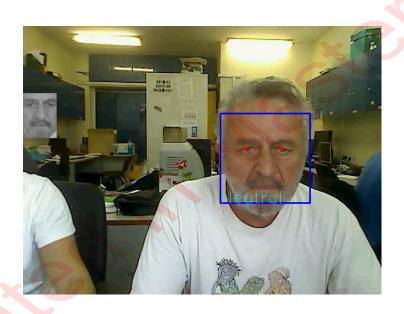


#### **Test phase/Evaluation/Online process**



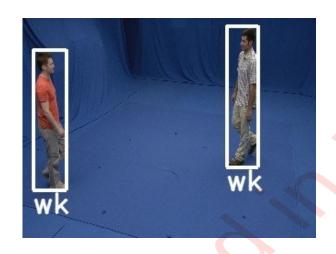


#### **Face Analysis**





#### **Action recognition/analysis**



A Hough Transform-Based Voting Framework for Action Recognition

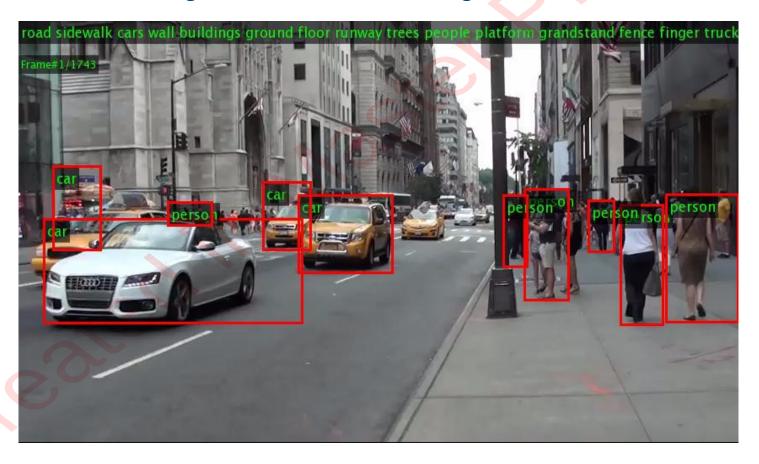
**CVPR 2010** 

Angela Yao, Juergen Gall and Luc Van Gool

Source: https://www.youtube.com/watch?v=sgcVOOZI8bw

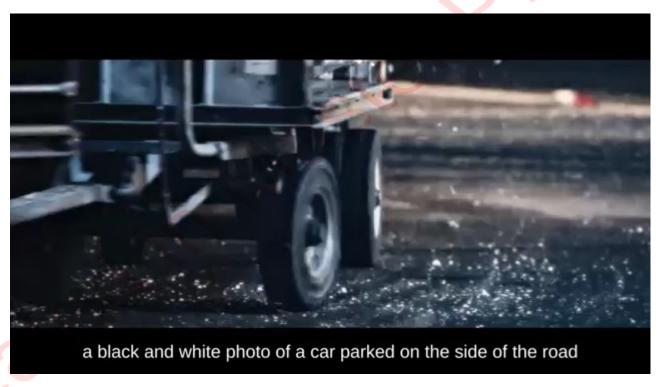


#### **Object localization/recognition, scene understanding**





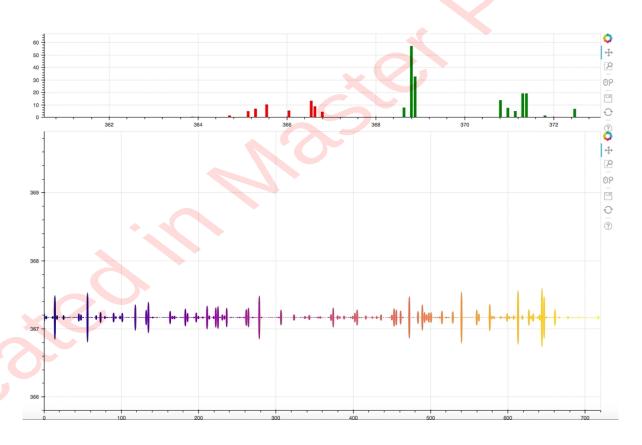
#### Image/video captioning



Source: <a href="https://www.youtube.com/watch?v=FmSsek5luHk">https://www.youtube.com/watch?v=FmSsek5luHk</a>



#### **Time-series forecasting**



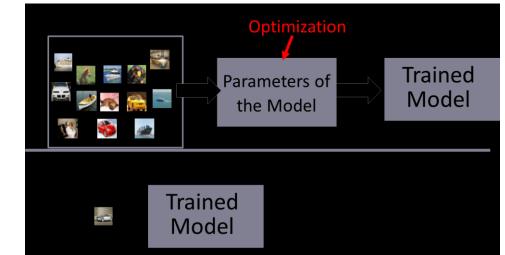


#### **Everyday examples**

- Natural language processing
- Affective computing
- Medical diagnosis
- > Robotics
- > Speech and handwritten recognition
- > Translation
- **)** ..

