

SCILA

Machine Learning in Market Surveillance

Måns Tillman
Jens Wiren

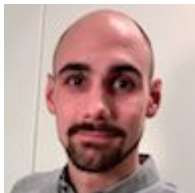
info@scila.se
jobs@scila.se



Market Surveillance

*“**Market surveillance** plays a crucial role in the field of consumer product safety as even the best rules are worth little if they are not enforced properly.”*

- European Commission



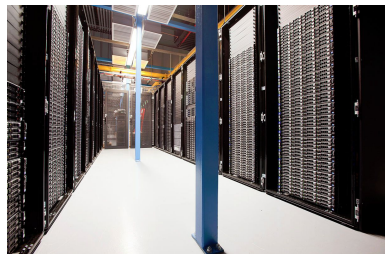
Måns Tillman, mans.tillman@scila.se, <http://www.scila.se>



Jens Wirén, jens.wiren@scila.se, <http://www.scila.se>

SCILA

Market Surveillance



Market messages



SCILA

ALERT
Insider
Trading

ALERT
Spoofing

ALERT
Circular
Trading

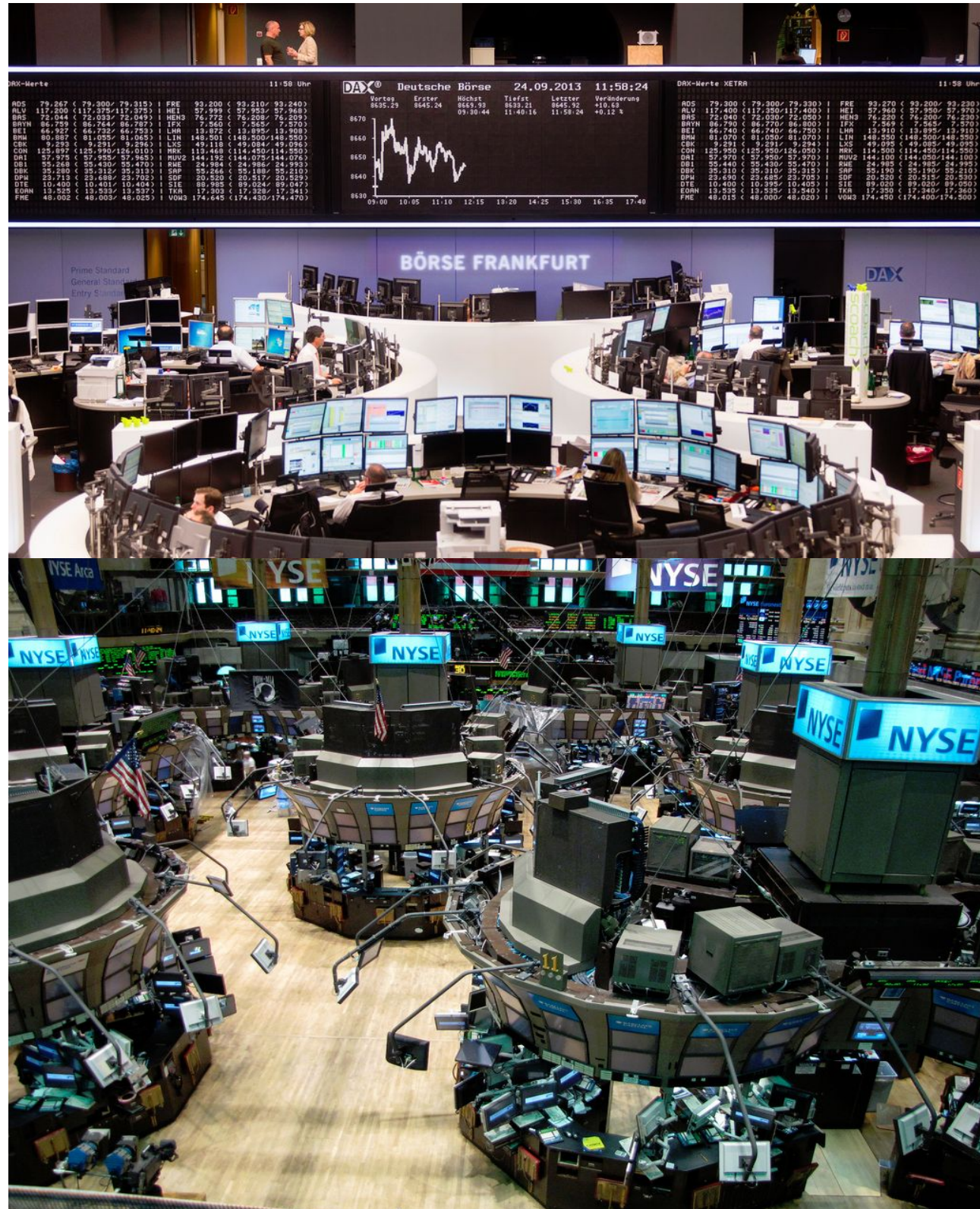


Triggered alerts



SCILA

A World of BIG DATA



Deutsche Börse

– 5 billion messages per day

New York Stock Exchange - NYSE

– 40 billion messages per day

– up to 3 million per second

SCILA

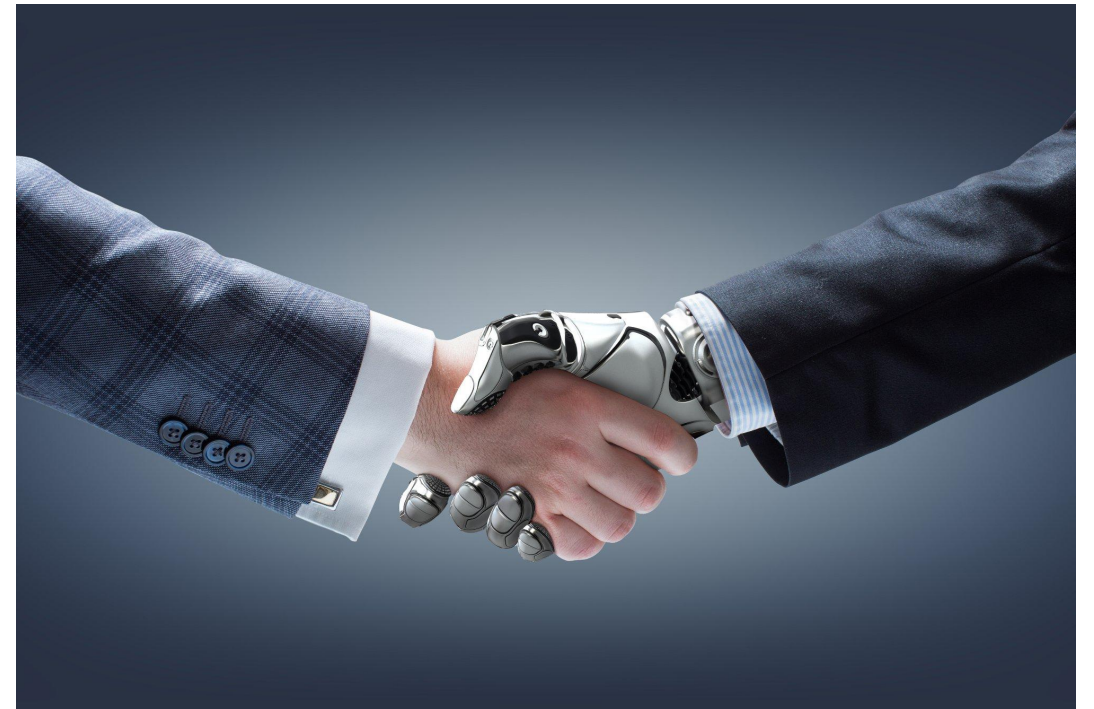
Drowning in Data: HowTo

- Requirement to trigger 'alerts' on all suspicious trading activity, enforced by regulation
