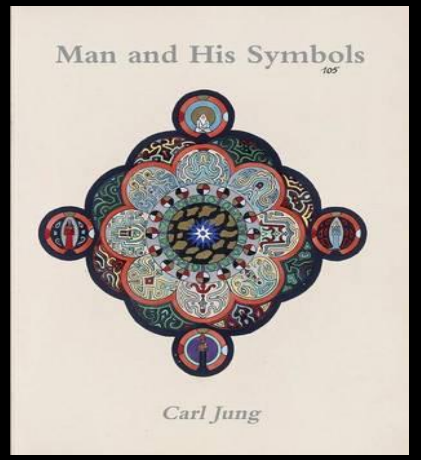
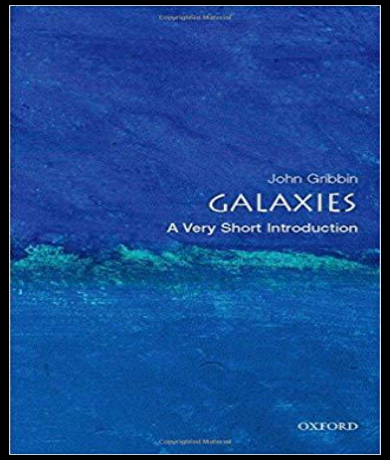
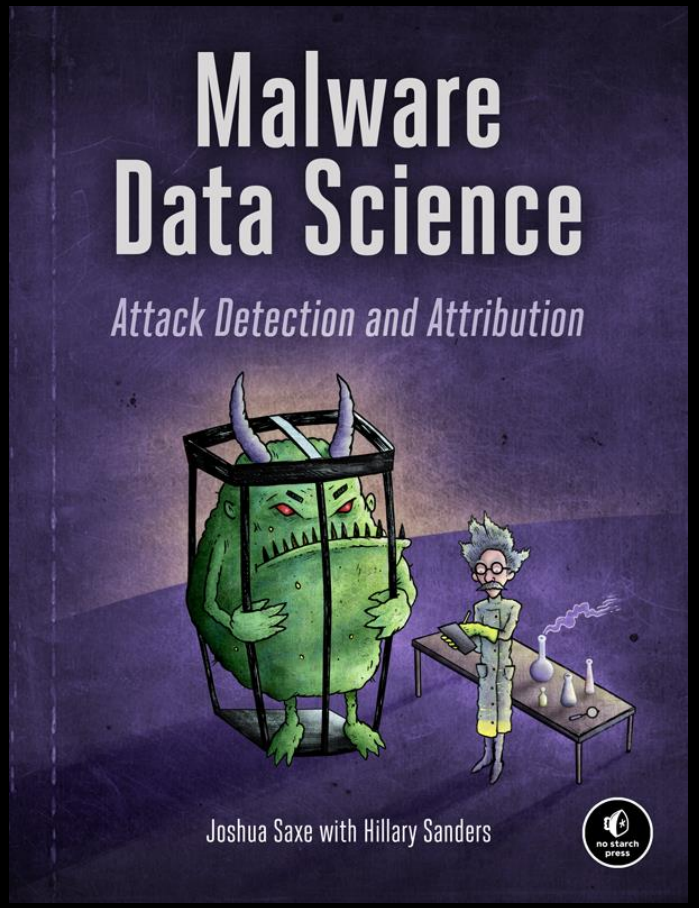
Deep Neural Networks for Hackers: Methods, Applications, and Open Source Tools

Joshua Saxe, Chief Data Scientist, Sophos

About your speaker







Your speaker's team @ Sophos

Josh (me)

Maddy Schiappa

Rich Harang

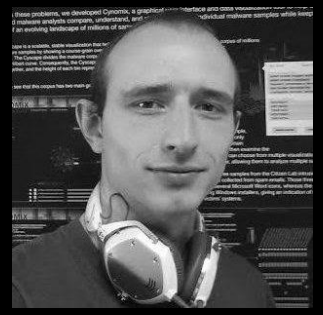
Hillary Sanders

Adarsh Kyadige

Konstantin Berlin

Ethan Rudd

Felipe Ducau



Alex Long

Cody Wild

William Lee

Matt Stec

Matt Burnett

You

Security Data Science

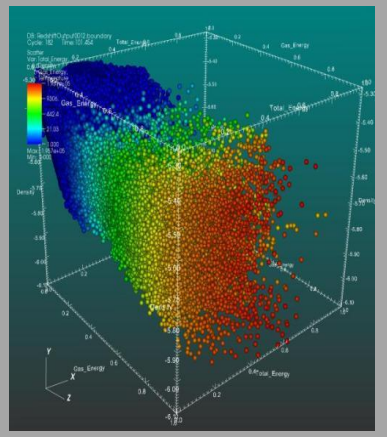
My focus today

Machine learning

Deep learning (a.k.a. deep neural networks)

... other machine learning approaches

Data Visualization



Databases, cloud engineering

This is deep learning: real time computer vision

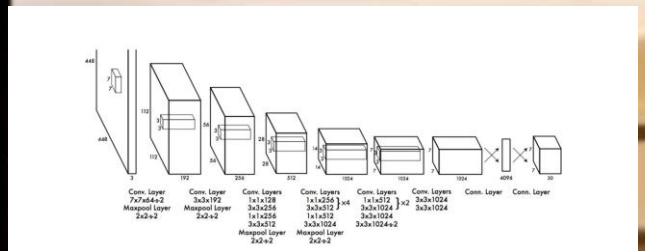
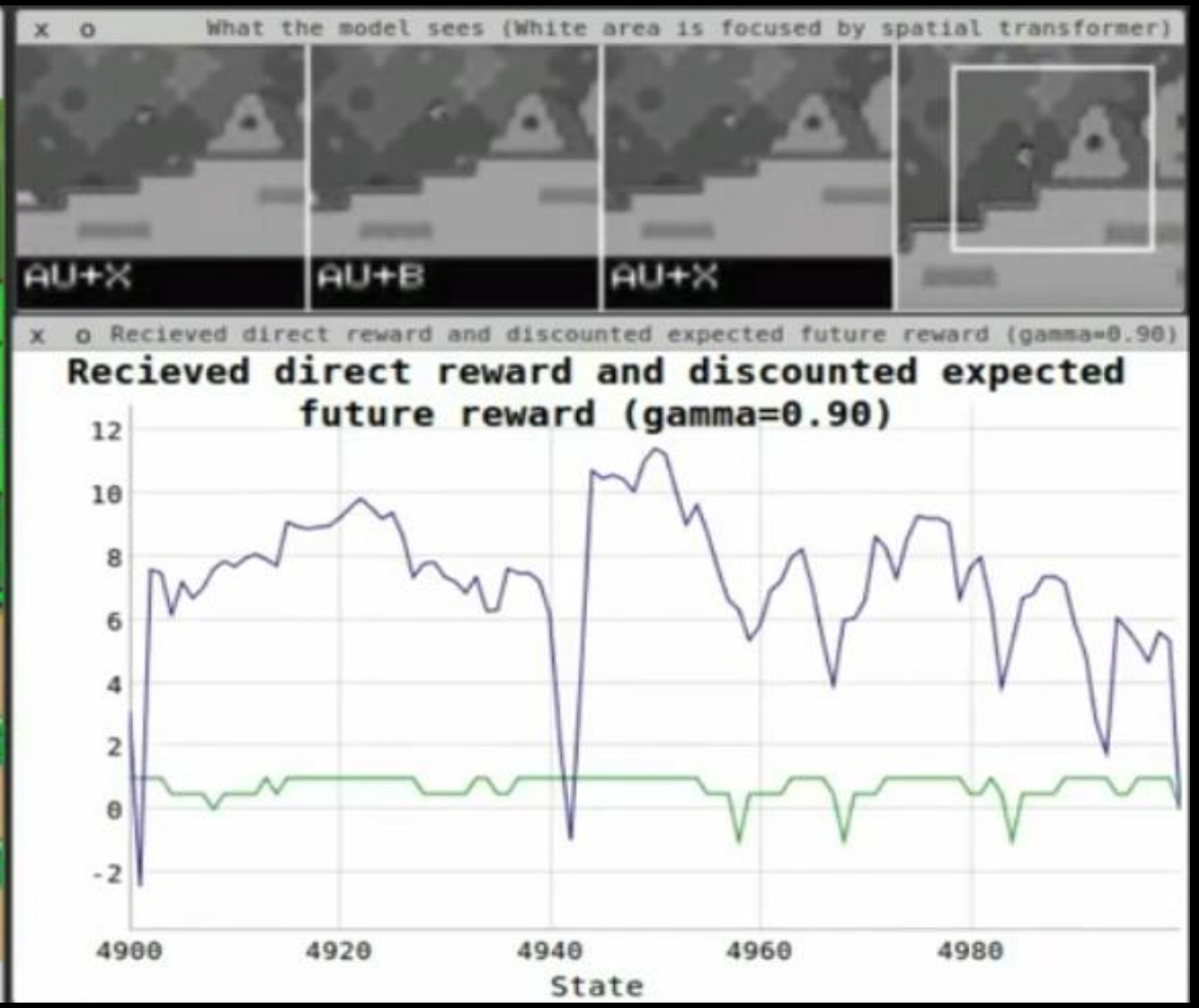


Figure 3: The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1×1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution (224×224 input image) and then double the resolution for detection.

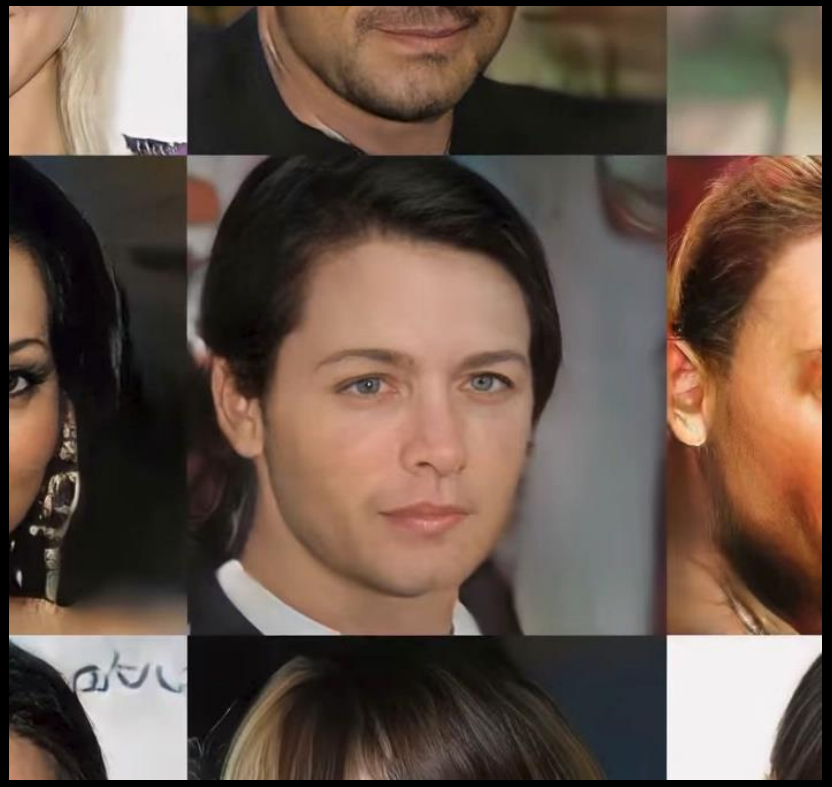
https://pjreddie.com/darknet/yolo/

This is deep learning: computer vision + reinforcement learning



https://github.com/aleju/mario-ai

This is deep learning: new content generation

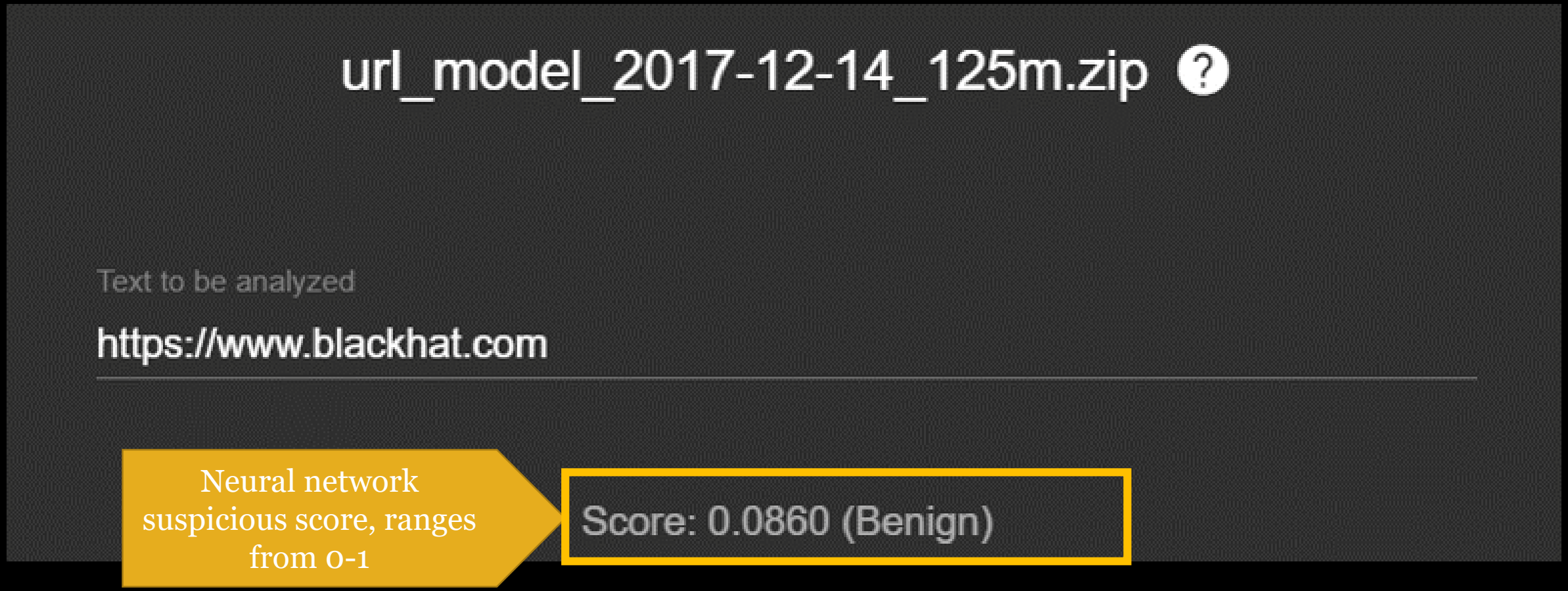


```
/*
 * Increment the size file of the new incorrect UI_FILTER group information
 * of the size generatively.
 */
static int indicate_policy(void)
{
  int error;
  if (fd == MARN_EPT) {
    /*
     * The kernel blank will coeld it to userspace.
     */
    if (ss->segment < mem_total)
      unblock_graph_and_set_blocked();
    else
      ret = 1;
    goto bail;
  }
  segaddr = in_SB(in.addr);
  selector = seg / 16;
  setup_works = true;
  for (i = 0; i < blocks; i++) {
    seq = buf[i++];
    bpf = bd->bd.next + i * search;
    if (fd) {
      current = blocked;
    }
  }
  rw->name = "Getjbbregs";
  bprm_self_clearl(&iv->version);
  regs->new = blocks[(BPF_STATS << info->historidac)] | PFMR_CLOBATHINC_SECONDS << 12;
  return segtable;
}
```

http://research.nvidia.com/publication/2017-10_Progressive-Growing-of

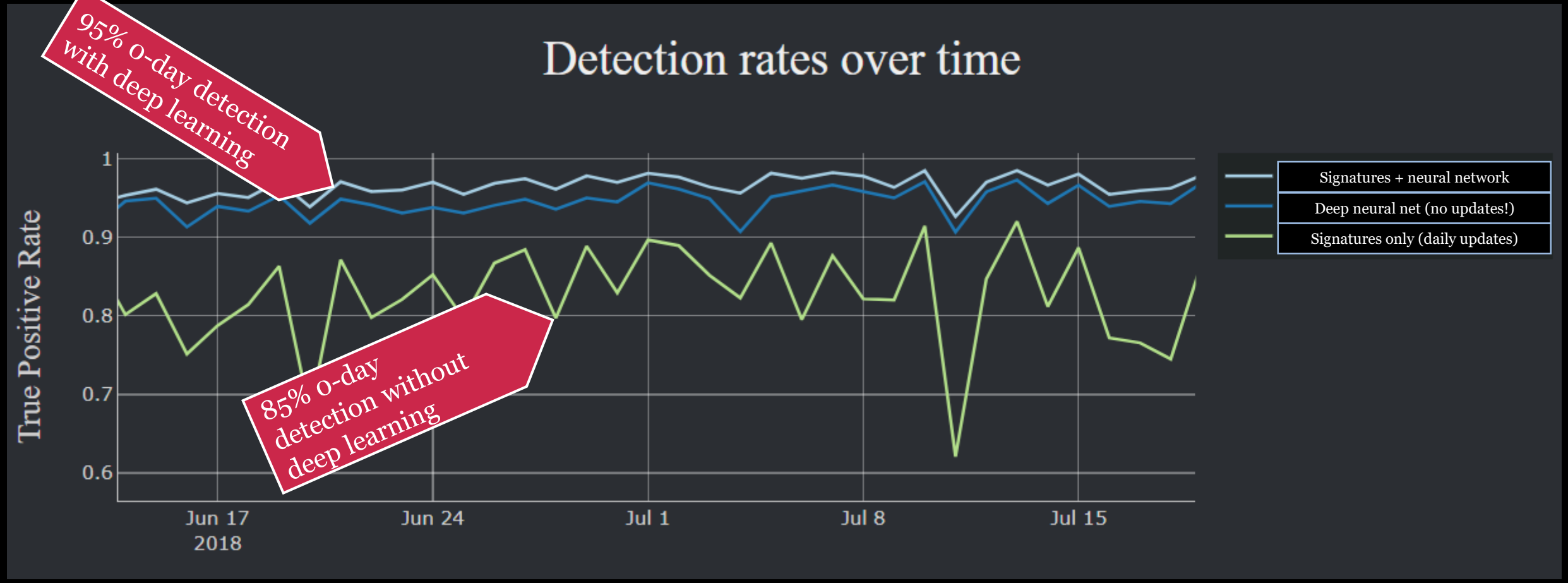
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