# Why should I learn about Optimization?

Machine learning models are in fact formed by a set of parameters which need to be 'optimized', i.e. to find good values for these parameters in order to achieve our goal.





# Why should I learn about Optimization?

Machine learning models are in fact formed by a set of parameters which need to be 'optimized', i.e. to find good values for these parameters in order to achieve our goal.

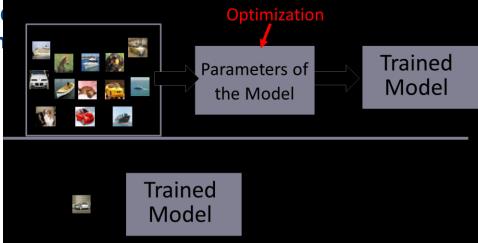
These parameters are usually stored in vectors and matrices. For example a linear model has the form:

$$\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i$$

W can be random values if x is high dimensions and y is low dimensions. For classifier, W cannot be random.

where the parameters of the linear model are stored in the matrix W.

In this case, the goal of optimization is to fine an 'optimal W', which means a W that, given the input  $x_i$  can generate a good (desired)  $y_i$ .





## **Optimization topics**

#### In the course, we will cover the following topics:

- Linear programming
- > Unconstrained optimization
- Nonlinear constrained optimization
- Solving linear equations
- > Unconstrained optimization: 1D search methods
- Unconstrained optimization: Gradient methods
- Global search methods



# Mathematical optimization

#### A large number of optimization problems can be written in the following form:

```
minimize f_0(x)
subject to f_i(x) \le b_i, where i = 1, ..., m
```

Contraints are used to define a solution that is practical.

- $x = (x_1, ..., x_n)$ , Optimization variables, or decision variables
- >  $f_0: \mathbb{R}^n \to \mathbb{R}$ , objective function
- $f_i(x) R^n \to R$ , where i = 1, ..., m: constraint functions
- $x^*$  optimal solution, smallest value of  $f_0$  among  $x = (x_1, ..., x_n)$ , satisfying the constraint



# Mathematical optimization

A large number of optimization problems can be written in the following form:

- $x = (x_1, ..., x_n)$ , Optimization variables, or decision variables
- >  $f_0: \mathbb{R}^n \to \mathbb{R}$ , objective function
- $f_i(x) R^n \to R$ , where i = 1, ..., m: constraint functions
- $x^*$  optimal solution, smallest value of  $f_0$  among  $x = (x_1, ..., x_n)$ , satisfying the constraint

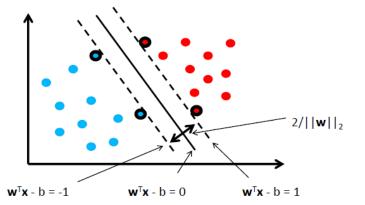


# Mathematical optimization

A large number of optimization problems can be written in the following form:

- $x = (x_1, ..., x_n)$ , Optimization variables, or decision variables
- >  $f_0: \mathbb{R}^n \to \mathbb{R}$ , objective function
- $f_i(x) R^n \to R$ , where i = 1, ..., m: constraint functions
- $x^*$  optimal solution, smallest value of  $f_0$  among  $x = (x_1, ..., x_n)$ , satisfying the constraint

### **Example of ML model: SVM**





# **Linear Programming**

#### Linear programming problems have the following form

minimize  $c^T x$ 

subject to  $a_i^T x \leq b_i \ i = 1, ..., m$ 

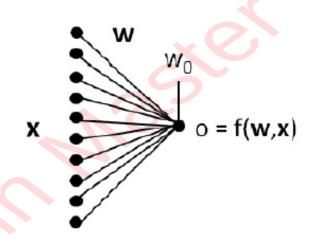
if we had c^T c then we would not be able to use linear programming

- > *no* analytic solution
- > reliable and efficient algorithms and software
- > computational time proportional to  $n^2m$ , where  $A \in \mathbb{R}^{m \times n}$ , in some case less than this
- > mature technology



# **Linear Programming**

**Example of ML model: Perceptron** 





## Unconstrained optimization

Unconstrained optimization is used in cases where there are no constraints on the model's parameters