- For humans to process the output
- Trading automation has increased message rate dramatically
- The work increases proportionally



Drowning in Data: How **Not** To

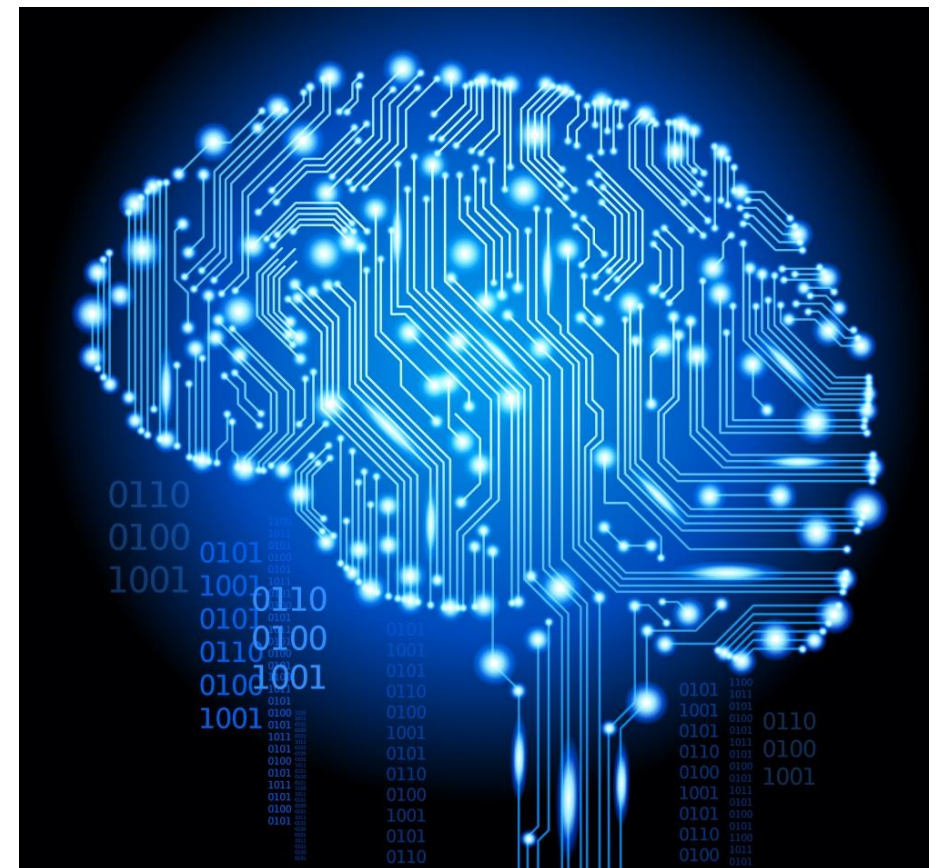
- Requirement to trigger 'alerts' on all suspicious trading activity, enforced by regulation
- For humans to process the output
- Trading automation has increased message rate dramatically
- The work increases proportionally
- Those who want to stay afloat have asked for help



SCILA

Machine Learning to the Rescue

- Machine Learning: new phenomenon in fintech
- Classifying rather than removing 'false positives' is preferred
- Every alert needs to be handled
 - => Lots of labelled data
 - => Text book case for supervised learning!



SCILA

Accuracy measures

Two primary measures

- Recall: “What percentage of all *actual* ‘positives’ did we capture?”
- Precision: “What percentage of our classified ‘positives’, are *actual* ‘positives’?”

