

examples of computer generated disease modeling in...

alessandro verri

disi – università degli studi di genova

via dodecaneso 35, 16146 genova



<http://slipguru.disi.unige.it>



the group



16/01/2008

festival della scienza 2007



summary

- two problems: teach a computer to
 - identify the inflamed synovia
 - find a few needles in a genomic haystack
- issue we have to address
- our guide
- our strategy
- conclusion



where is the inflamed synovia?

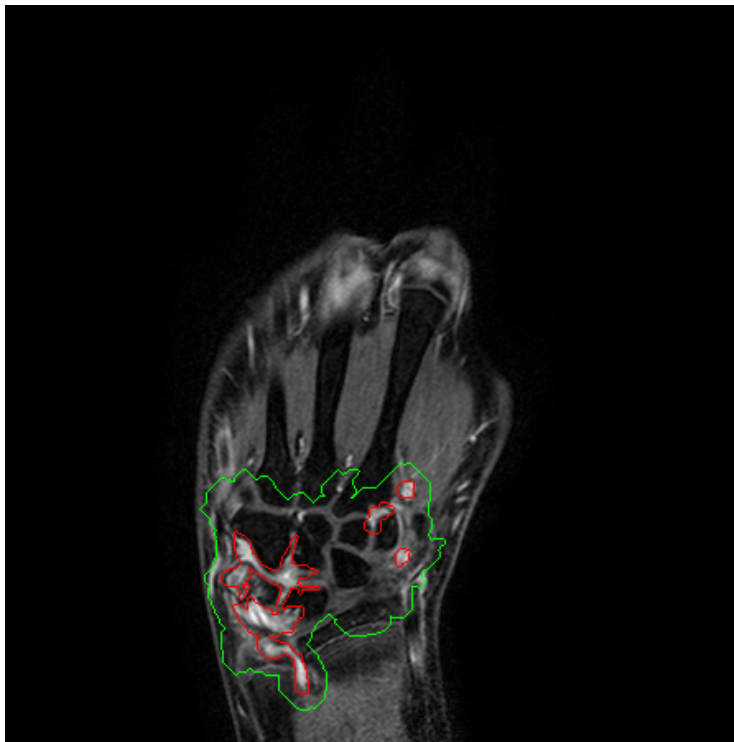


the volume of the inflamed synovia could be an early predictor for the severity of the disease

difficult and costly if assessed by hand (and operator dependent)



we ask the doctor for help a few times



two lists of examples:
the first consists of small volumes of inflamed synovia (positive examples, within the **read** contours)
the second of small volumes just outside (negative examples)

problems: small volumes of the two lists look alike, image intensities vary with the exam and the machine, it's hard to teach anatomy to a pc,...



what a computer really sees...



221	234	243	255	230	209	213	15 6	17	20
233	213	224	245	232	201	154	34		16
209	233	212	212	212	90	34	34	45	10
205	200	212	245	84	67	65	43	18	9
221	223	203	156	67	64	18	15	13	20
234	223	225	226	60	55	44	15	10	2
190	225	217	236	57	65	145	15 6	178	153
249	217	204	222	78	75	101	78	144	123
237	234	217	245	73	74	22	90	144	112



looking for needles in a genomic haystack

- in another case, along with clinical info and imaging, we have DNA micro-array - i.e., information on human gene expression (~ 30.000) from brain tumor cells
- starting from a gene regulatory network known only partially, the problem is to find the group of genes the expression of which make it possible to identify homogeneous groups of patients



which are the right features?

- we have two groups of people we can look at and ask questions to
- we know they can be distinguished for some features but we don't know which

group 1: alice, camilla, paola

group 2: charlie, john, luca, lorenzo



which are the right features?

- we have two groups of people we can look at and ask questions to
- we know they can be distinguished for some features but we don't know which

group 1:



group 2:





which are the right features?