For example a linear unconstrained problem has the form

minimize  $c^T x$ 



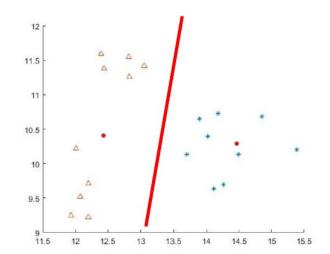
# Unconstrained optimization

Unconstrained optimization is used in cases where there are no constraints on the model's parameters

For example a linear unconstrained problem has the form

minimize  $c^T x$ 

**Example of ML model: Linear classifiers** 





# Solving linear equations

A system of linear equations has the form:

$$Ax = b$$

**Example of ML model: Linear/nonlinear Regression** 

$$\mathcal{J}_s = \sum_{i=1}^N (\mathbf{w}^T \mathbf{x}_i - b_i)^2 = \|\mathbf{X}^T \mathbf{w} - \mathbf{b}\|_2^2$$

### 1D search methods

Very often, optimization cannot be obtained using a closed form solution. Then we need to follow an iterative process which at every step gives a set of parameters which are better than before

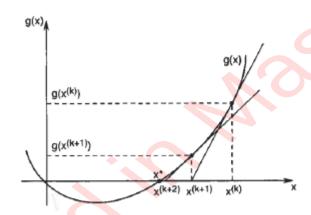


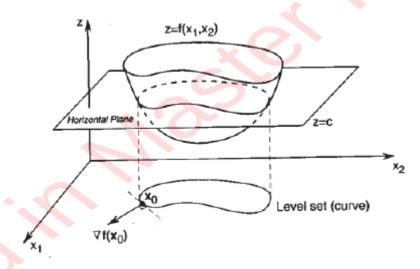
Figure 7.8 Newton's method of tangents.

$$x^{(k+1)} = x^{(k)} - \frac{g(x^{(k)})}{g'(x^{(k)})}$$



### **Gradient methods**

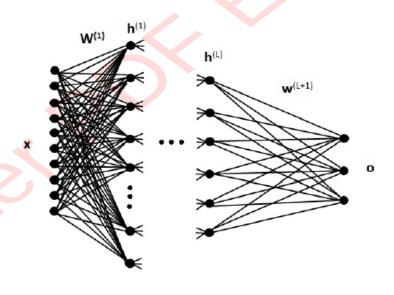
When the number of parameters is higher than one, the iterative optimization process uses the gradient of the function w.r.t. the parameters for updating:

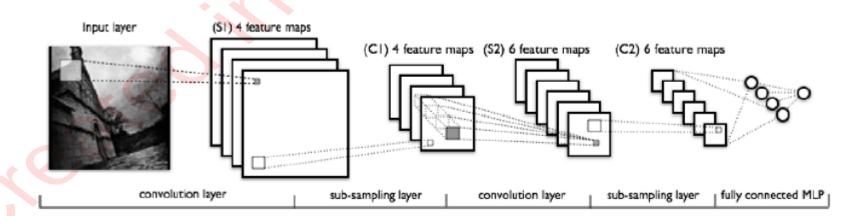




### **Gradient methods**

**Example of ML model: Neural networks** 

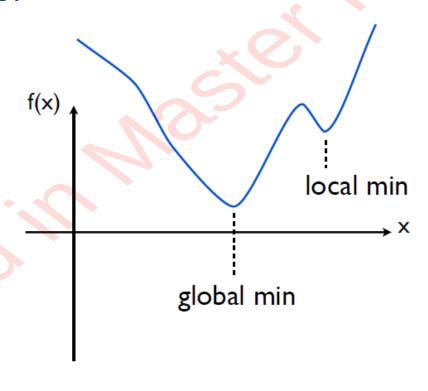






## Global search methods

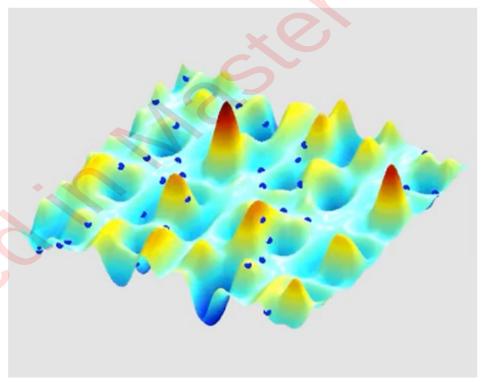
One of the problems of iterative optimization methods is that their performance depends on the starting point:





## Global search methods

One of the problems of iterative optimization methods is that their performance depends on the starting point:



https://www.youtube.com/watch?v=VAASmSSsFaY