Combating spam

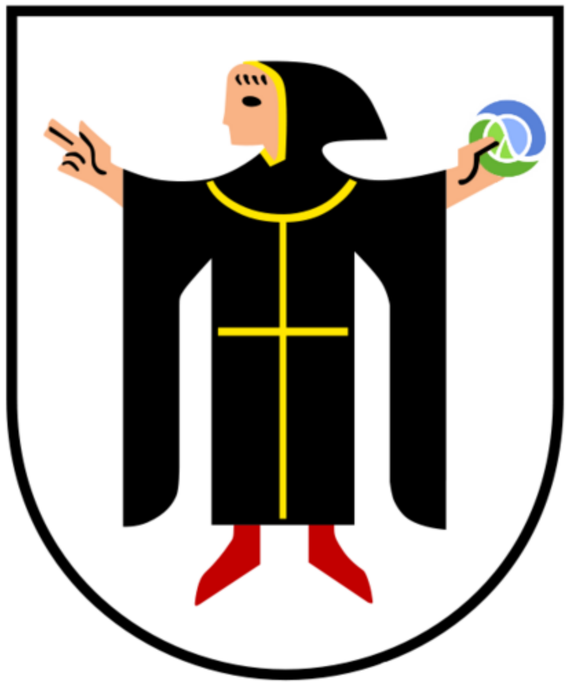
 or 

how I befriended the Killer Rabbit of Caerbannog

with **Jan Stępień**

@janstepien

@innoQ



@cljmuc


333-A
NO JUNK MAIL
333
NO
JUNK
MAIL

Messaging, Malware and Mobile Anti-Abuse Working Group

M³AAWG Email Metrics Program:
The Network Operators' Perspective

Report #16 – 1st Quarter 2012 through 2nd Quarter 2014
(Issued November 2014)