Stay positive

Two primary measures

- Recall: “What percentage of all *actual* ‘positives’ did we capture?”
- Precision: “What percentage of our classified ‘positives’, are *actual* ‘positives’?”

“Each triggered alert can indicate a serious criminal offense”



Frameworks of the Trade

- TensorFlow - Google's AI monster, soon available in Java
- DL4J - Advanced neural network builder for commercial use
- Spark - Apaches general ML framework
- SMILE - General ML framework



DEEPLEARNING4J



SCILA

Tools of the Trade

- Support Vector Machines - Transparent but requires domain knowledge
- Neural Networks - Jack-of-all-trades
 - RBF Networks - Nonlinear (shallow) neural networks
 - Convolutional Networks - great at finding cats on youtube! (Also Googles related image search)
- Bayesian Classifiers - Useful but requires huge data sets



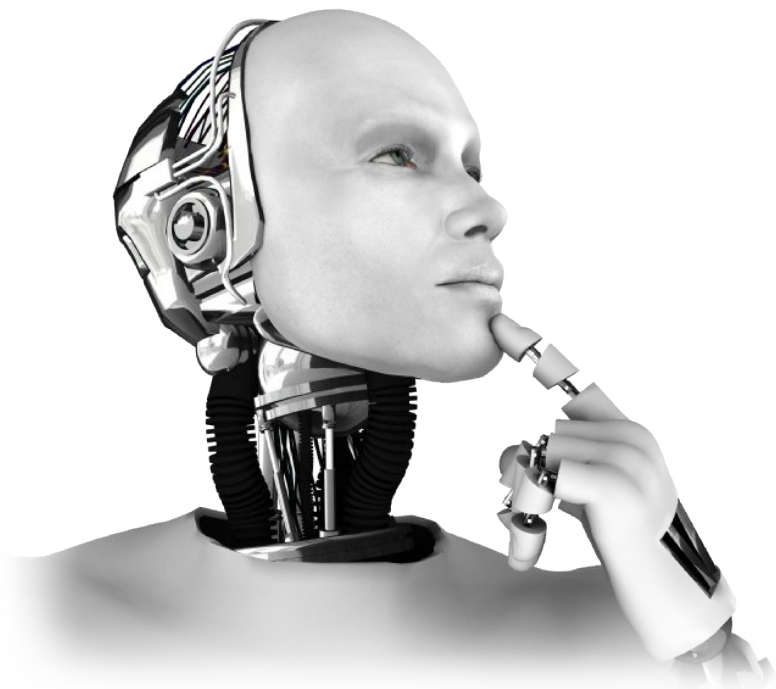
SVM vs Neural nets

SVM

- + Transparent algorithm
- + Only two hyper-parameters
- + Uses domain knowledge
 - Requires domain knowledge
 - Must specify features
 - Computationally expensive

Neural Networks

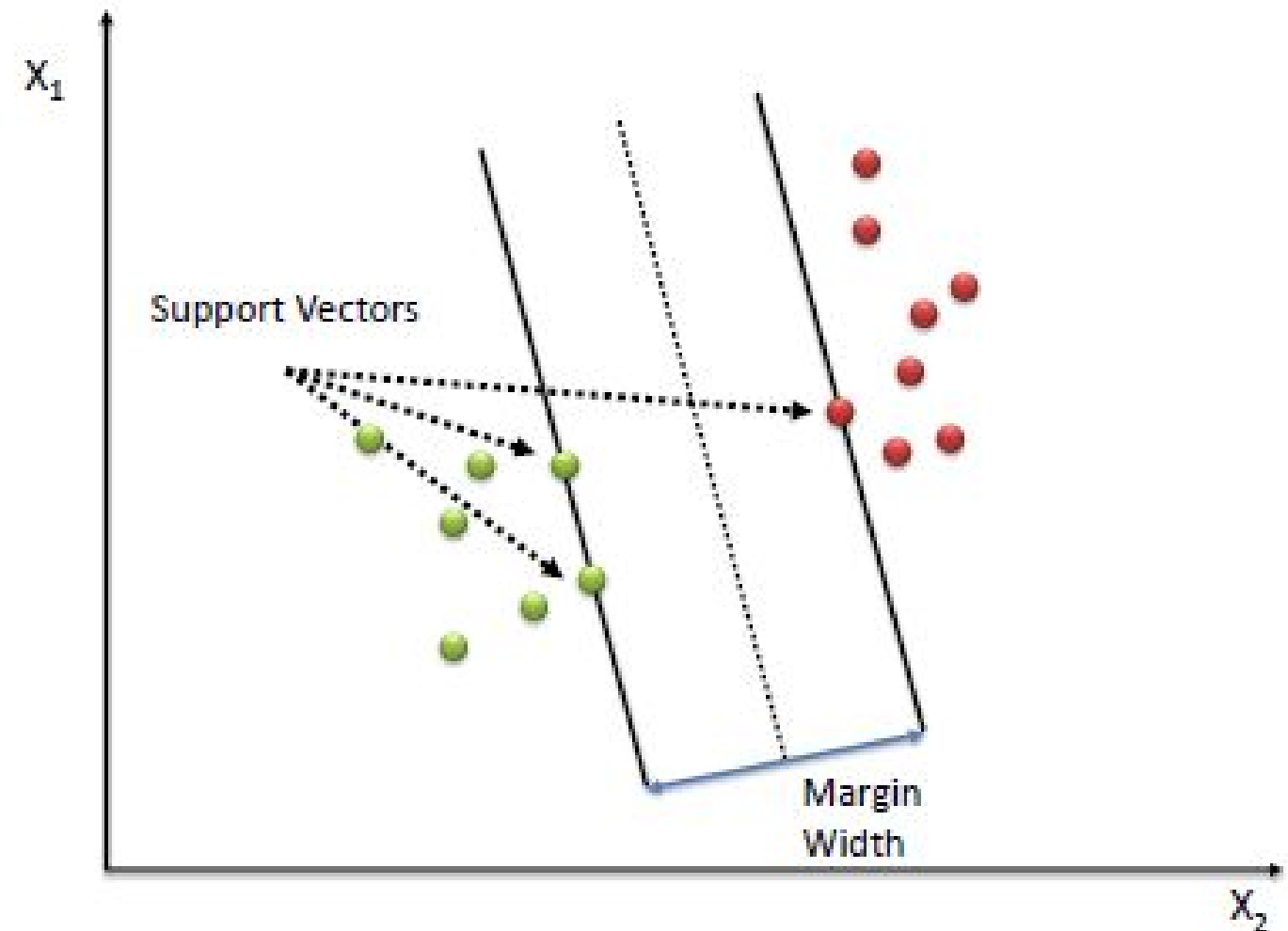
- + Requires no domain knowledge
- + Inherent feature selection
 - Black box
 - Many hyper-parameters
 - Very computationally expensive