- ❑ what if the features are as many as 30000?
- ❑ If the feature number is about 30 the number of possible combinations is about

1.000.000.000.000.000.000.000.000.000.000.
000.000.000.000.000.000.000.000.000.000.
000.000.000.000.000.000.000.000.000.000.
000.000.000.000.000.000.000.000.000.000

a grid consisting a billion of computers a billion times faster of the fastest existing computer today and working for a billion of years in a billion of planets won't even scratch the surface of the problem

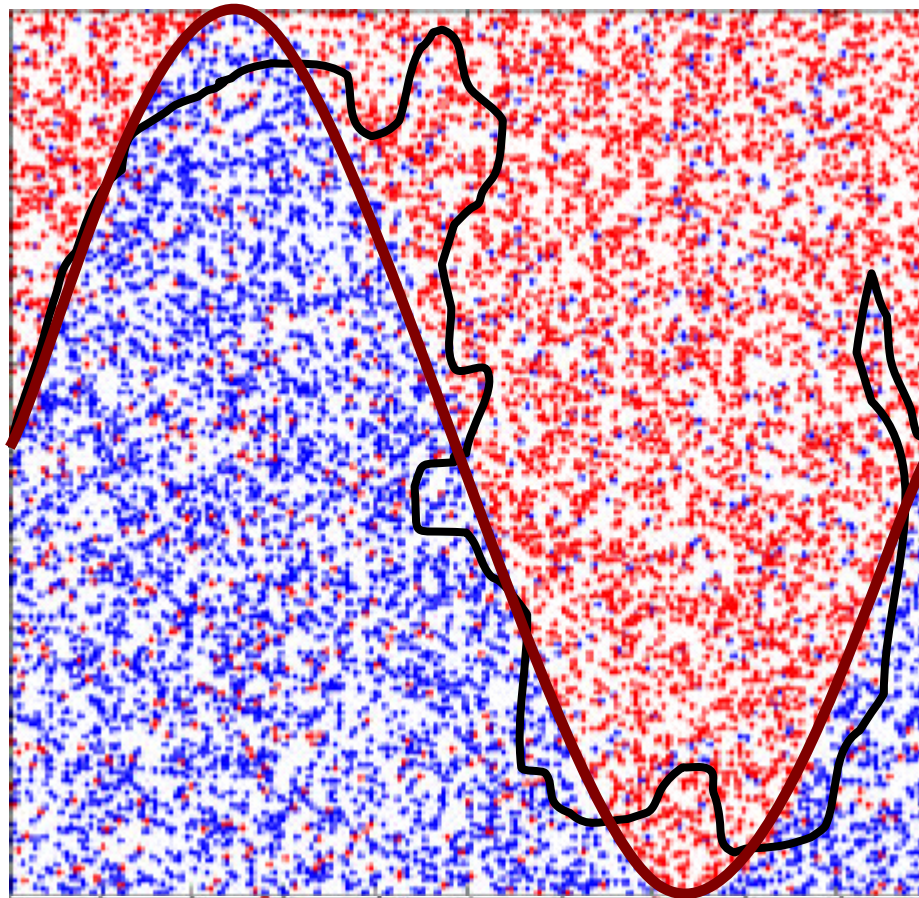


inference is pretty much like science

- guess a law (*starting from part of the available data*)
- compute the consequences (*learn how to make predictions*)
- verify the agreement with experiments (*compare the obtained predictions against the remaining available data*)



if dimensions were only two...





as for the synovia we are already there





as for genomics we are ready to go

- ❑ a sophisticated math is needed for finding a solution which is the best trade-off between staying close to the available data and obtain good generalization rates (functional analysis, probability theory, calculus and optimization)
- ❑ the key is to look for **smooth** (not too complex) solutions and **weigh** instead of **counting** features
- ❑ permutation tests come to the rescue



conclusion

- research does not have collateral effects and is good for the brain and the rest of the body too
- research applied to the biomedical world even more