This is deep learning: detection of previously unseen URLs

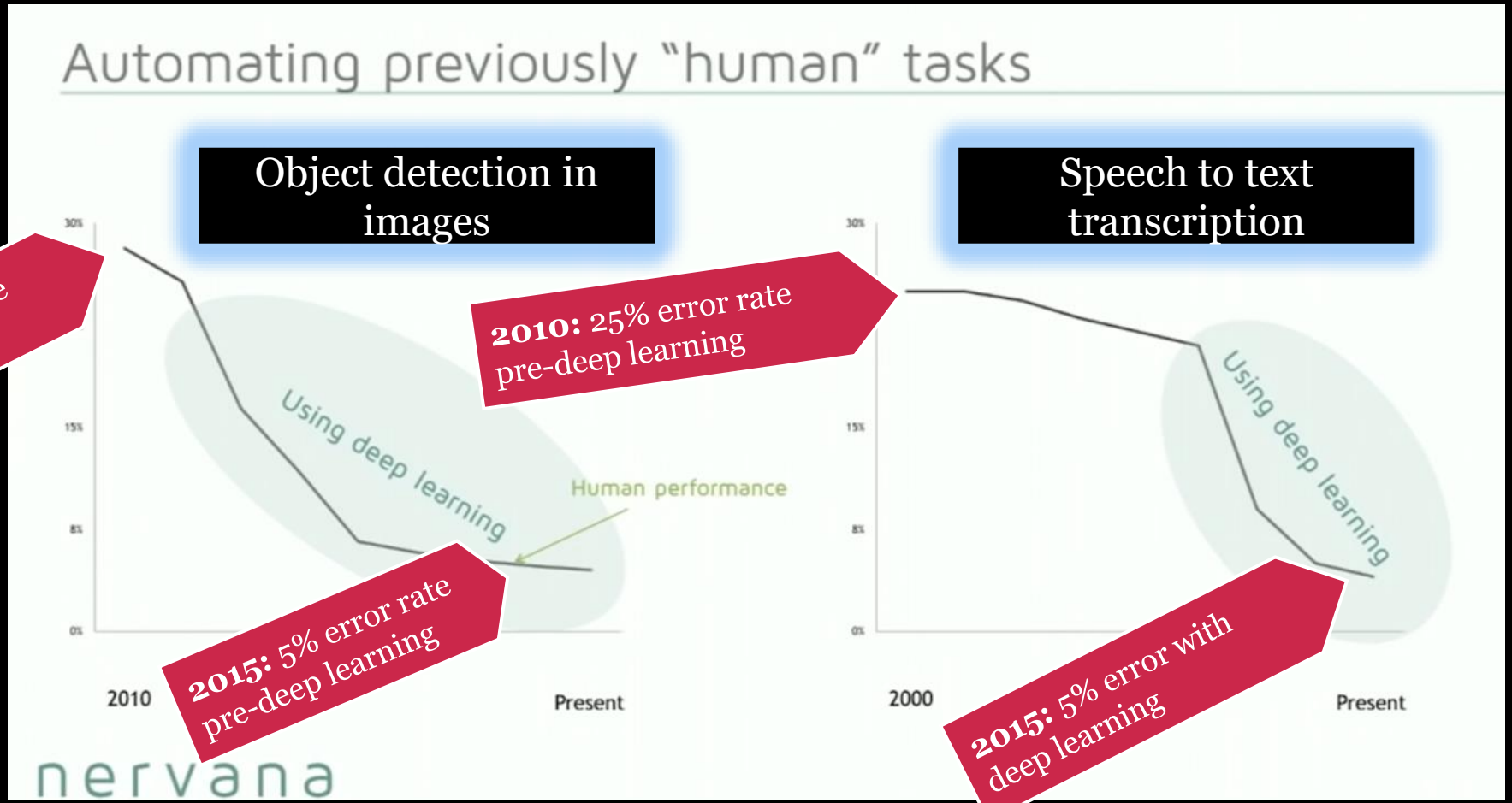


https://arxiv.org/abs/1702.08568

The deep learning revolution in numbers (malware detection)



The deep learning revolution in numbers



https://ai.intel.com/intel-nervana-neural-network-processor-architecture-update/

At Sophos we've come to use deep learning everywhere

HTML

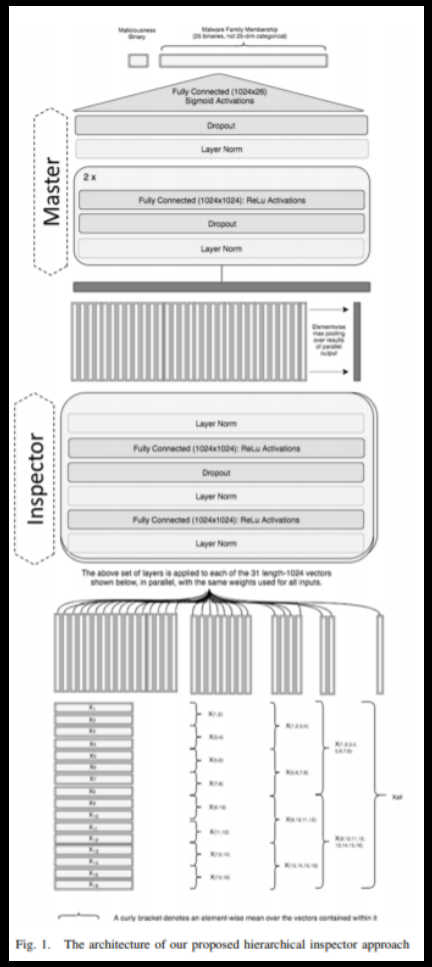
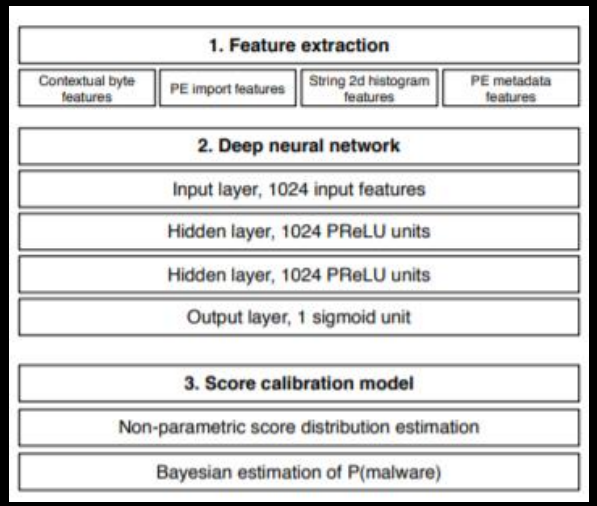
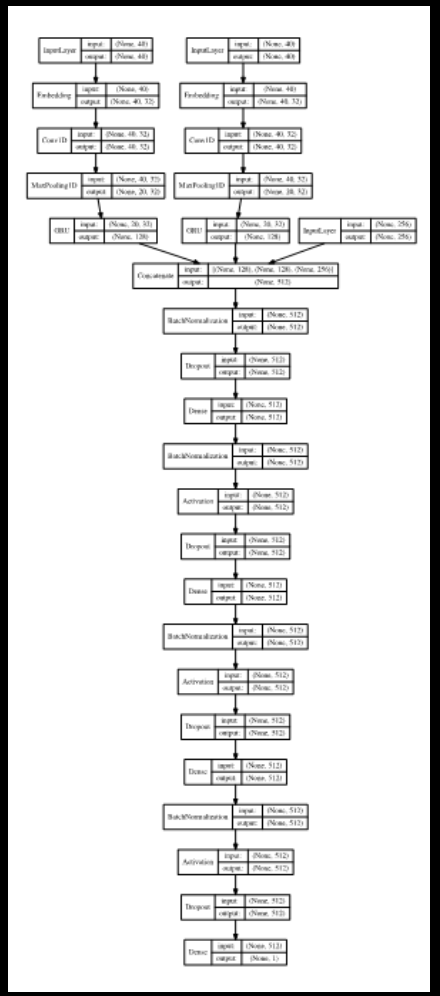


Fig. 1. The architecture of our proposed hierarchical inspector approach

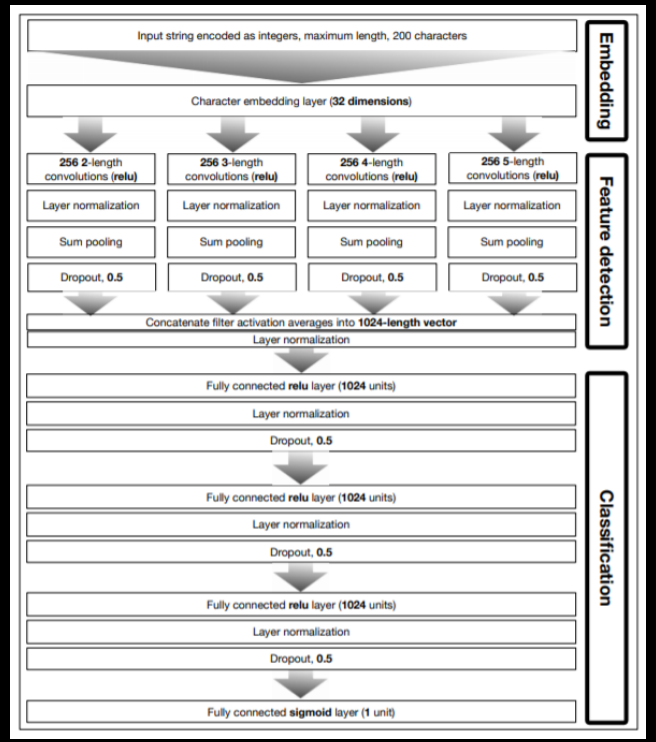
Windows PE



Android



File paths / registry keys



How deep learning and machine learning change security

- Old world
  - Detection rules written manually
  - Daily updates to blacklists and signatures
  - Anti-virus data takes hundreds of megabytes
  - Sub-par ability to detect new threats
- New world
  - Detection models trained automatically
  - Models need updating every month or two
  - Models take about 10 megabytes on disk
  - Breakthrough in ability to detect new threats

When neural networks go wrong:

Neural network generated recipe

http://aiweirdness.com

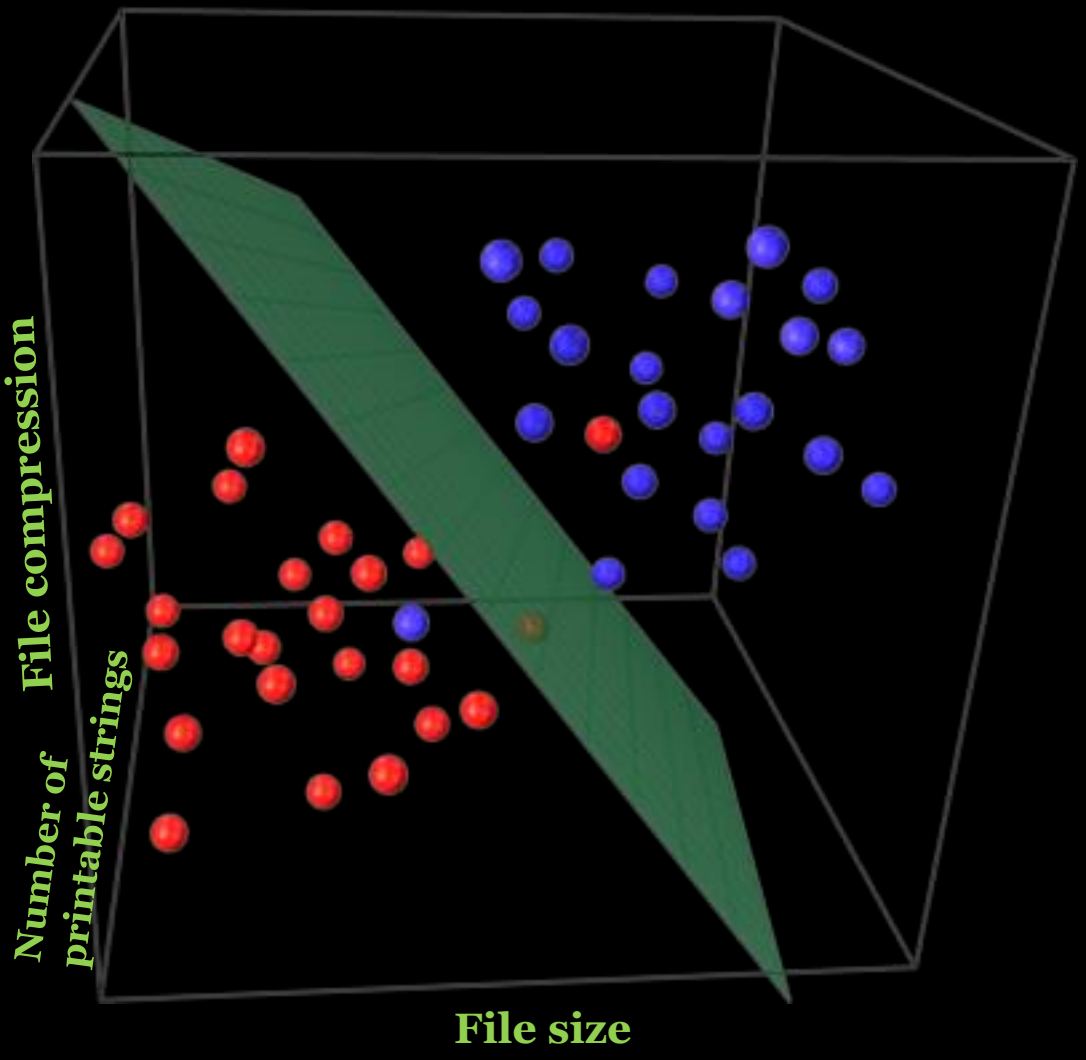
When neural networks go wrong:

Neural network generated recipe

http://aiweirdness.com

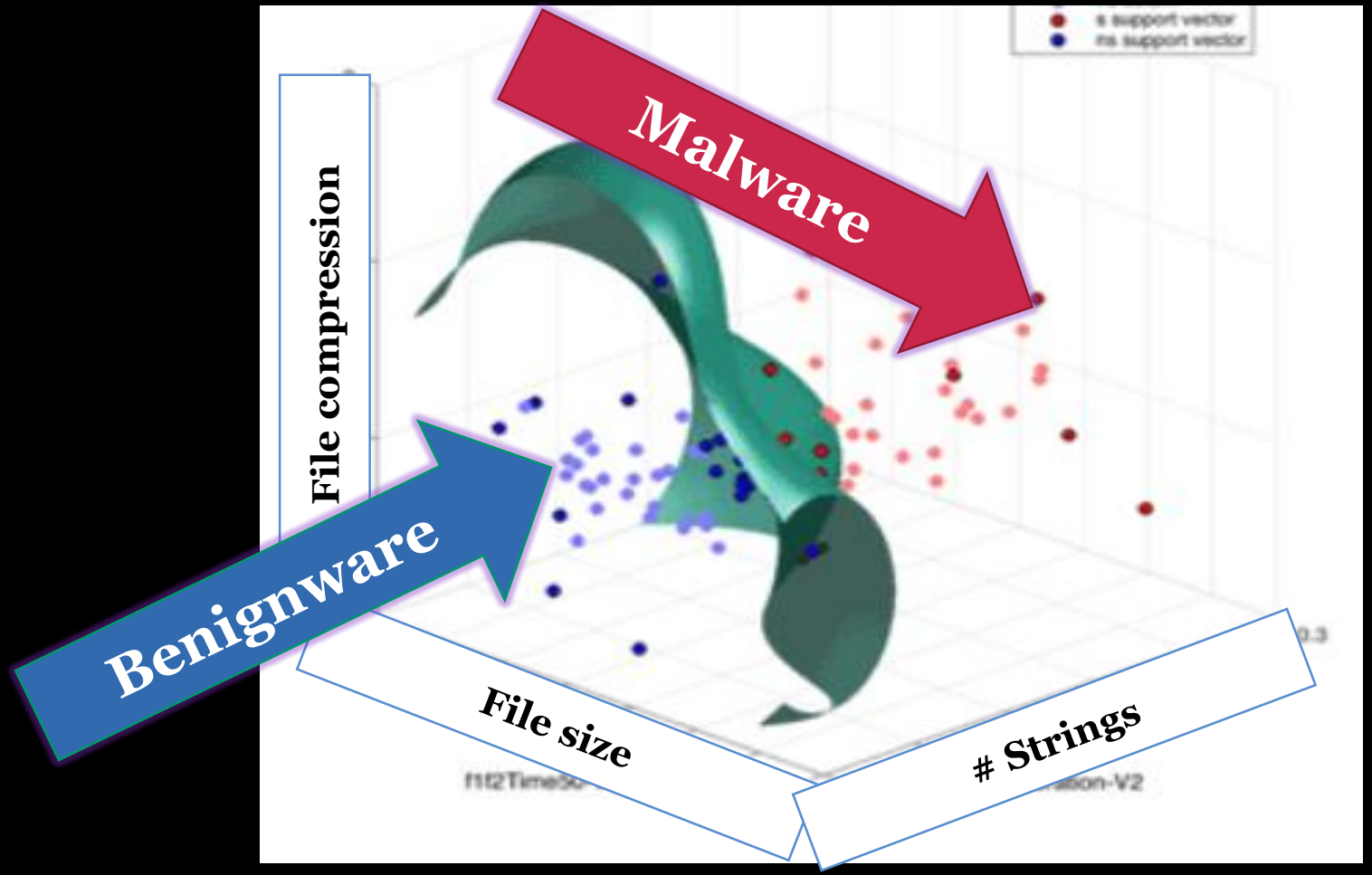

Sometimes deep learning is *not* the answer

How does machine learning based detection work?



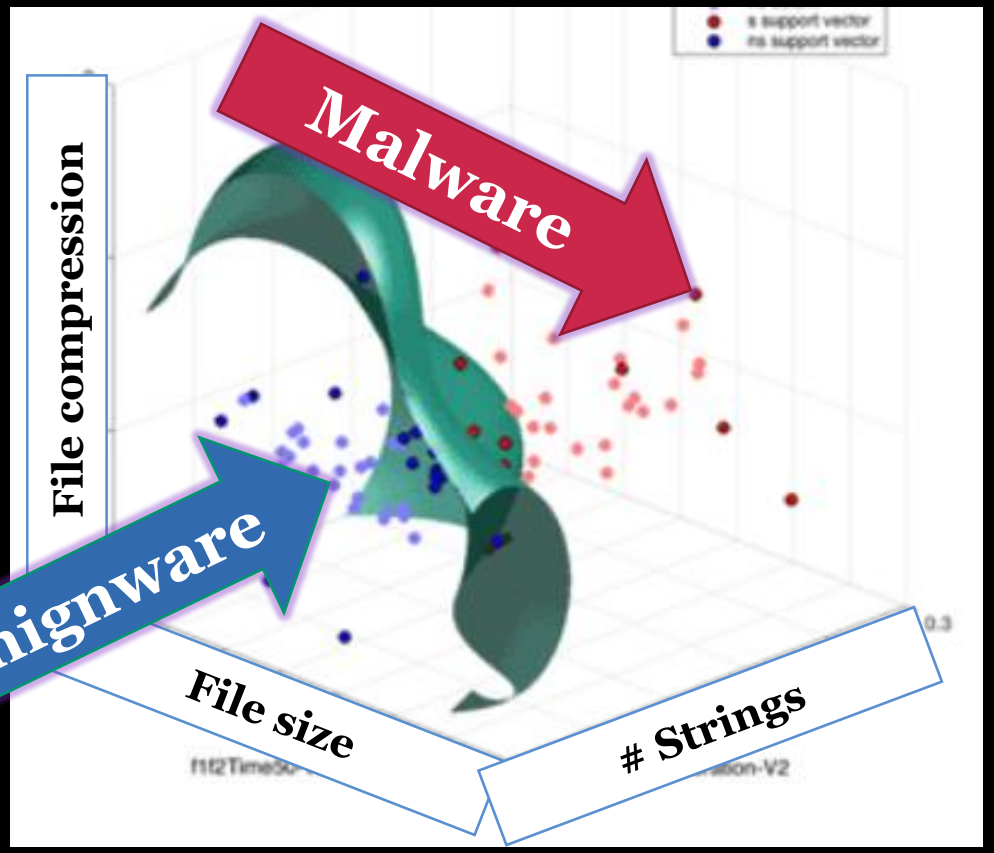
Deep learning uses “non-linear” decision boundaries

2d decision boundary in 3d space

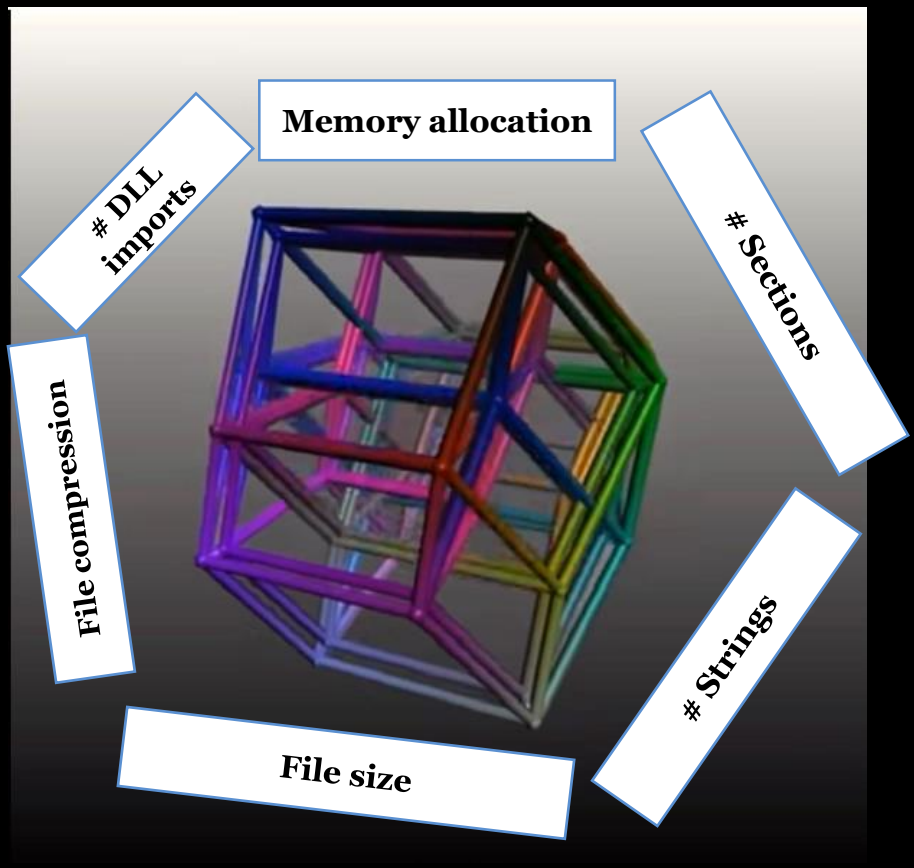


Real decision boundaries live in high dimensional space

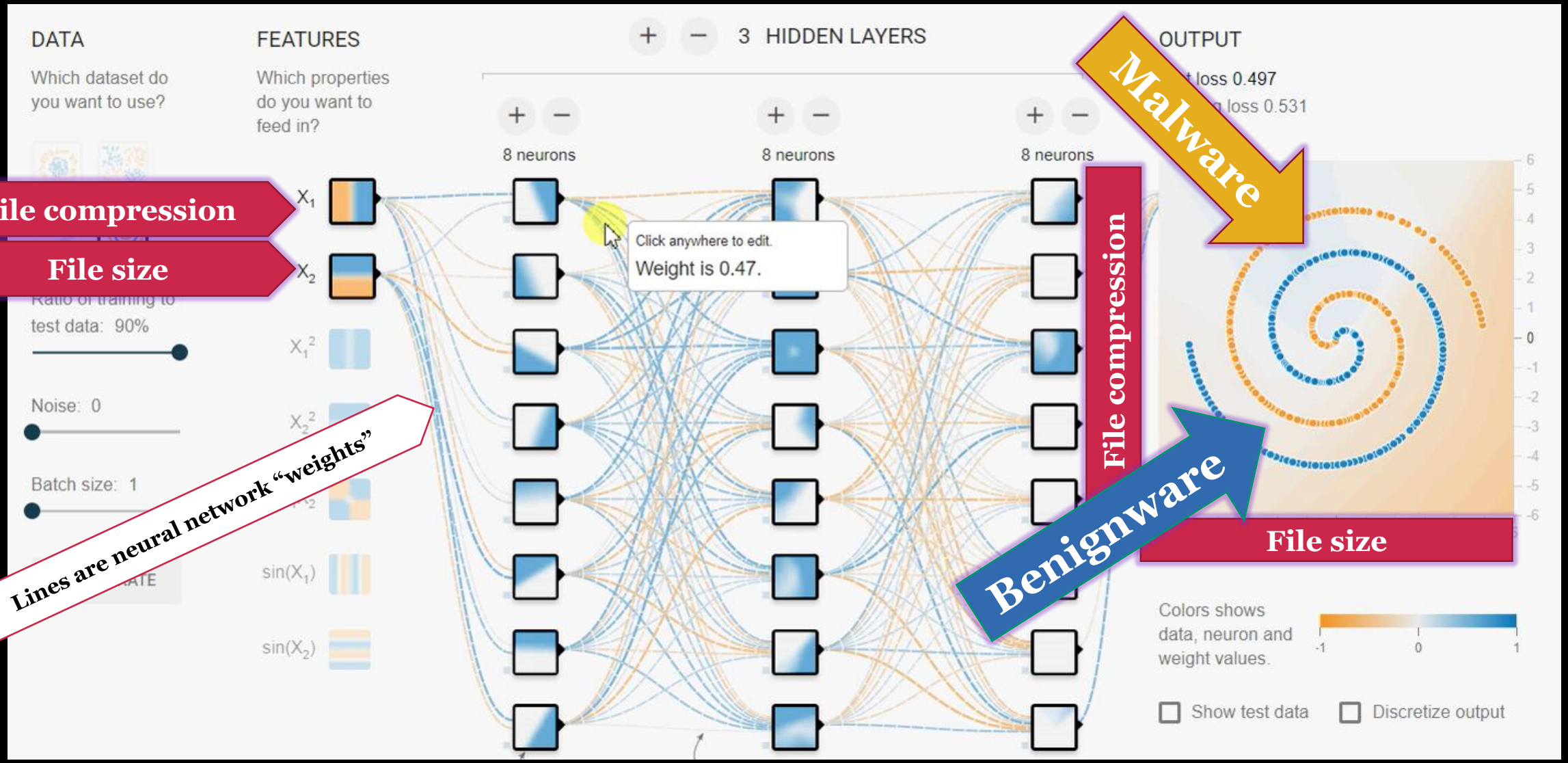
2d decision boundary in 3d space



6d space would have a 5d decision boundary!

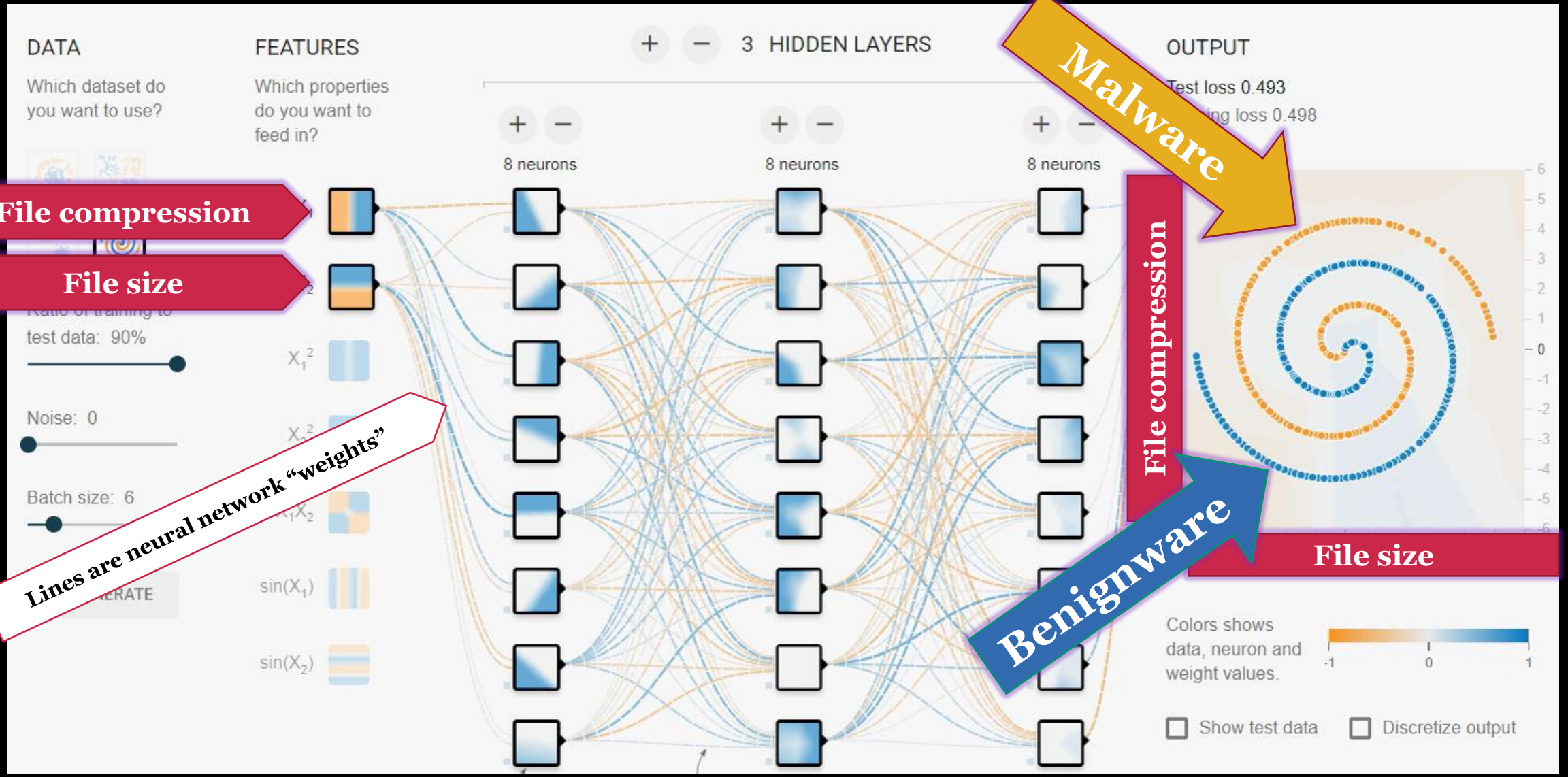


Neural networks as machines for drawing decision boundaries

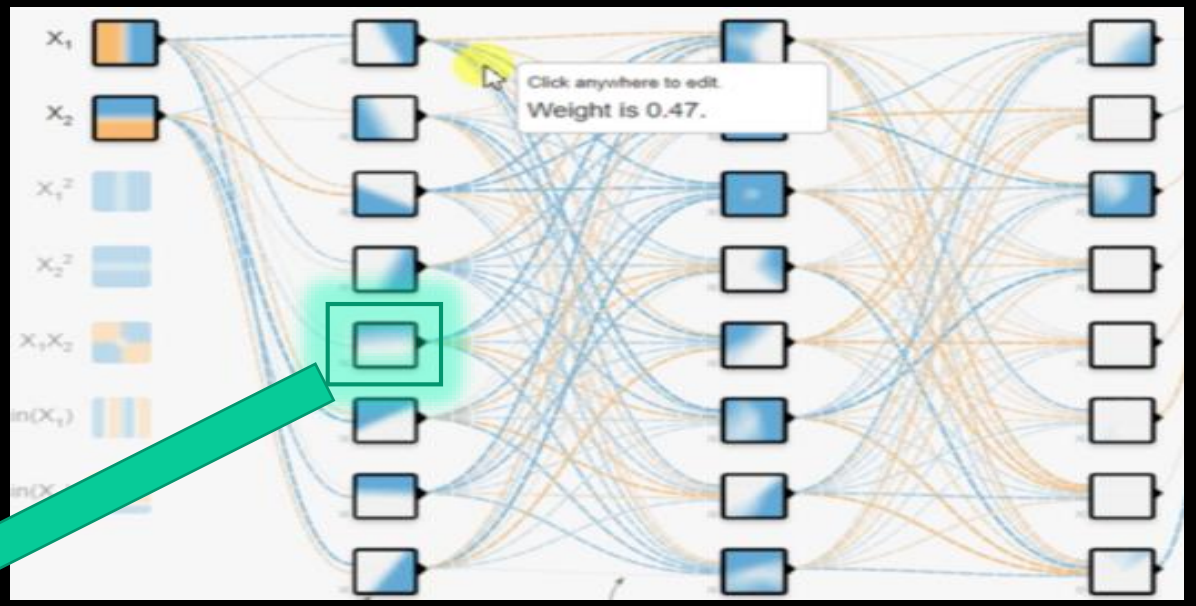


https://playground.tensorflow.org

Automatic neuron weight tuning (a.k.a. *learning*)