SCILA

Support Vector Machines

- We have domain knowledge
- Features are alert parameters
- POC gave ~70% accuracy
- Transparent algorithm
- Real-time classification



Support Vector Machines: IRL

```
SVM.Trainer svmTrainer = new SVM.Trainer<>(new GaussianKernel(sigma), c)
    .setNumEpochs(numberOfEpochs);

SVM svm = svmTrainer.train(trainingData, trainingLabels);
svm.finish();

svm.trainPlattScaling(trainingData, trainingLabels);

int predictedLabel = svm.predict(testInstance);
```

- Structured grid search for optimal hyper parameters



Neural Networks in SMILE: IRL

```
Trainer neuralNetworkTrainer = new Trainer(  
    ErrorFunction.LEAST_MEAN_SQUARES,  
    ActivationFunction.LOGISTIC_SIGMOID,  
    ...numUnits: 100,100,100,2);  
  
neuralNetworkTrainer.setLearningRate(0.05)  
    .setMomentum(0)  
    .setNumEpochs(100)  
    .setWeightDecay(0);  
NeuralNetwork neuralNetwork = neuralNetworkTrainer  
    .train(trainingData, trainingLabels);  
  
int predictedLabel = neuralNetwork.predict(testInstance);
```

- Easy to setup but too simplistic



Neural Networks in DL4J: IRL

```
NeuralNetConfiguration netConfiguration = new NeuralNetConfiguration.Builder()
    .iterations(1)
    .weightInit(WeightInit.XAVIER)
    .activation("relu")
    .optimizationAlgo(OptimizationAlgorithm.STOCHASTIC_GRADIENT_DESCENT)
    .learningRate(0.05)
    // ... other hyperparameters
    .layer(new DenseLayer.Builder().nIn(784).nOut(250)
        .build())
    .build();
```