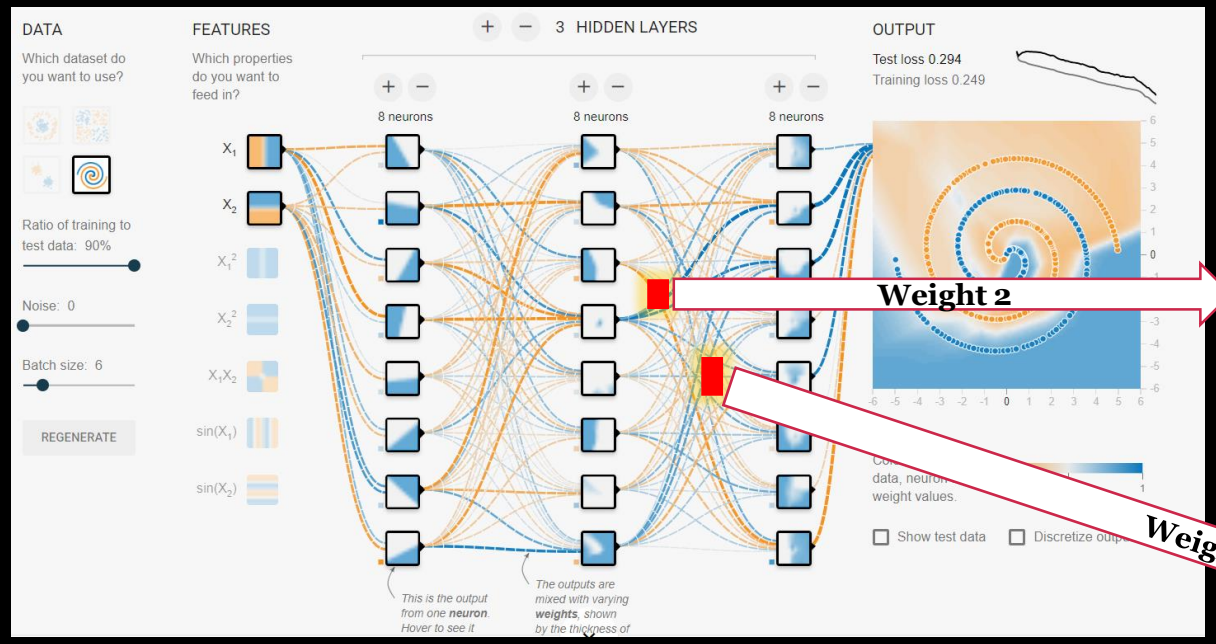
How neurons work as simple computational units

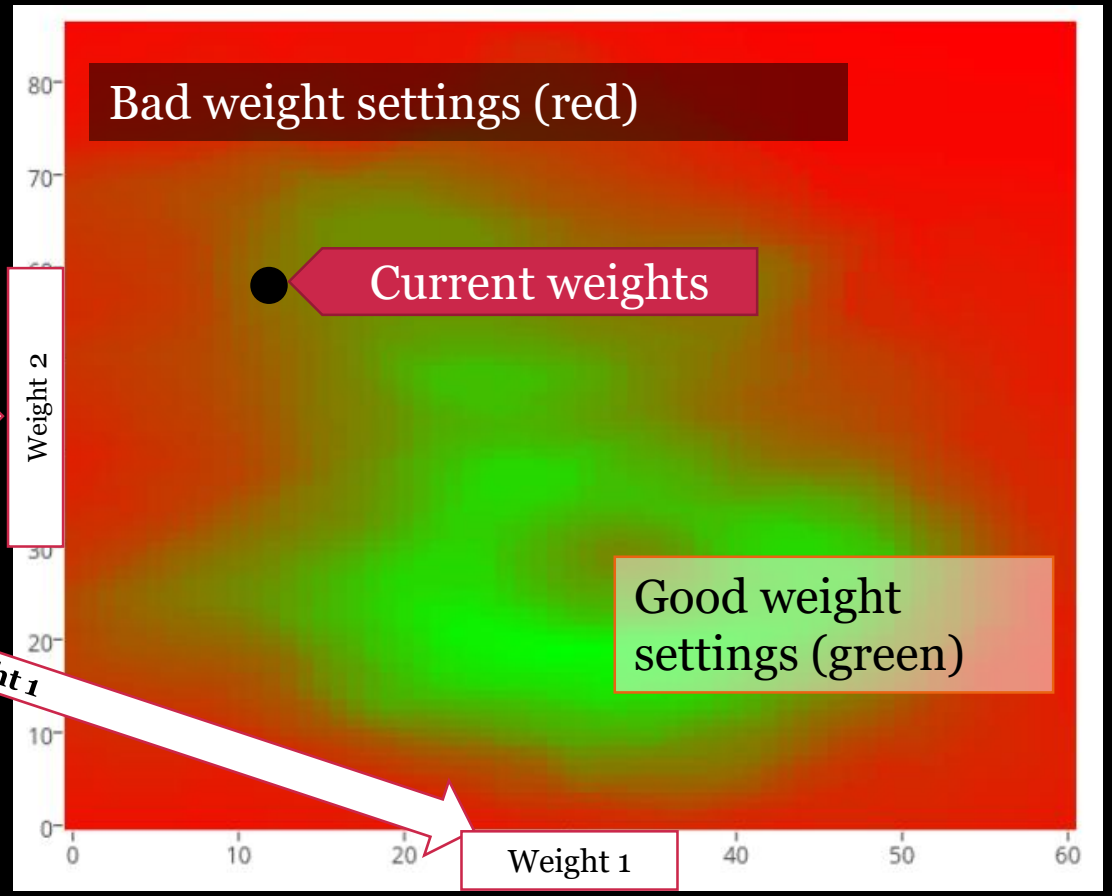



```python
#!/usr/bin/python

def compute_neuron(
    neuron_weights=[], # connection strengths,
    neuron_inputs=[], # data coming into neuron,
    num_incoming_connections=1, # how many data points incoming,
    bias=0 # starting energy of our neuron
    ):

    energy = 0 + bias # initial neuron energy

    for idx in range(num_incoming_connections):
        weight = neuron_weights[idx]
        neuron_input = neuron_inputs[idx]
        energy += weight * neuron_input # add weight * input to energy

    if energy > 0:
        # "activate" if energy is more than 0
        return energy
    else:
        # otherwise don't activate (return 0)
        return 0
```

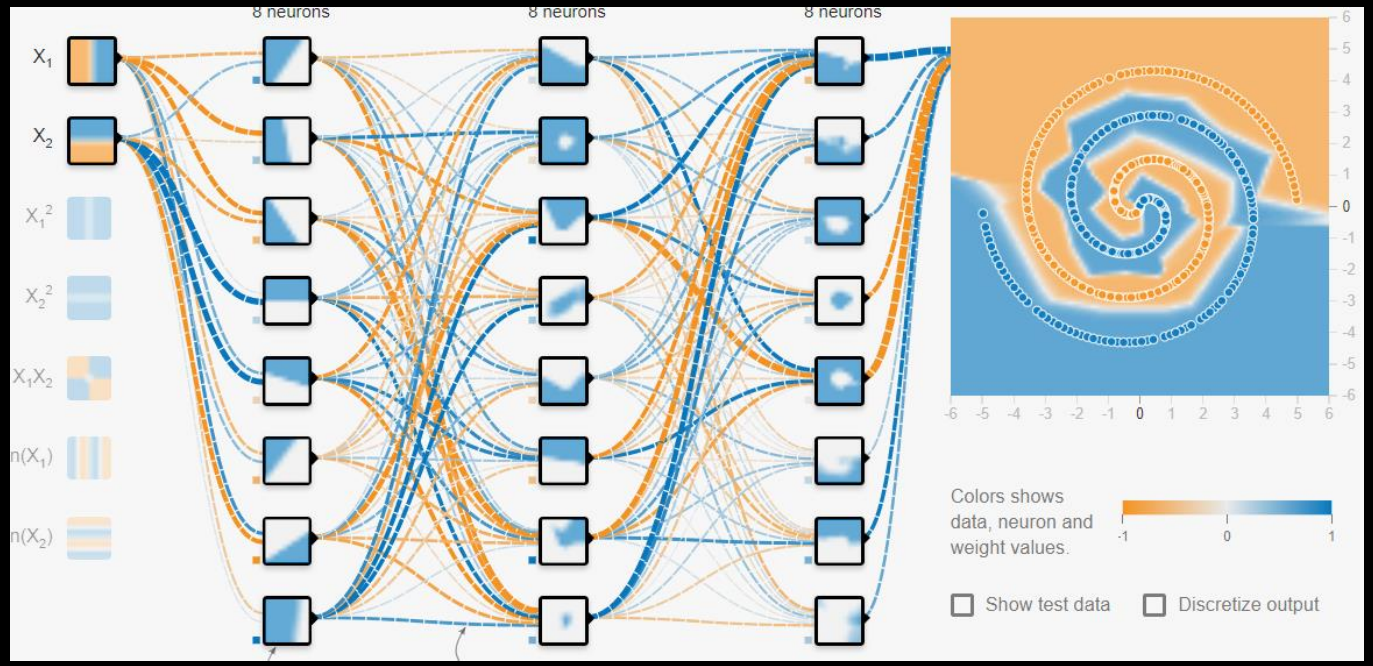
```
In [19]: compute_neuron(neuron_weights=[1,1,1], neuron_inputs=[5,5,5], num_incoming_connections=3)
```

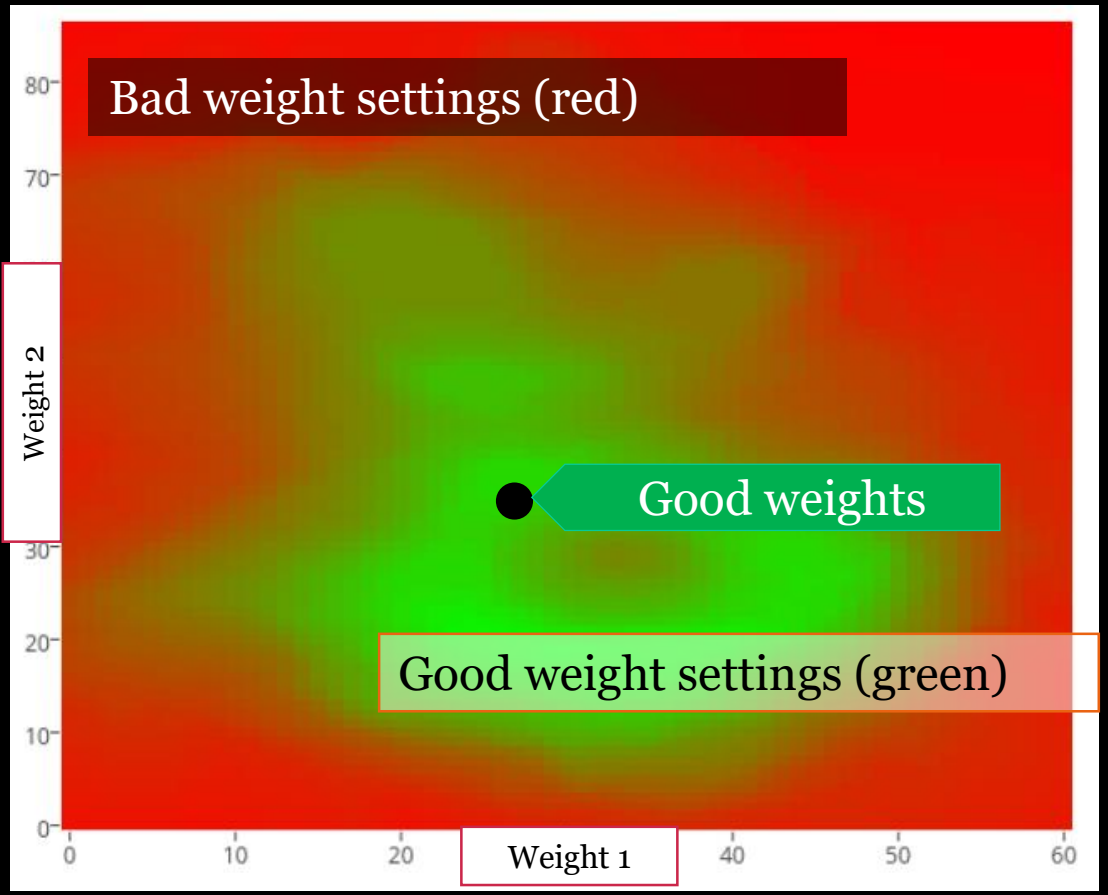
Behind the decision boundary “learning” process



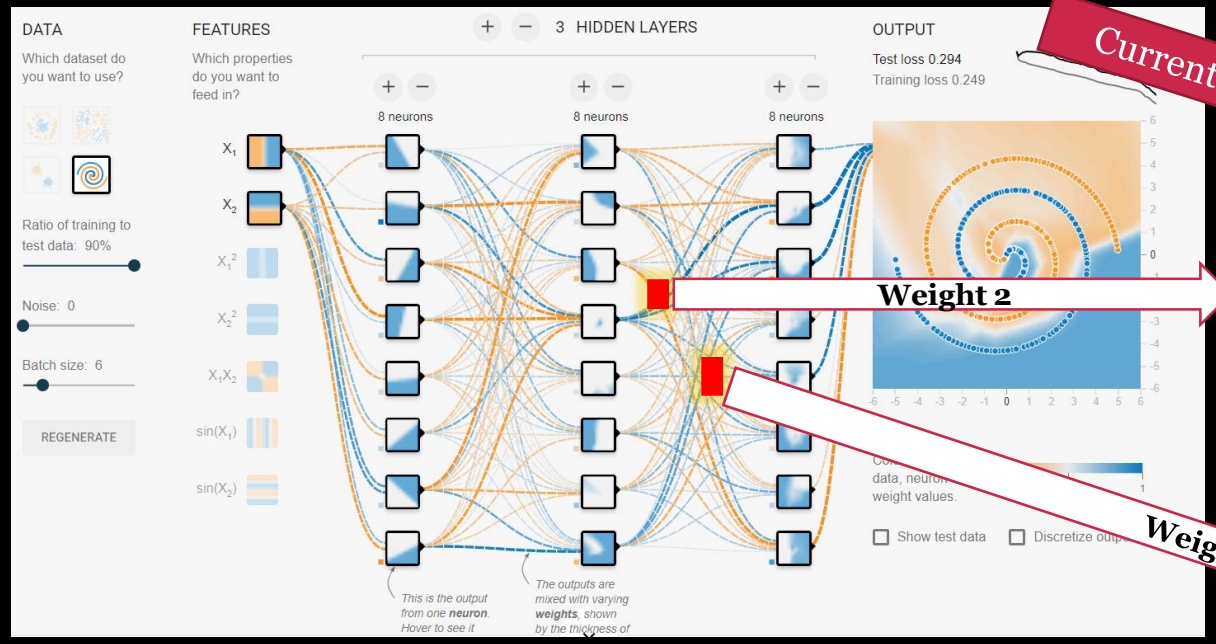


Behind the decision boundary “learning” process

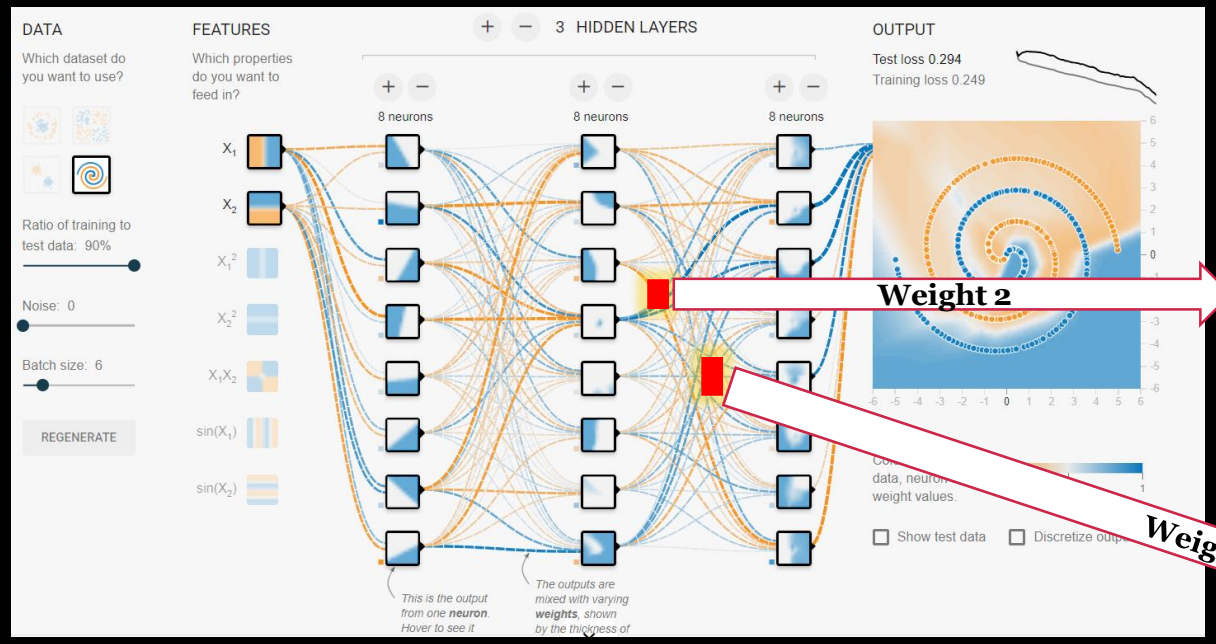


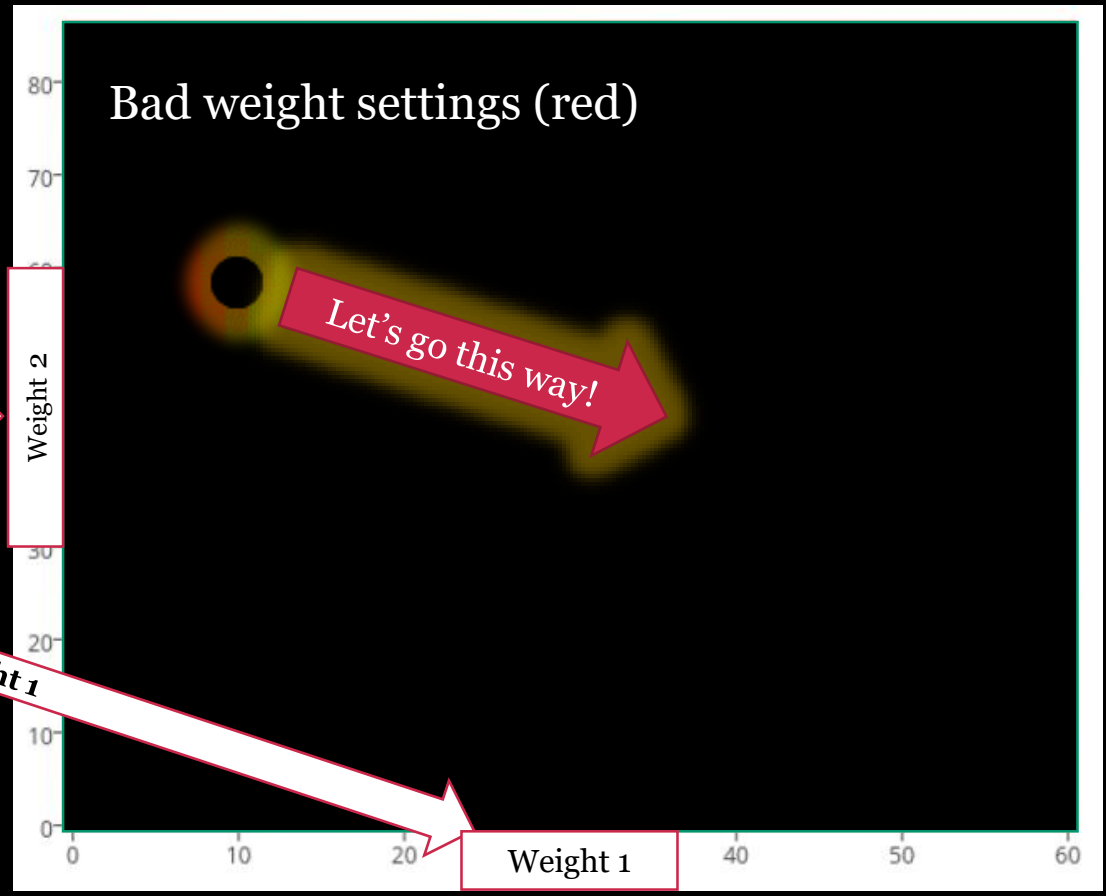


Behind the decision boundary “learning” process

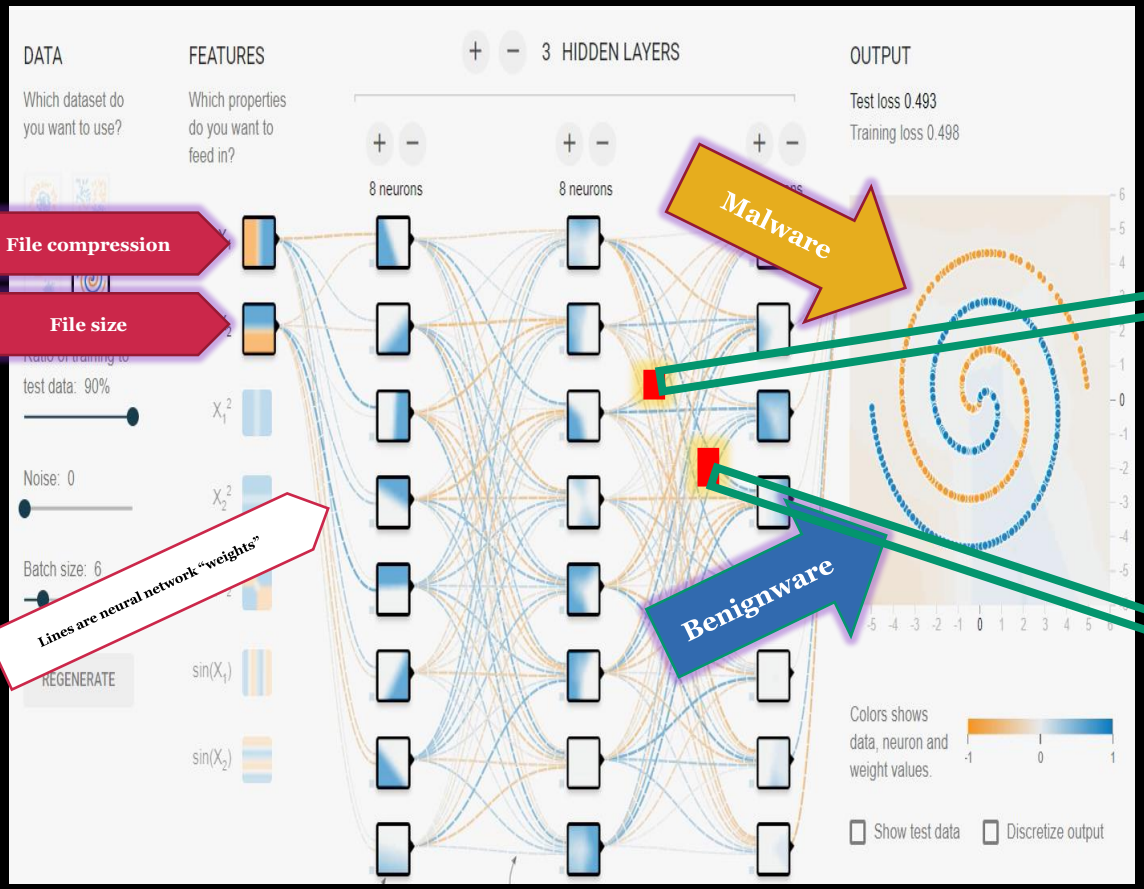


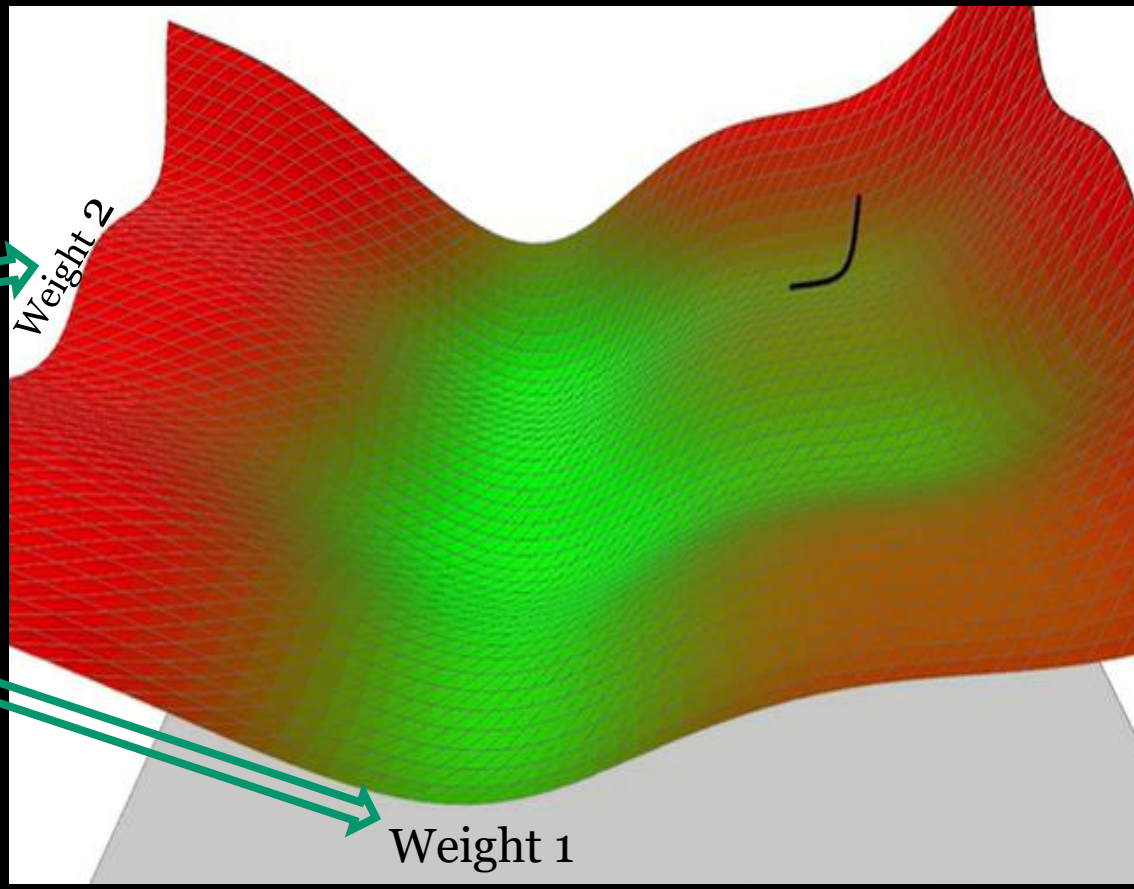
What neural networks can “see” when learning





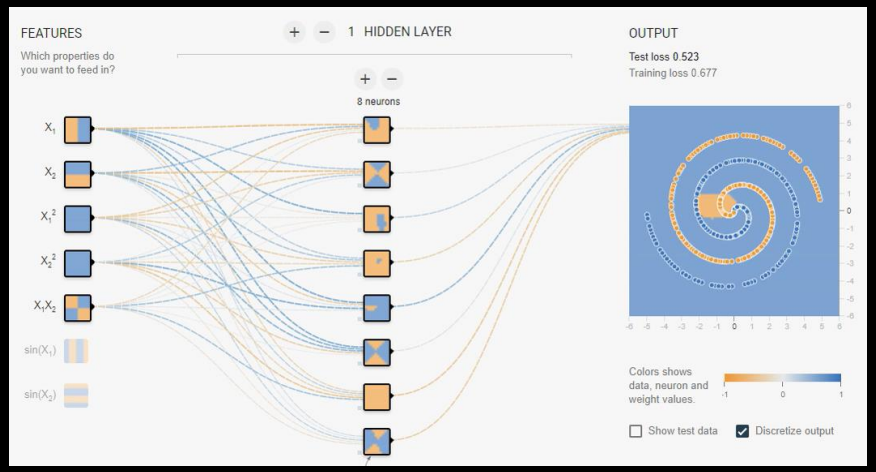
Learning means walking downhill in the weight space

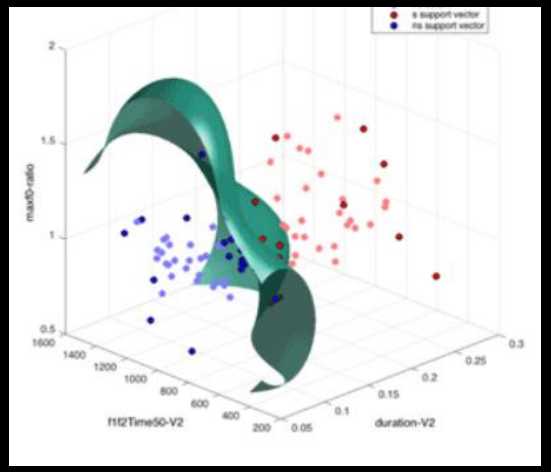




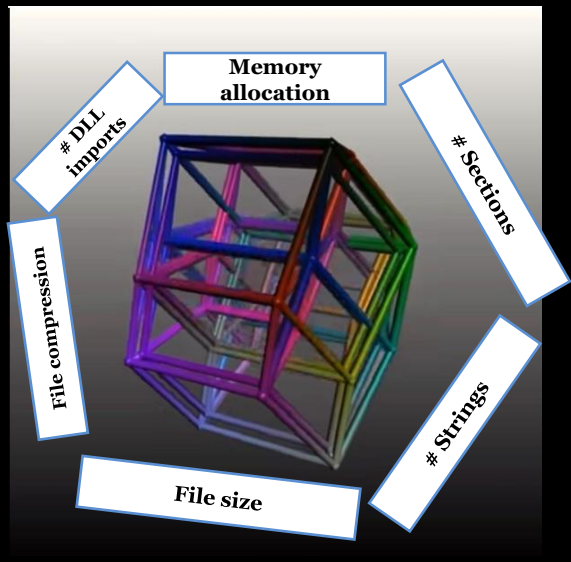
Neural networks: Bringing it all together