- More complicated but fully customizable

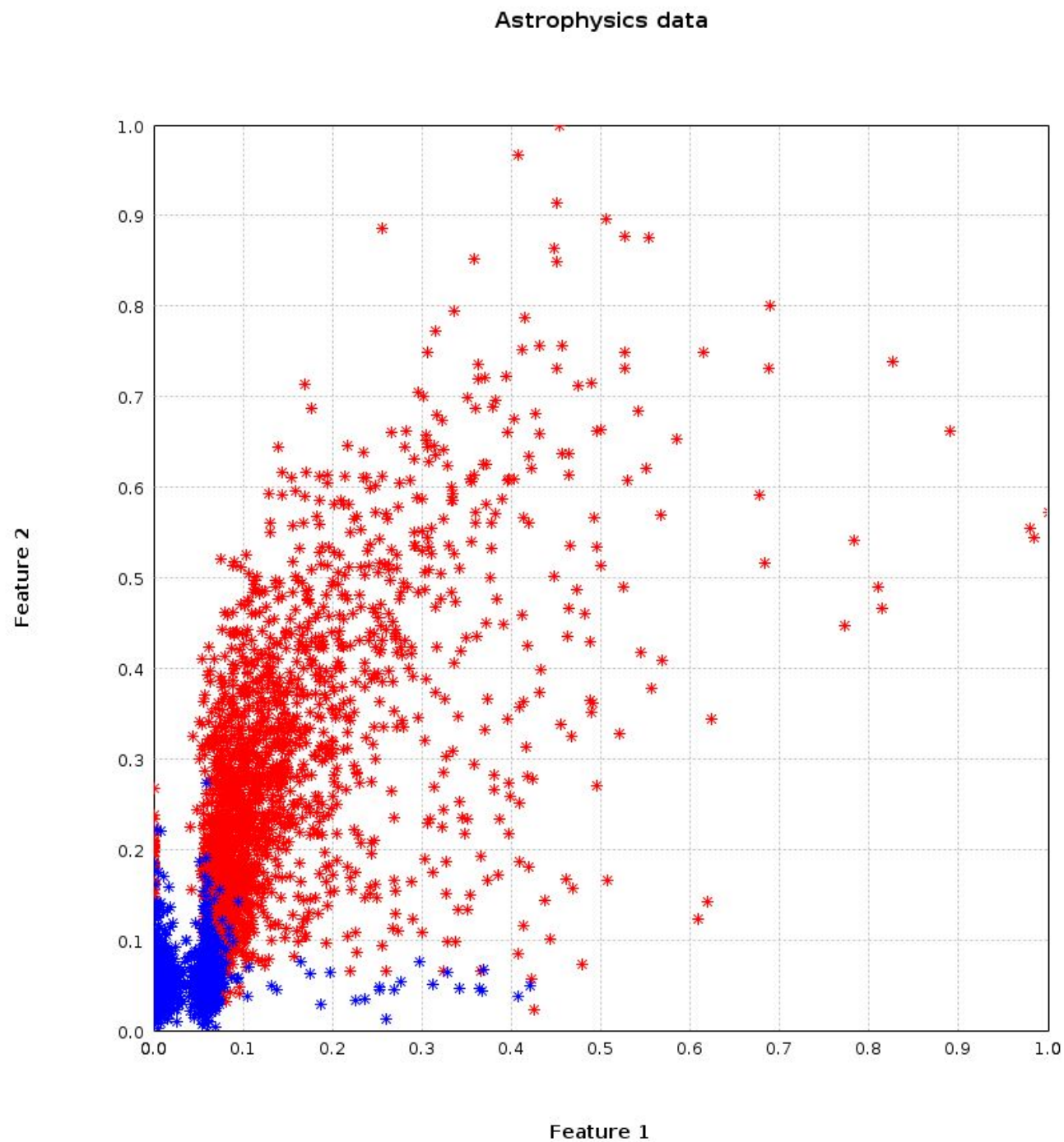


A Word on Data

- Results from a Support Vector Machine are plotted as a function of σ and c
- Financial data too sensitive to show!
- Two illustrative open source datasets:
 - Astrophysics simulation: muon vs other particle
 - Results: 97% (3880/4000) correctly classified
 - Liver-disease: disease vs no disease
 - Results: 62,5% (125/200) correctly classified



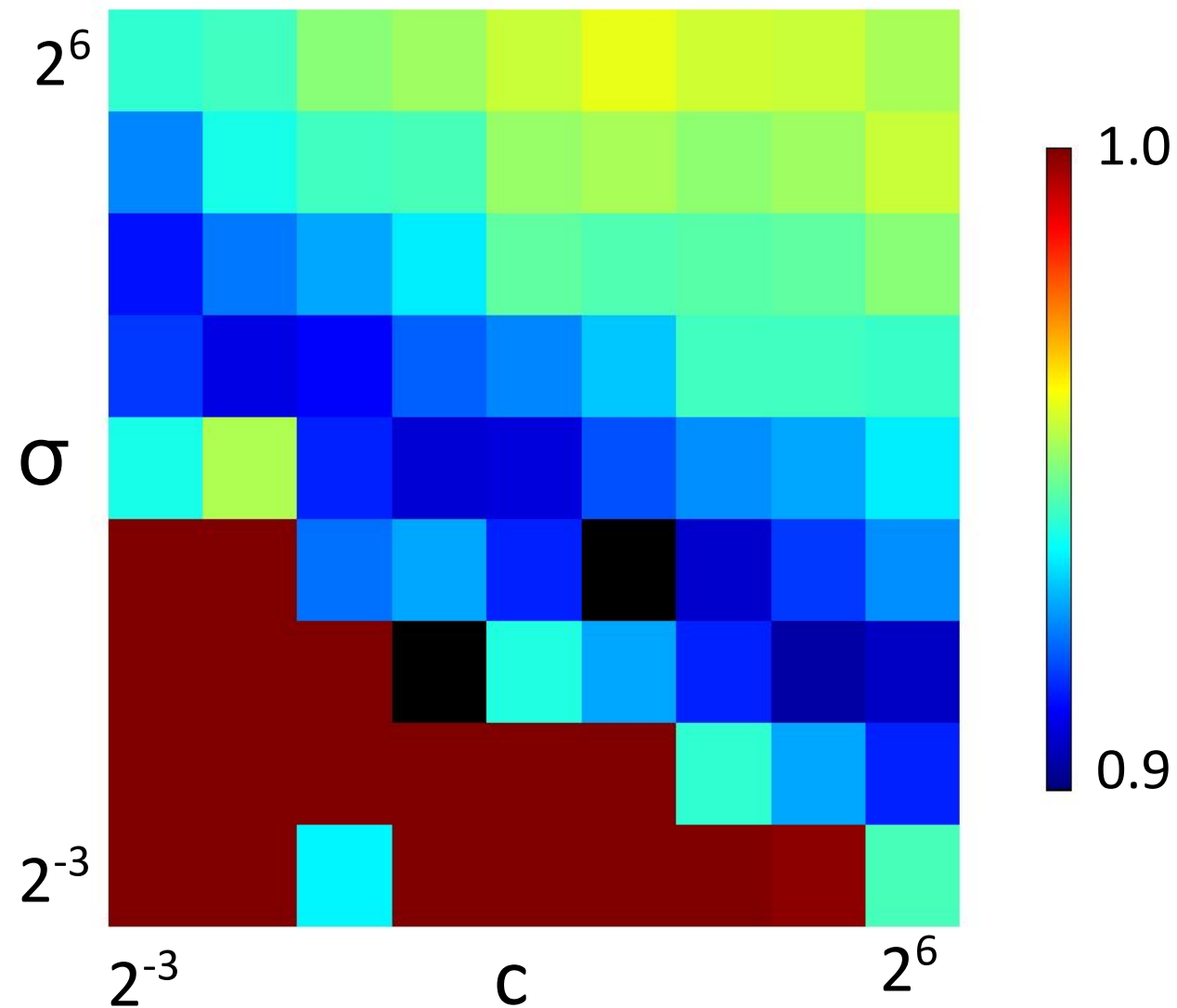
Some data is good...



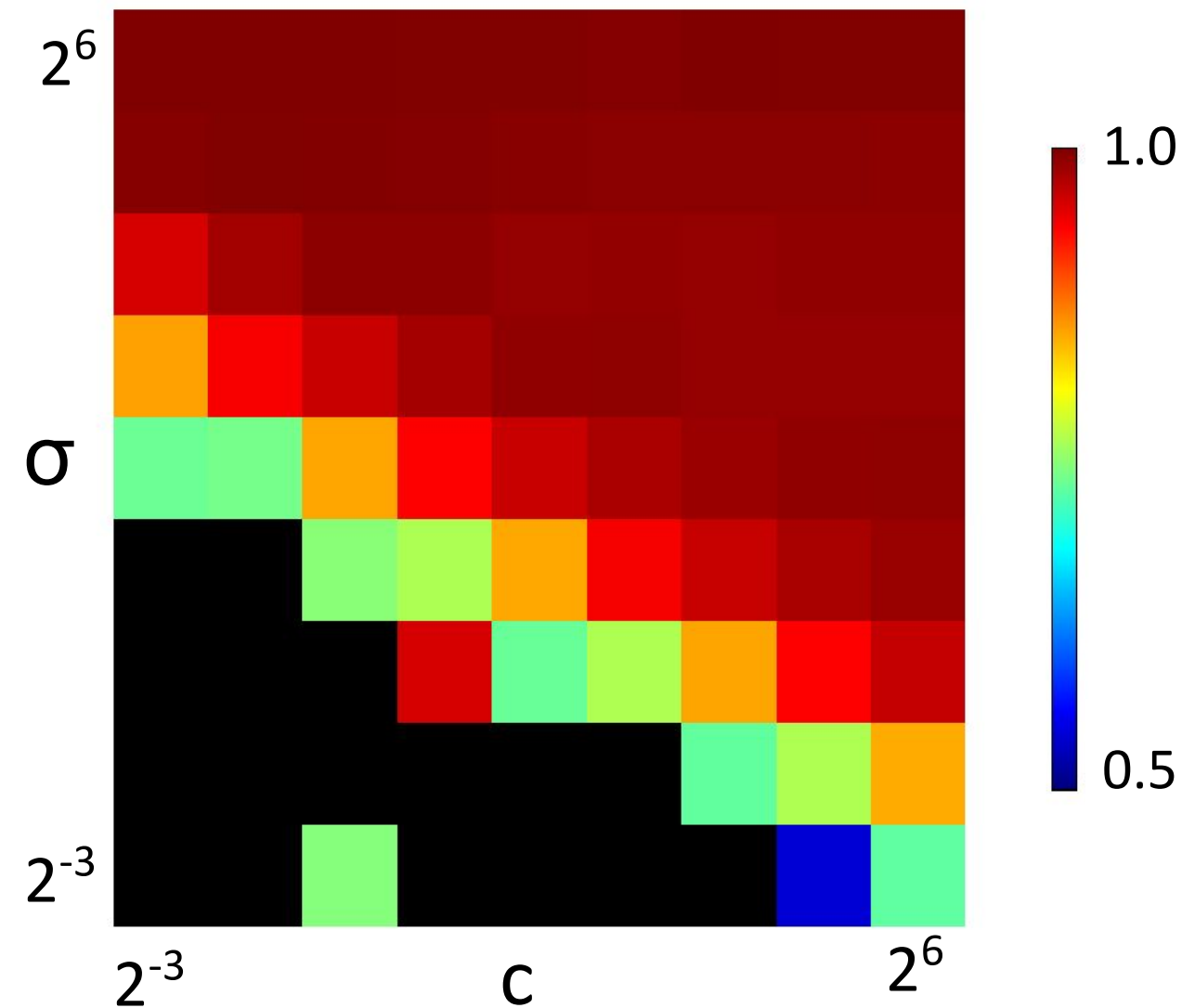
- Astrophysics data
- 4-dimensional

Some data is good...