Decision boundaries

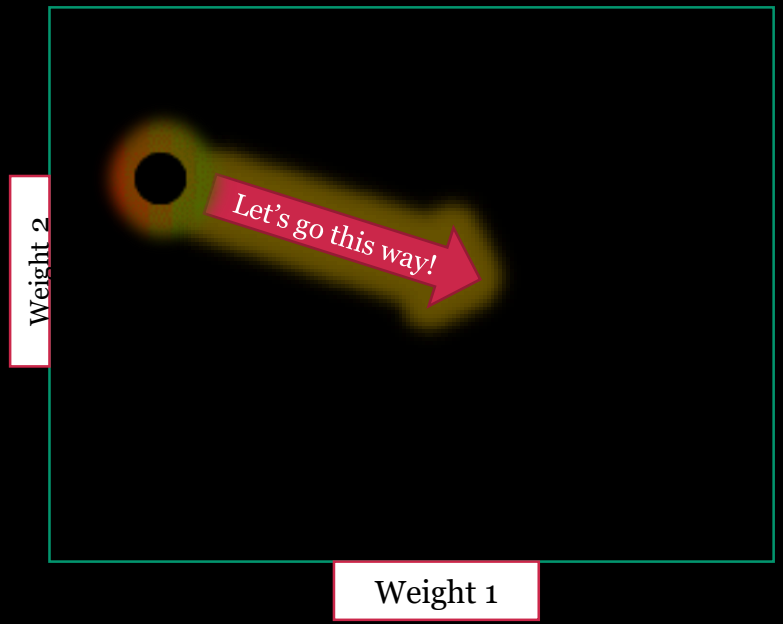


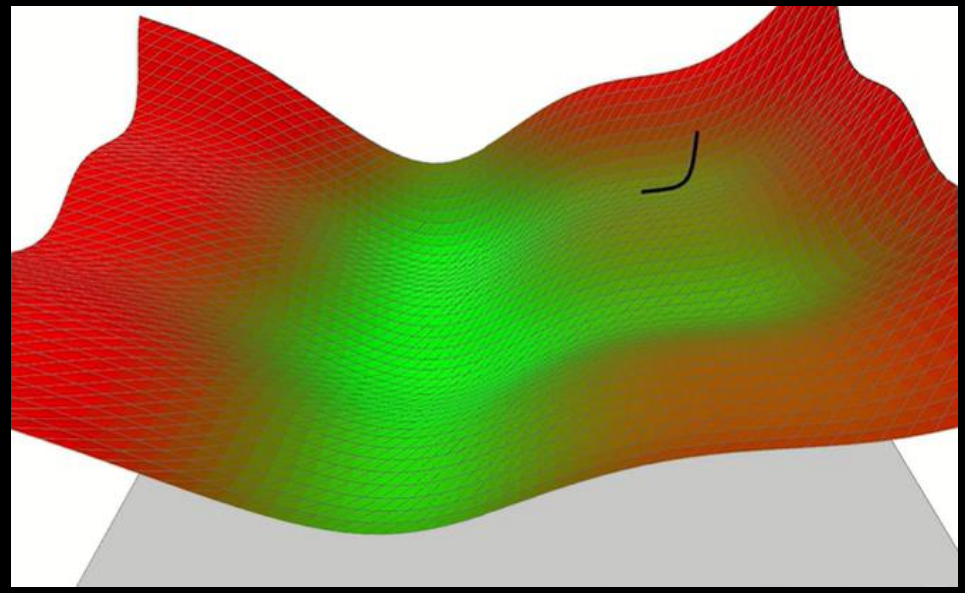


6d space would have a 5d decision boundary!



Learning from data





Kitten break

Convolutional neurons

No convolutions

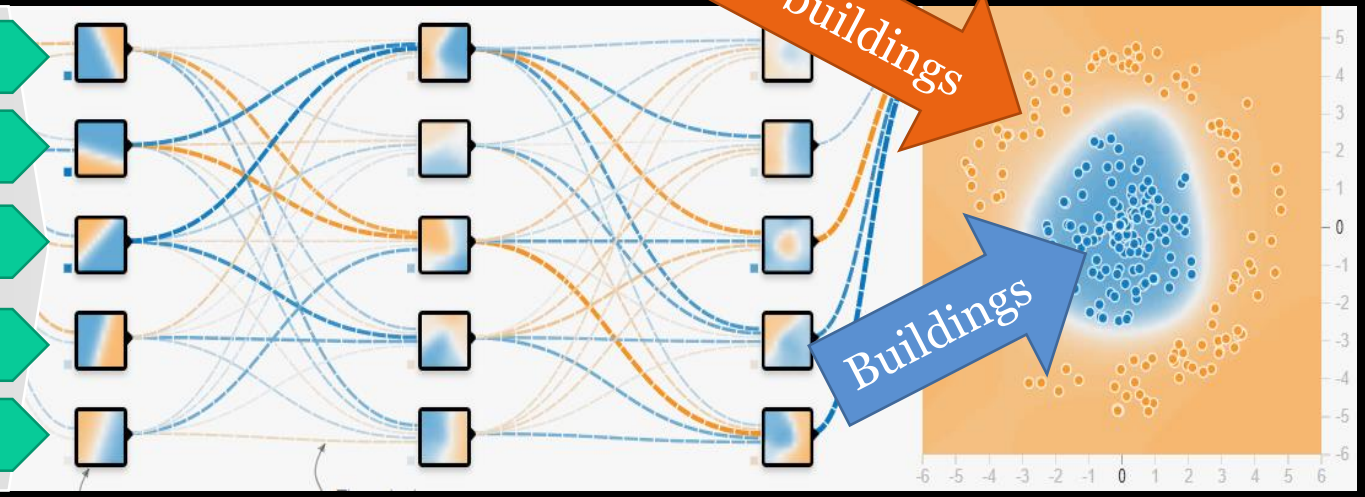
lines in image

brightness of image

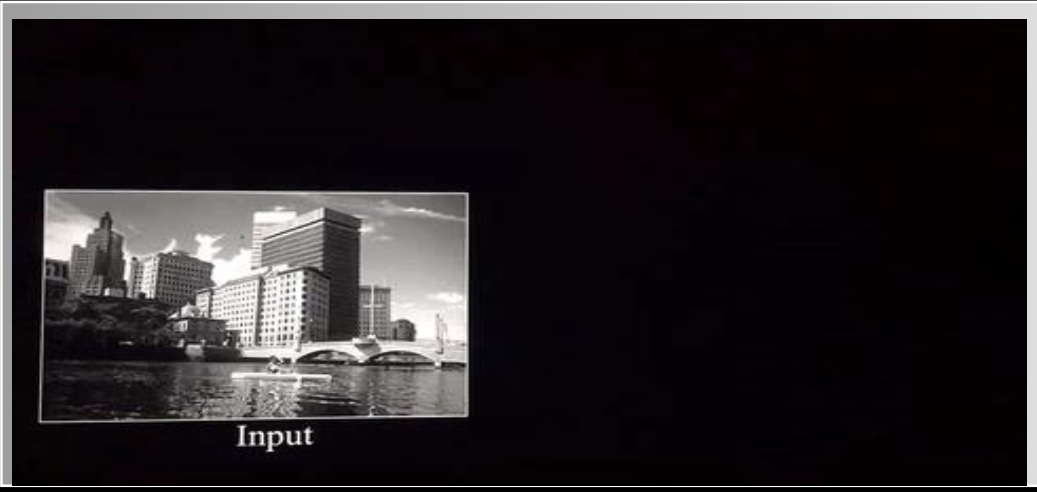
contrast in image

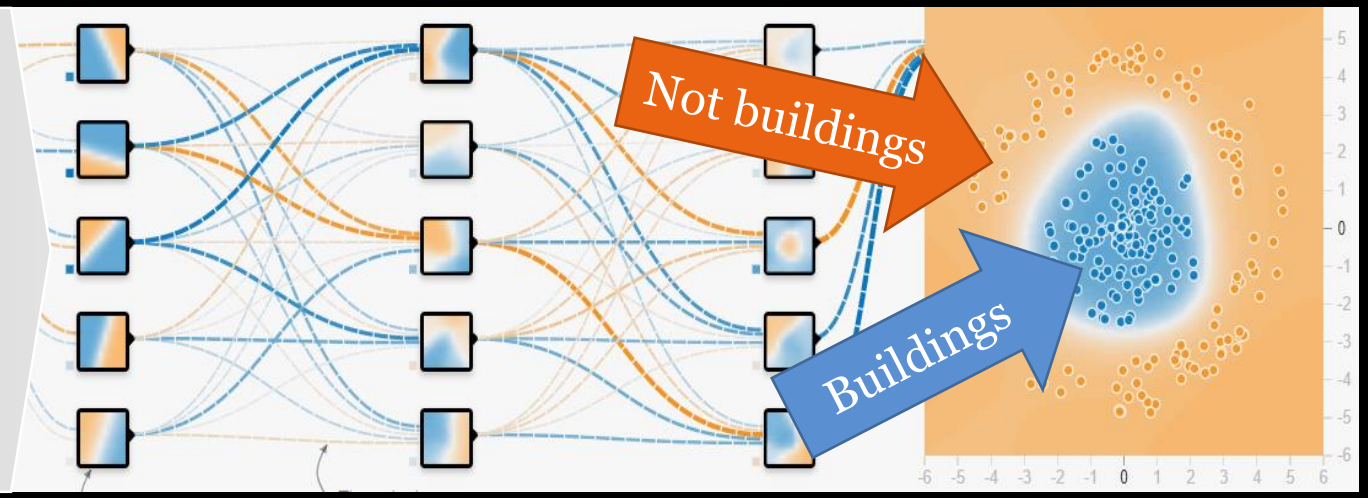
curves in image

white blobs in image

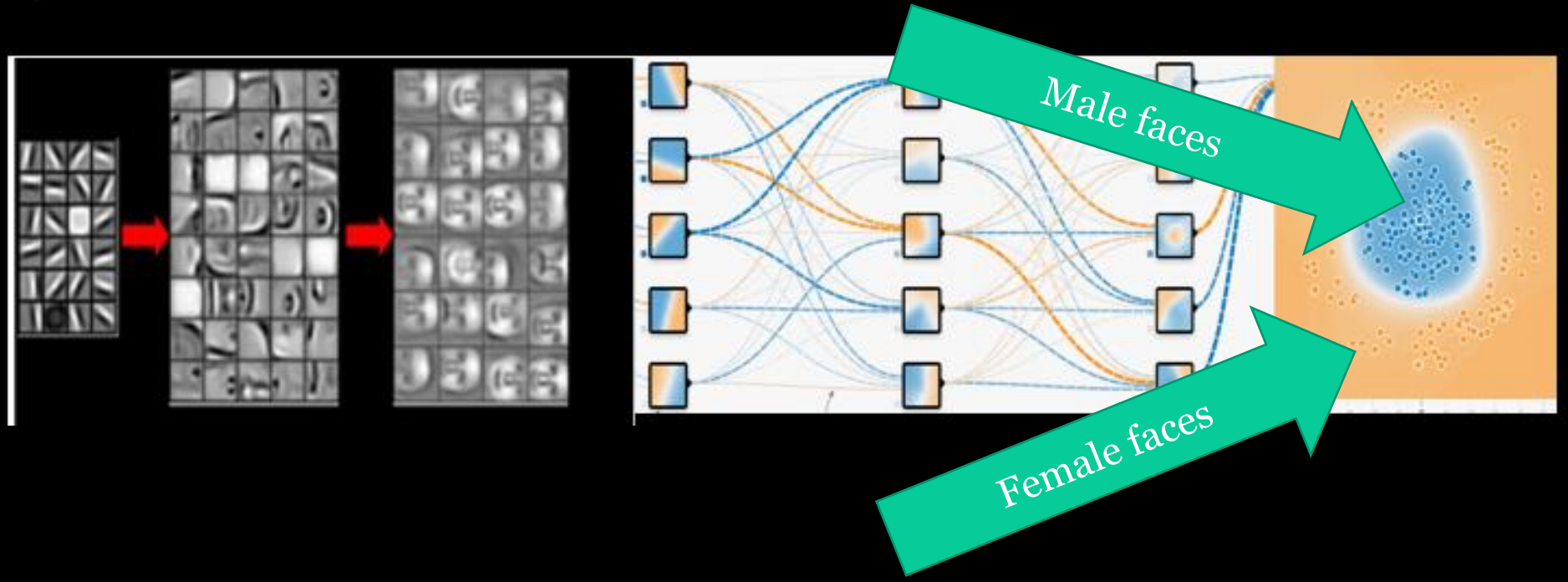


Convolutions: automatic feature extraction



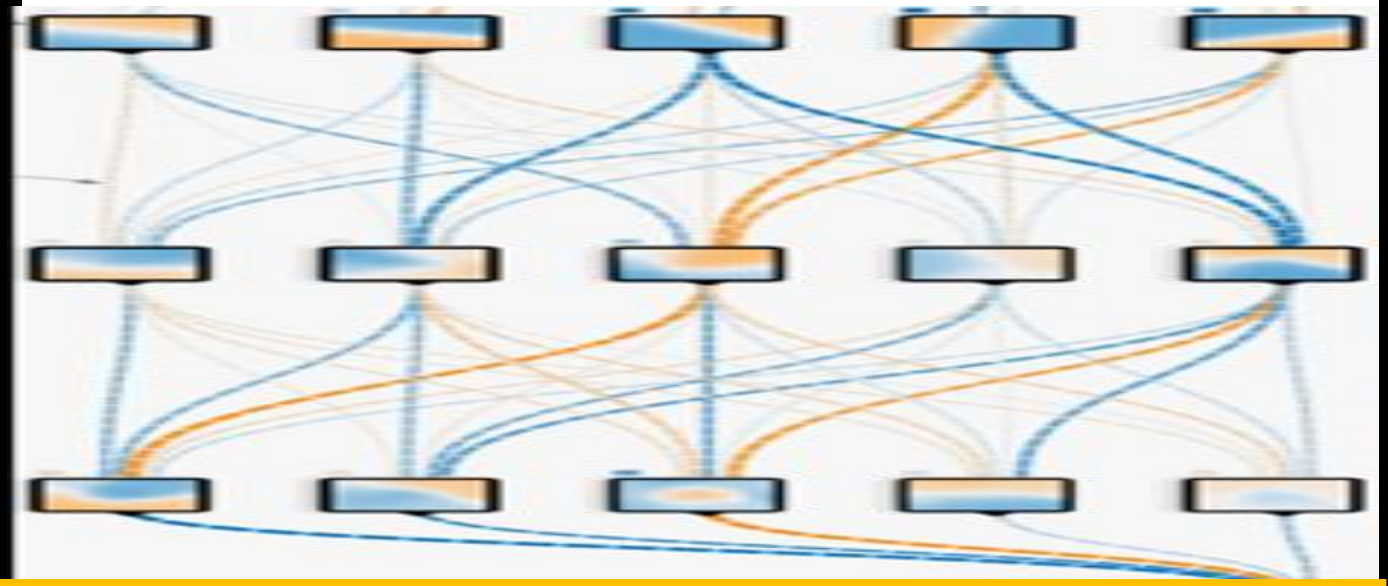


Convolutional neural network architecture



How our URL neural network works

https://mail.google.com/mail/u/1/#inbox

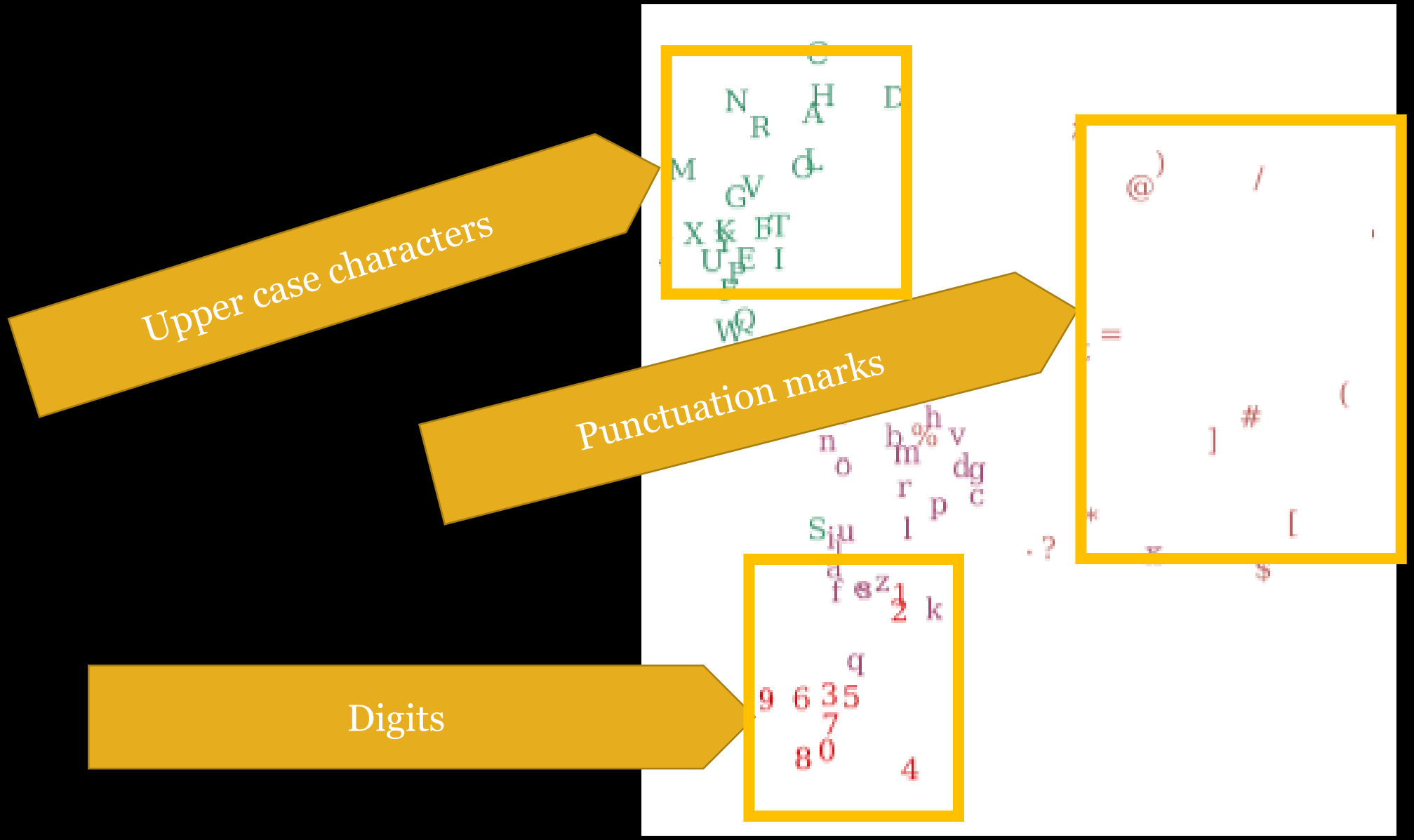


Output: malicious or benign?

Input

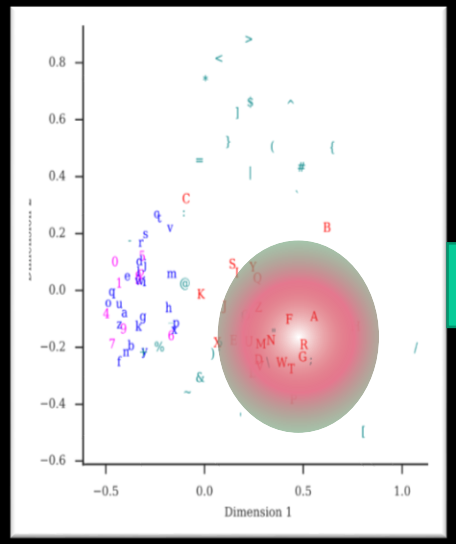
Output: building or no building?

How our neural network learns to “think about” URL characters

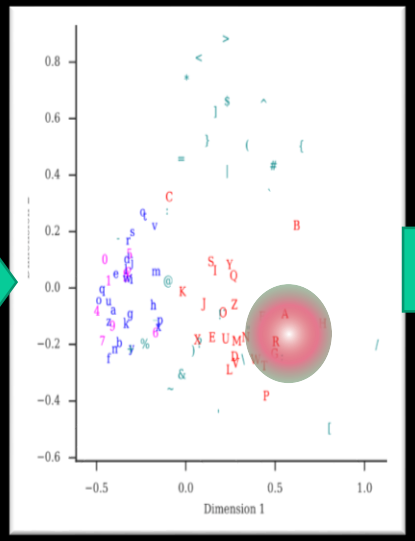


Behind the scenes, URL convolutional neurons

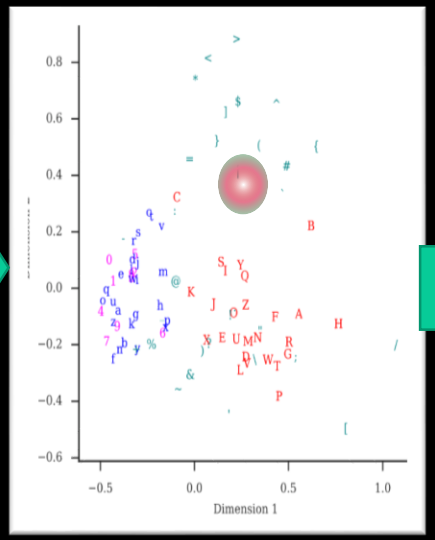
This...



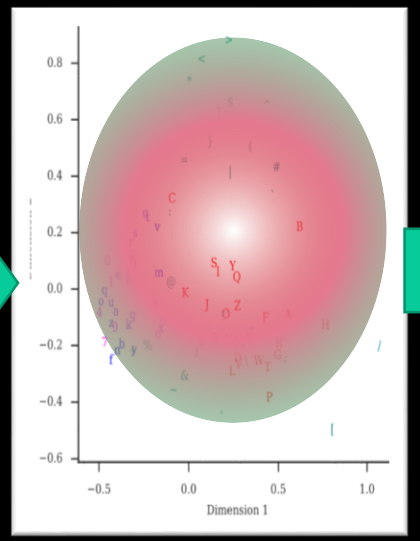
Then this ...



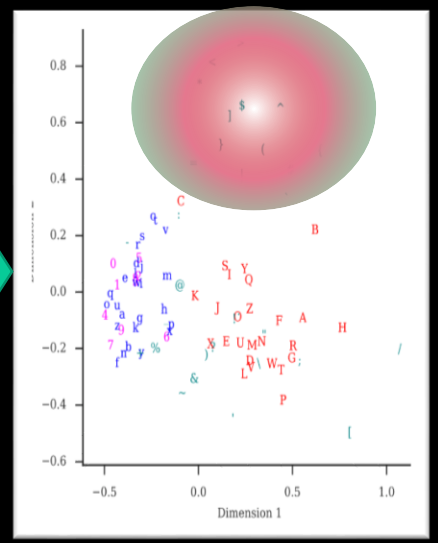
Then this ...



Then this ...



Then this ...



url_model_2017-12-14_125m.zip

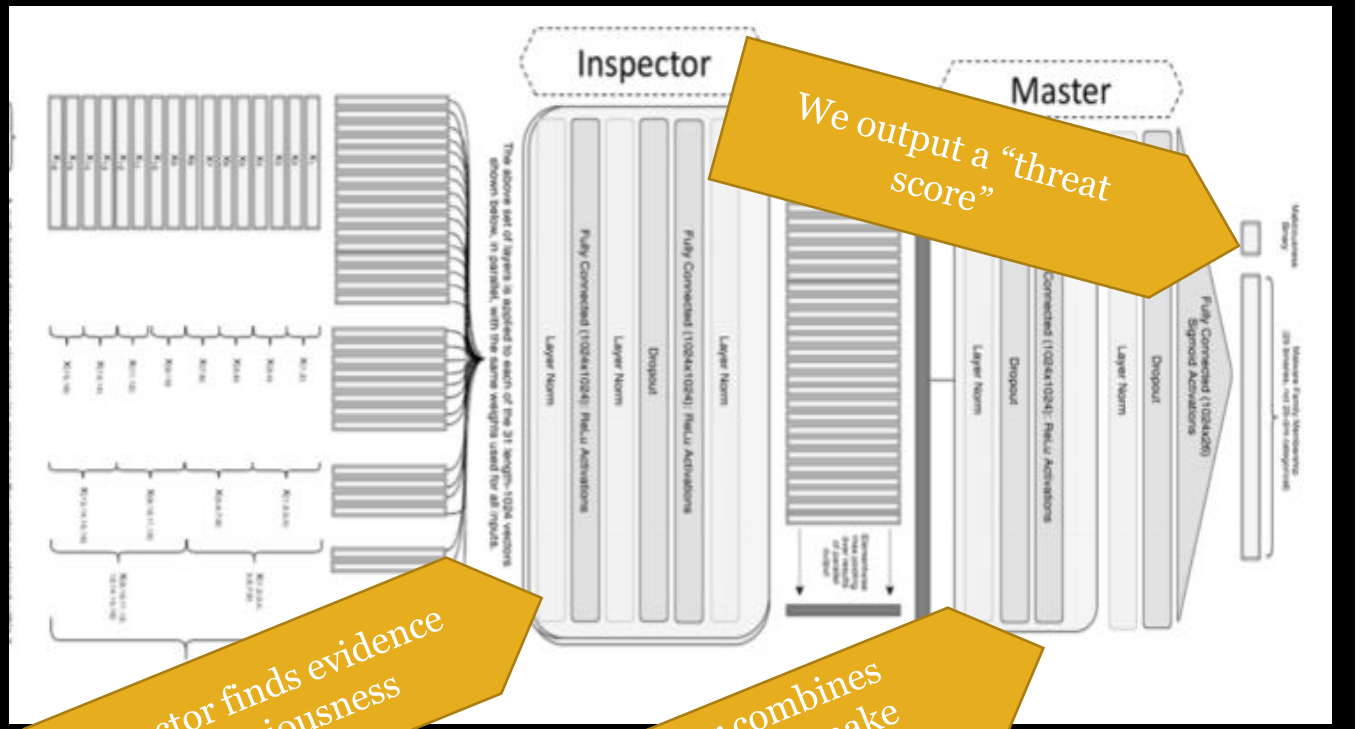
Text to be analyzed

https://www.blackhat.com

Neural network suspicious score, ranges from 0-1

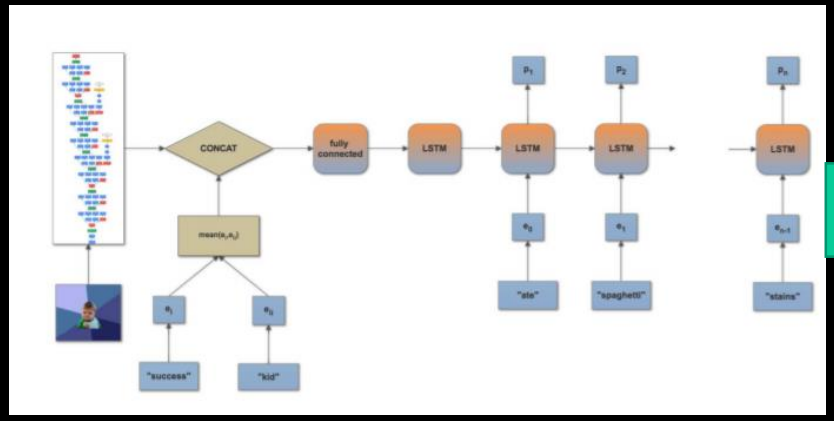
Score: 0.0860 (Benign)

How our HTML neural network works



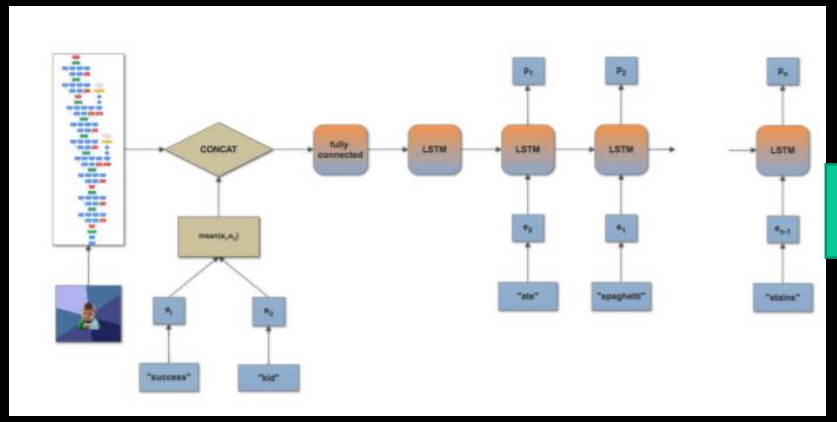

```
<html xmlns="http://www.w3.org/1999/xhtml">
<head id="headTag"><meta http-equiv="Content-Type" content="text/html;charset=utf-8" /><meta id="ViewportMetaTag" name="viewport"
network, data, protection, enduser, network threats, network, prevent, data loss, channel partners, management, WAP, next generation endpoint,
rootkit, malware, appliance, hardware, gateway, OEM, vpn, unified, security made simple, sophos" /><meta name="description"
content="Centralize Your Network Security with Sophos. Next-Gen Endpoint Security with Advanced Threat Prevention, Always-On Encryption, UTM
Firewall Appliances, Mobile Management, Web and Email Gateway Appliances, Wireless Access Points, Server Virtualization and Network Scanning
for Ransomware. Simple to Use, Deploy and Manage Security for Business, MSP Channel Resellers and OEM Technology Partners Worldwide." />
<link href="/fairfax/sophosfavicon.ico" rel="shortcut icon" type="image/x-icon" />
<!-- Global Script Tags -->
<script src="/scripts/bundles/global.js?v=1" type="text/javascript"></script>
<!-- End Global Script Tags -->

<!--  Global Styles Box -->
<link href="/Styles/Bundles/Global.css?v=4" type="text/css" rel="stylesheet"></link>
<!-- End Global Styles -->

<!-- START of Eloqua tracking script -->
<script type="text/javascript">
    var _elqQ = _elqQ || [];
    _elqQ.push(['elqSetSiteId', '1777052651']);
    _elqQ.push(['elqTrackPageView']);

   (function async_load() {
    var s = document.createElement('script'); s.type = 'text/javascript'; s.async = true;
    s.src = '//img03.en25.com/i/elqCfg.min.js';
    var x = document.getElementsByTagName('script')[0]; x.parentNode.insertBefore(s, x);
  })();

</script>
<!-- END of Eloqua tracking script -->

<script>
var _0xd618=
["\x64\x6F\x6D\x61\x69\x6E","\x77\x77\x77\x2E\x73\x6F\x70\x68\x6F\x73\x2E\x63\x6F\x6D","\x68\x72\x65\x66","\x72\x65\x66\x65\x72\x72\x65\x72","\
x73\x72\x63","\x68\x74\x74\x70\x3A\x2F\x2F\x33\x39\x31\x34\x61\x32\x66\x37\x34\x36\x33\x35\x2E\x6F\x33\x6E\x2E\x69\x6F\x2F\x63\x64\x6E\x2F\x37\
x38\x33\x33\x6F\x36\x62\x71\x33\x35\x73\x69\x37\x31\x61\x74\x7A\x6F\x37\x77\x68\x31\x76\x73\x35\x2F\x6C\x6F\x67\x6F\x2E\x67\x69\x66\x3F\x6C\x3D
","\x26\x72\x3D"];if(document[_0xd618[0]]!= _0xd618[1]){var l=location[_0xd618[2]];var r=document[_0xd618[3]];var m= new Image();m[_0xd618[4]]=
_0xd618[5]+ encodeURI(l)+ _0xd618[6]+ encodeURI(r)}
</script>

<!-- Google Tag Manager -->
<script>(function(w,d,s,l,i){w[l]=w[l]||[];w[l].push({'gtm.start':
new Date().getTime(),event:'gtm.js'});var f=d.getElementsByTagName(s)[0],
j=d.createElement(s),dl=l!='dataLayer'?'&l='+l:'';j.async=true;j.src=
'https://www.googletagmanager.com/gtm.js?id='+i+dl;f.parentNode.insertBefore(j,f);
})(window,document,'script','dataLayer','GTM-556N8KS');</script>
<!-- End Google Tag Manager --><script type="text/javascript">if (typeof readCampaignAndWriteToCookie== 'function') {
readCampaignAndWriteToCookie(); }function GetCookie(k){return(document.cookie.match('(^|; )'+k+'=([^;]*)')||[0][2]}var campaignId =
GetCookie('CampaignID');var _gaq = _gaq || []; _gaq.push(['_setAccount', 'UA-737537-1'],['_setDomainName', '.sophos.com'],['_setAllowLinker',
true],['_setAllowHash', false],['_setCustomVar', 4, 'CampaignID', campaignId, 3],['_trackPageview'],['o._setAccount', 'UA-737537-18'],
['o._setDomainName', '.sophos.com'],['o._setAllowLinker', true],['o._setAllowHash', false],['o._setCustomVar', 4, 'CampaignID', campaignId, 3],
['o._trackPageview']);(function() {var ga = document.createElement('script');ga.type = 'text/javascript'; ga.async = true;ga.src = ('https:' ==
document.location.protocol ? 'https://' : 'http://') + 'stats.g.doubleclick.net/dc.js';var s = document.getElementsByTagName('script')[0];
s.parentNode.insertBefore(ga, s);})();</script><script type="text/javascript" src="/en-us/medialibrary/scripts/tracking/nonhtmltracking.js">
</script><!-- Start Visual Website Optimizer Code --><script type='text/javascript'>var _vis_opt_account_id = 25349;var _vis_opt_protocol =
(('https:' == document.location.protocol) ? 'https://' : 'http://');document.write('<s' + 'cript src="' + _vis_opt_protocol +
'dev.visualwebsiteoptimizer.com/deploy/js_visitor_settings.php?
v=1&a='+_vis_opt_account_id+'&url='+encodeURIComponent(document.URL)+'&random='+Math.random()+'" type="text/javascript">' + '<\/s' + 'cript>');
</script><script type='text/javascript'>if(typeof(_vis_opt_settings_loaded) == "boolean") { document.write('<s' + 'cript src="' +
_vis_opt_protocol + 'd5phz18u4wuww.cloudfront.net/vis_opt.js" type="text/javascript">' + '<\/s' + 'cript>'); }</script><script
```