Recall



Precision

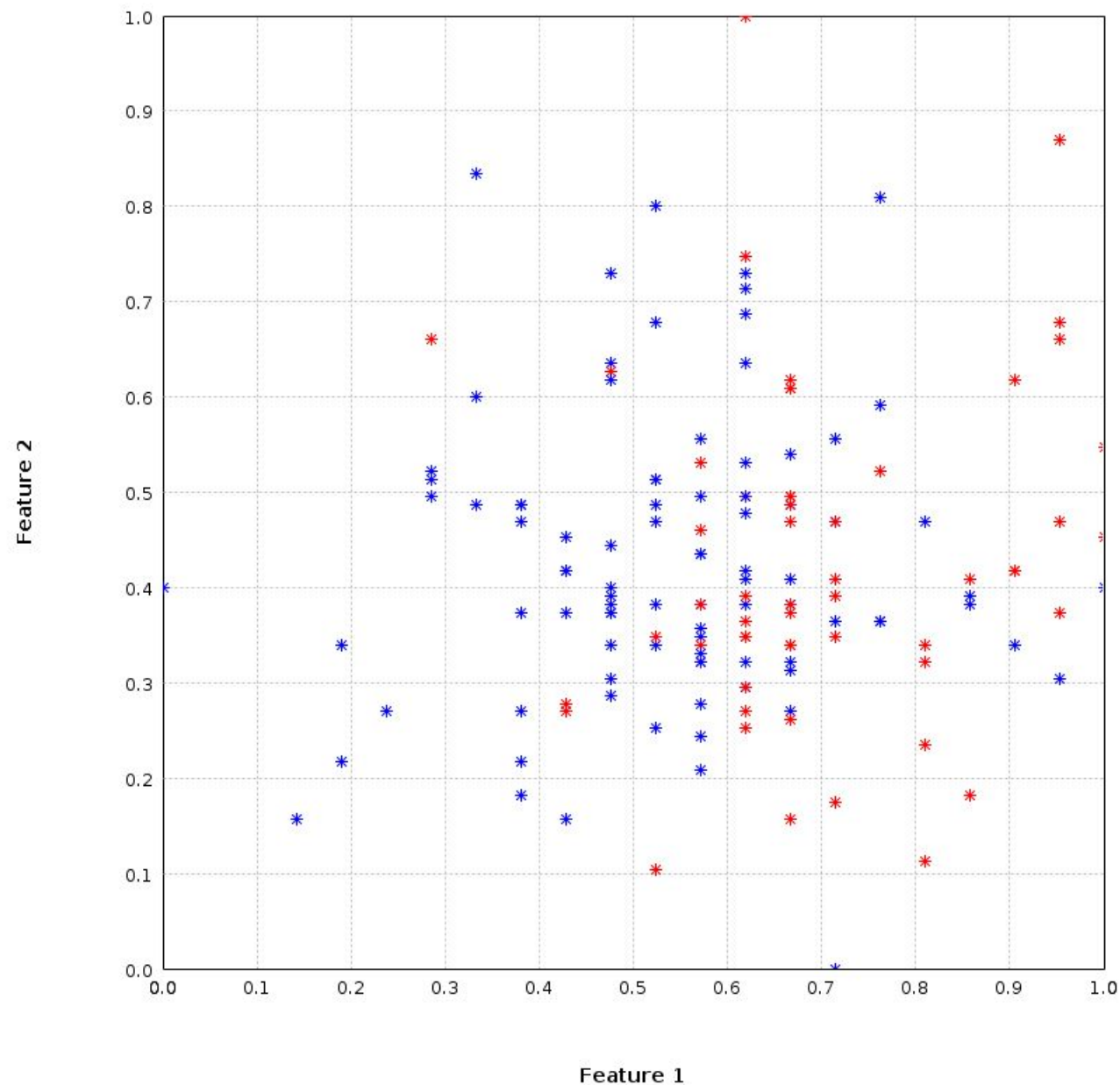


Results: 97% (3880/4000) correctly classified

SCILA

...and some data is bad!