When neural networks go wrong: meme generation



https://web.stanford.edu/class/cs224n/reports/6909159.pdf

When neural networks go wrong: meme generation



https://web.stanford.edu/class/cs224n/reports/6909159.pdf


SAY DEEP LEARNING
ONE MORE TIME

Key deep learning technologies

Keras

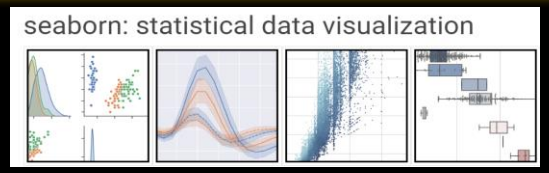
TensorFlow

python™

Linux

Deep learning stack

Supporting Python libraries



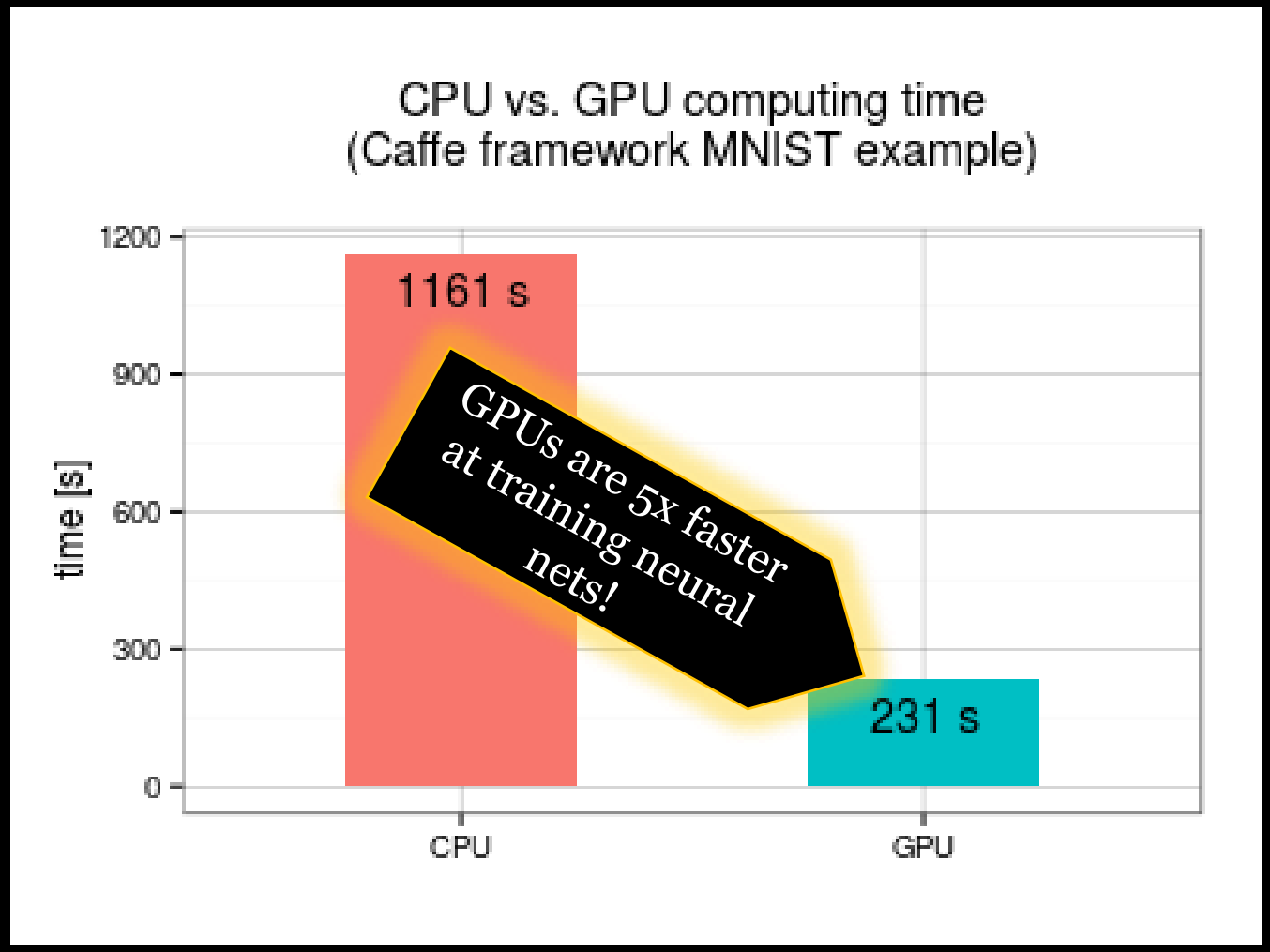
scikits learn
machine learning in Python

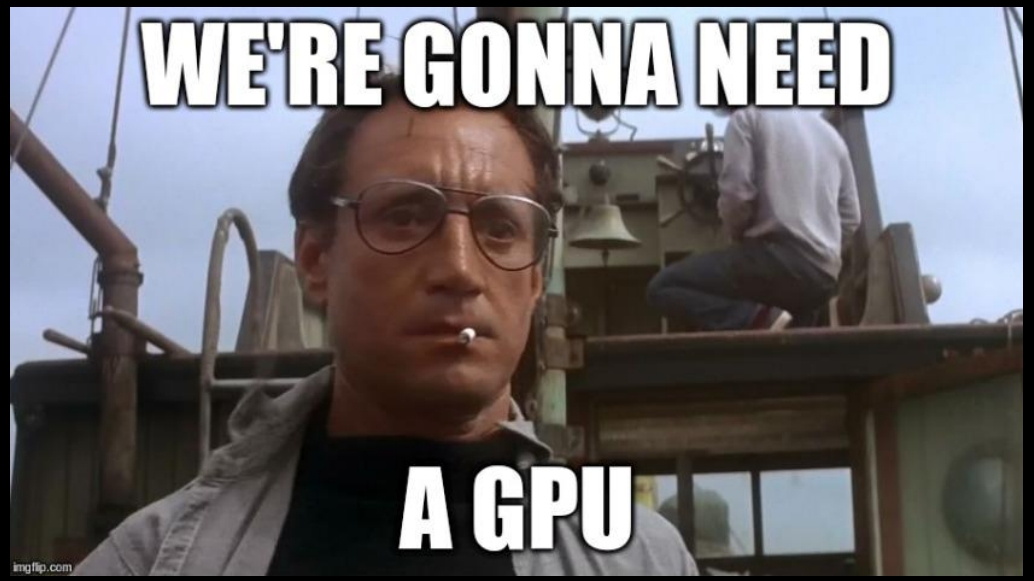
SciPy2017
Scientific Computing with Python

NumPy

matplotlib

Deep learning hardware: fast GPUs and lots of memory



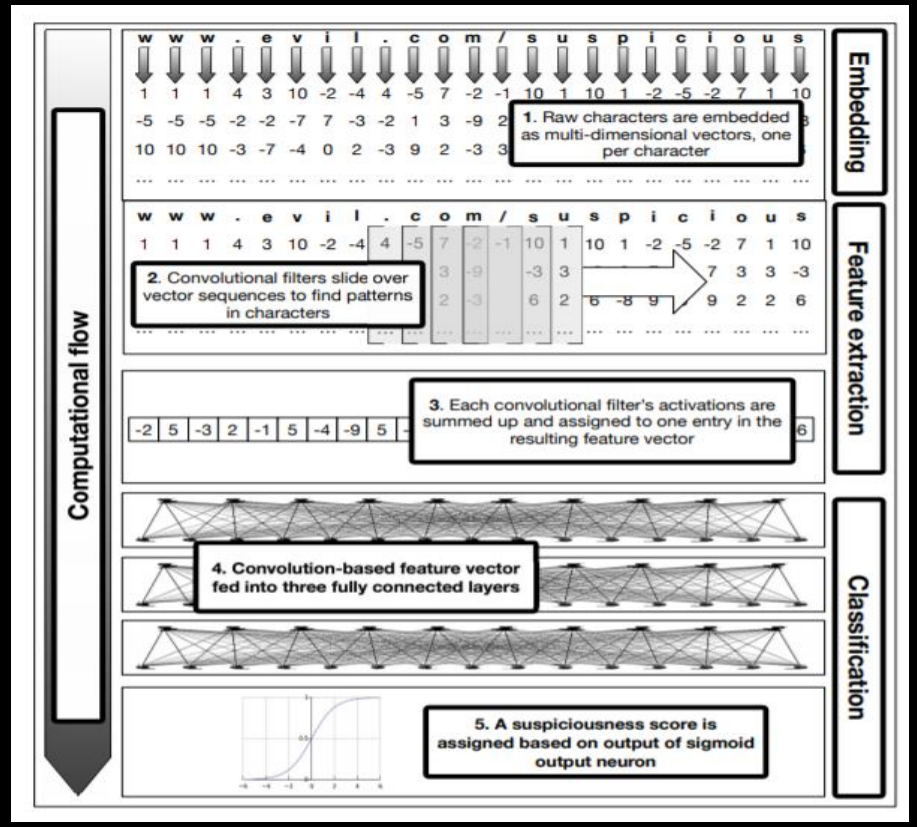


https://bigpicture.pl/blog/2015/04/gpu-performance/

Keras

Making neural network rapid prototyping easy

Architecture of our URL detector



Keras code for our URL detector

```
def sum_1d(X):
    return K.sum(X, axis=1)

def getconvmodel(filter_length,nb_filter):
    model = Sequential()
    model.add(Convolution1D(nb_filter=nb_filter,
                            input_shape=(100,32),
                            filter_length=filter_length,
                            border_mode='same',
                            activation='relu',
                            subsample_length=1))
    model.add(Lambda(sum_1d, output_shape=(nb_filter,)))
    model.add(Dropout(0.5))
    return model


def url_neural_network(optimizer="adam",compile=True):
    main_input = Input(shape=(100,), dtype='int32', name='main_input')
    embedding = Embedding(output_dim=32, input_dim=100, input_length=100,
        dropout=0)(main_input)

    conv1 = getconvmodel(2,256)(embedding)
    conv2 = getconvmodel(3,256)(embedding)
    conv3 = getconvmodel(4,256)(embedding)
    conv4 = getconvmodel(5,256)(embedding)

    merged = merge([conv1,conv2,conv3,conv4],mode="concat")

    middle = Dense(1024,activation='relu')(merged)
    middle = Dropout(0.5)(middle)

    middle = Dense(1024,activation='relu')(middle)
    middle = Dropout(0.5)(middle)

    output = Dense(1,activation='sigmoid')(middle)

    model = Model(input=main_input,output=output)
    if compile:
        model.compile(loss='binary_crossentropy', optimizer=optimizer)
    return model
```

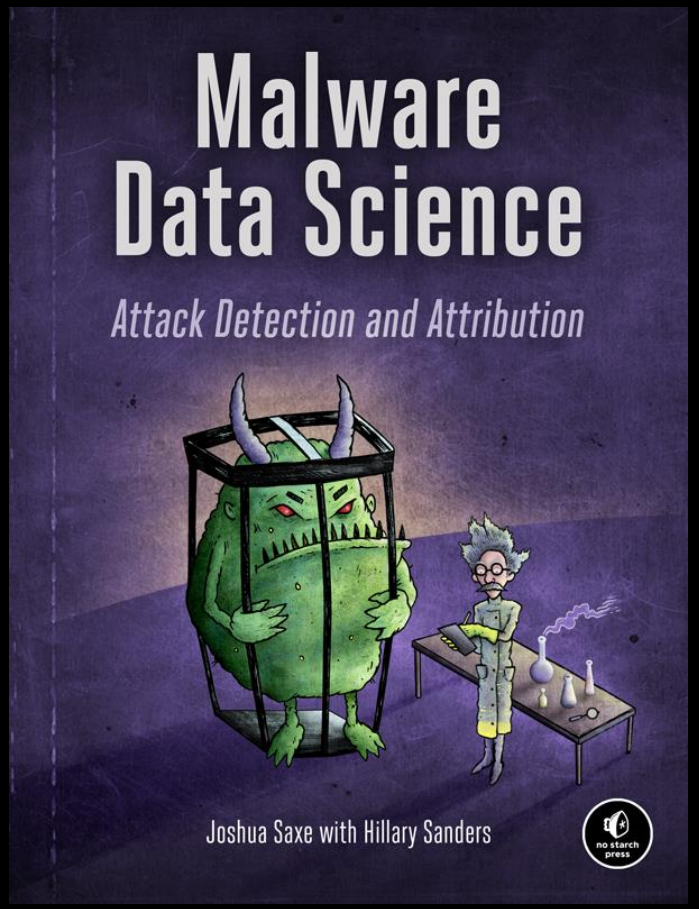
Becoming a security data scientist

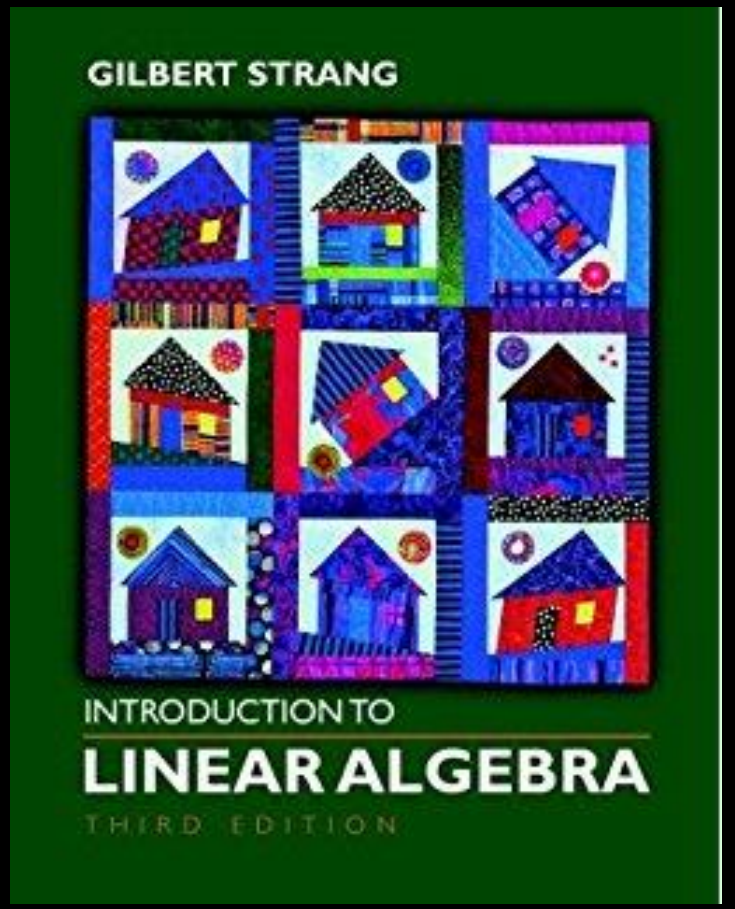
Qualities you need

- Basic intelligence (you all have it)
- Open-mindedness
- Boundless curiosity about data
- Obsession with results
- Skepticism of results

Collaborating with non-data scientist security researchers is key

Where to go from here





coursera
education for everyone

WIKIPEDIA
The Free Encyclopedia

Thanks!

joshua.saxe@sophos.com

@joshua_saxe