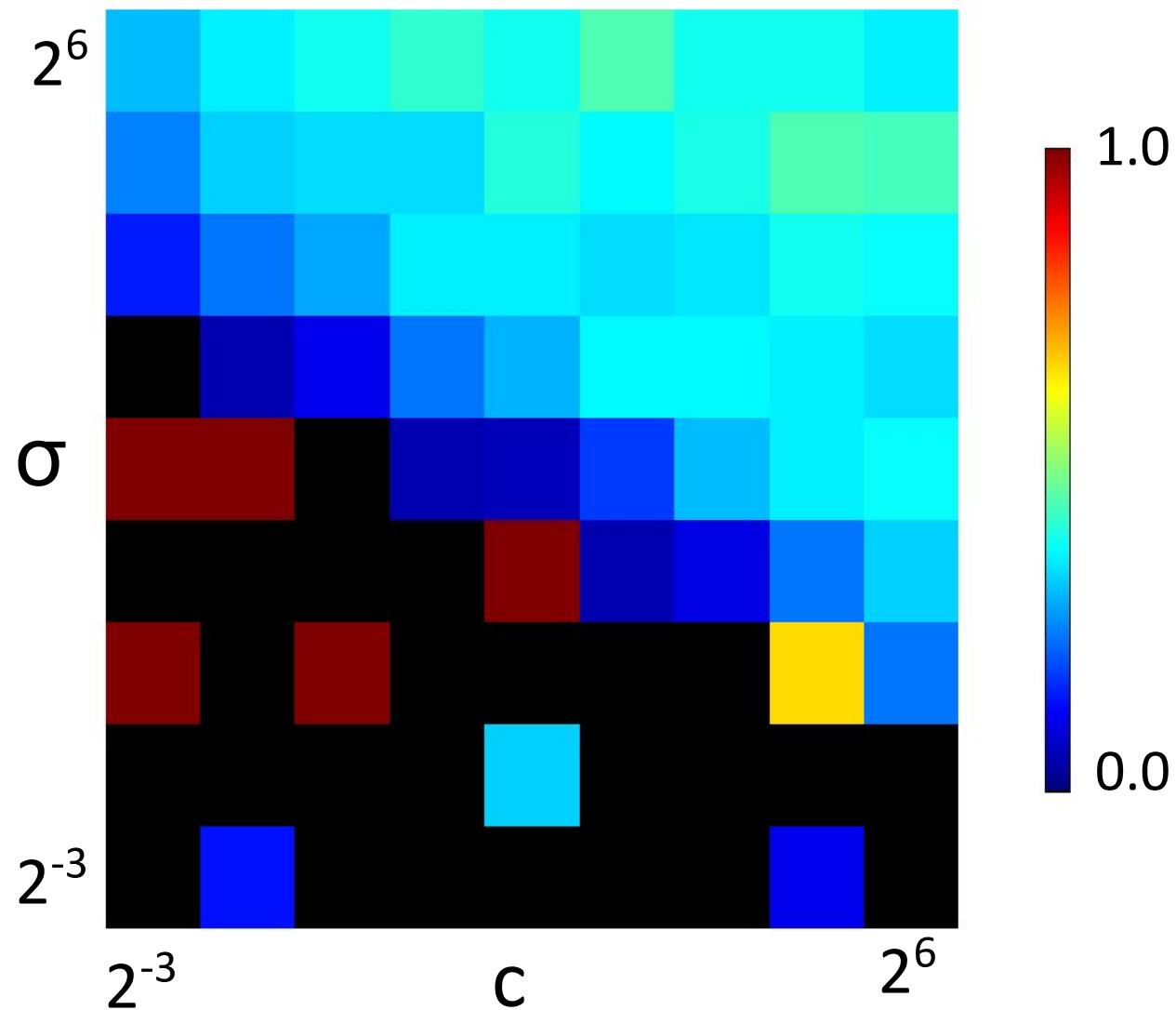
Liver-disease data



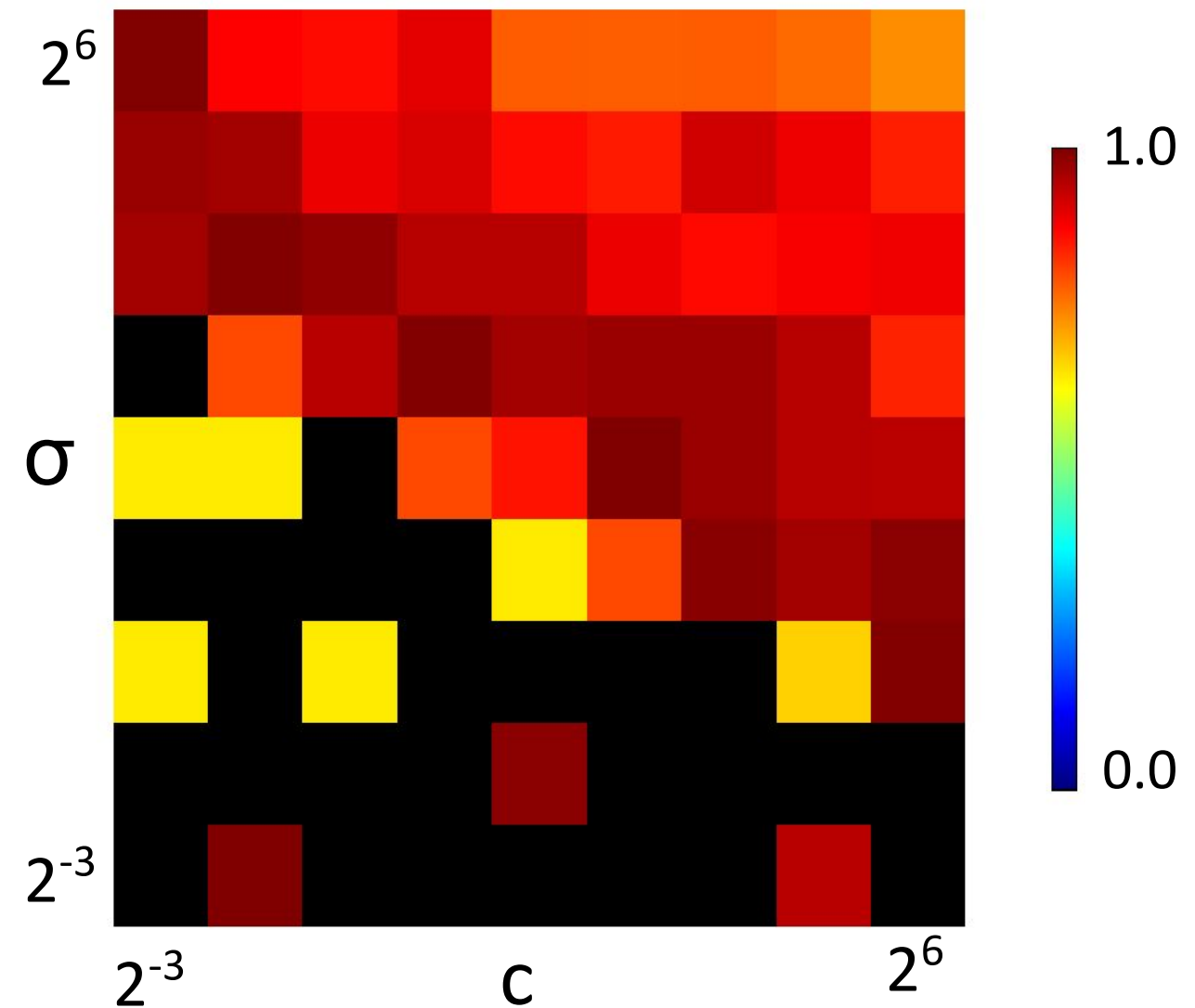
- Liver-disease data
- Features are blood tests
- 5-dimensional

...and some data is bad!

Recall



Precision



Results: 62,5% (125/200) correctly classified

SCILA

Tips of the Trade

- Analyze your data!
- Sampling || Weighting
- Normalize the data
- Understandable accuracy metrics



Tips of the Trade

- Analyze your data!
- Sampling || Weighting
- Normalize the data
- Understandable accuracy metrics

```
private Method getTakeHomeMessage() {  
    return domainKnowledge ? Method.SVM : Method.NEURAL_NETWORKS;  
}
```



Final slide: A catchy headline

Thank you for your time!

Questions? Come talk to us!

```
developersAtJFokus.stream()  
  .parallel()  
  .filter(Developer::isNice)  
  .filter(Developer::isTalented)  
  .filter(Developer::isKeenToLearnMore)  
  .map(Developer::getCv)  
  .map(Cv::toMail)  
  .reduce("jobs@scila.se", (a,m) -> m.sendTo(a), (a,b) -> a);
```



References

- Where we're from: www.scila.se
- Where you should email: jobs@scila.se

Links to frameworks of the trade:

- <http://haifengl.github.io/smile/>
- <https://deeplearning4j.org/index.html>
- <http://spark.apache.org/mllib/>
- <https://www.tensorflow.org/tutorials/>
- <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