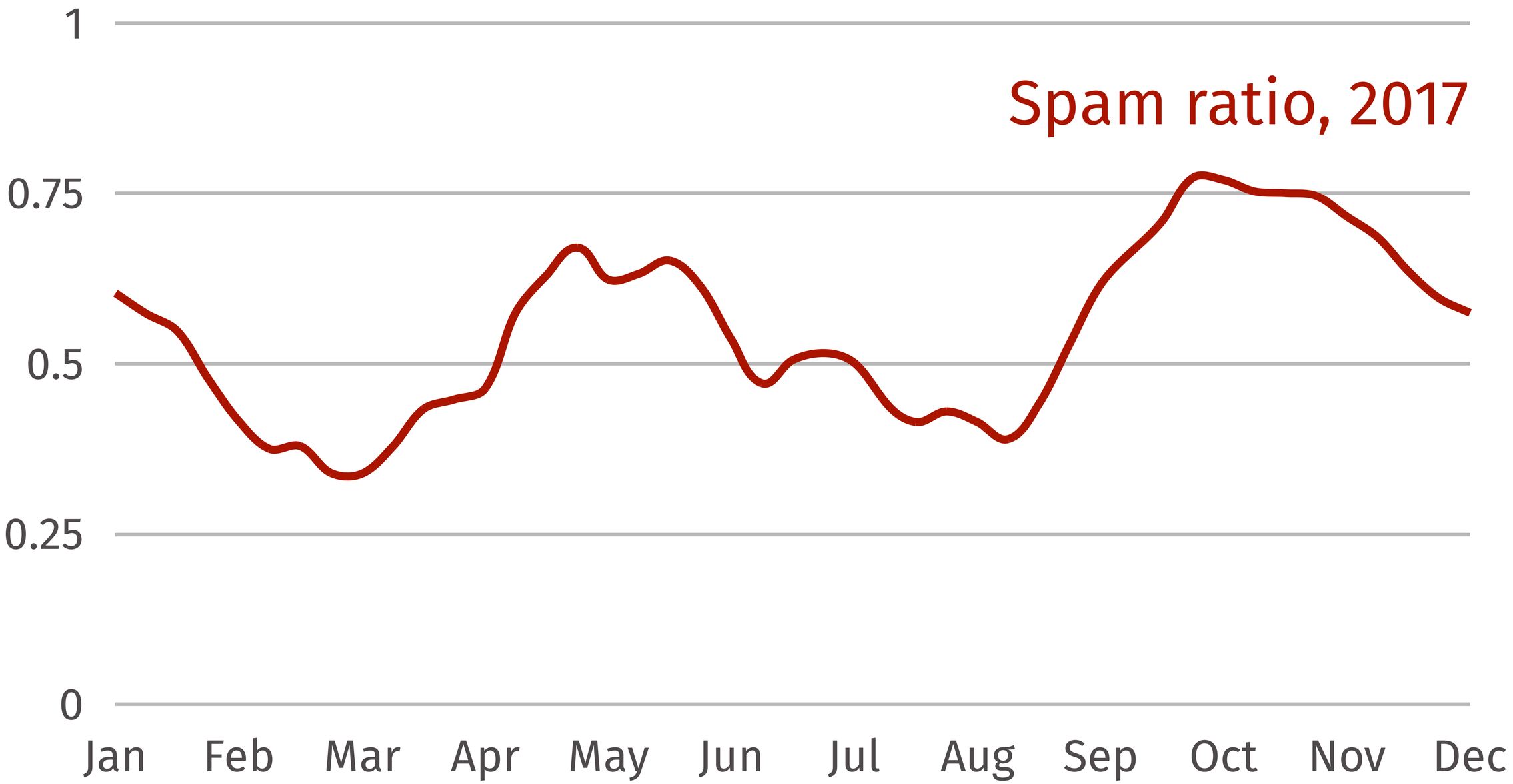
Executive Summary

This is the sixteenth report in the M³AAWG Email Metrics Program incorporating m[illegible] da[illegible] all of 2012 through the second quarter of 2014, with aggregated metrics from January 2012 through June 2014. For this combined 30-month period, abusive email remained fairly consistent, ranging from about 87.1% to 90.2%. This ongoing high rate of abusive messaging is a reminder that the industry must remain vigilant to disruptive online activities and continue its cooperative efforts to provide users a stable messaging environment.

About the M³AAWG Email Metrics Program

At the request of government and private agencies worldwide, the Messaging, Malware and Mobile Anti-Abuse Working Group (M³AAWG) developed these quarterly reports as a guide to understanding the effectiveness of the industry's efforts in obstructing abusive emails before they reach users. The metrics data reported here is provided voluntarily and confidentially by Internet service providers, network operators and email providers that have come together in M³AAWG to work against online abuse. However, M³AAWG members are under no obligation to supply

jan@stepien.cc


Spam ratio, 2017
1
0.75
0.5
0.25
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

Spam filtering
is an arms race

Solved since 2014*

* *for now*

```
Return-Path: <FlirtLife@happysillyfeetman.net>
X-Original-To: jan@stepien.cc
Delivered-To: jan@stepien.cc
Received: from ktxd8z.happysillyfeetman.net (ktxd8z.happysillyfeetman.net [213.5.68.131])
    by r245-52.iq.pl (Postfix) with ESMTP id DCE364A61FC4
    for <jan@stepien.cc>; Wed, 14 Oct 2015 08:46:12 +0200 (CEST)
Received: from 05f541e2.ktxd8z.happysillyfeetman.net (amavisd, port 7307)
    by ktxd8z.happysillyfeetman.net with ESMTP id 05LVF541OGE2;
    for <jan@stepien.cc>; Tue, 13 Oct 2015 23:46:07 -0700
To: <jan@stepien.cc>
Date: Tue, 13 Oct 2015 23:46:07 -0700
Message-ID: <530723999953959530767719911262707@ktxd8z.happysillyfeetman.net>
From: "FlirtLife" <FlirtLife@happysillyfeetman.net>
Subject: ATTN: You have (1) New Message
Content-Language: en-us
MIME-Version: 1.0
Content-Transfer-Encoding: 8bit
Content-Type: multipart/alternative;
    boundary="-----=Part.592.3860.1444805167"

------=Part.592.3860.1444805167
Content-Transfer-Encoding: 8bit
Content-Type: text/plain; charset="UTF-8"

You received a new message on Flirtlocal.com

From: adorableAira

I need a spark in my sexual life...i want someone who can...
```

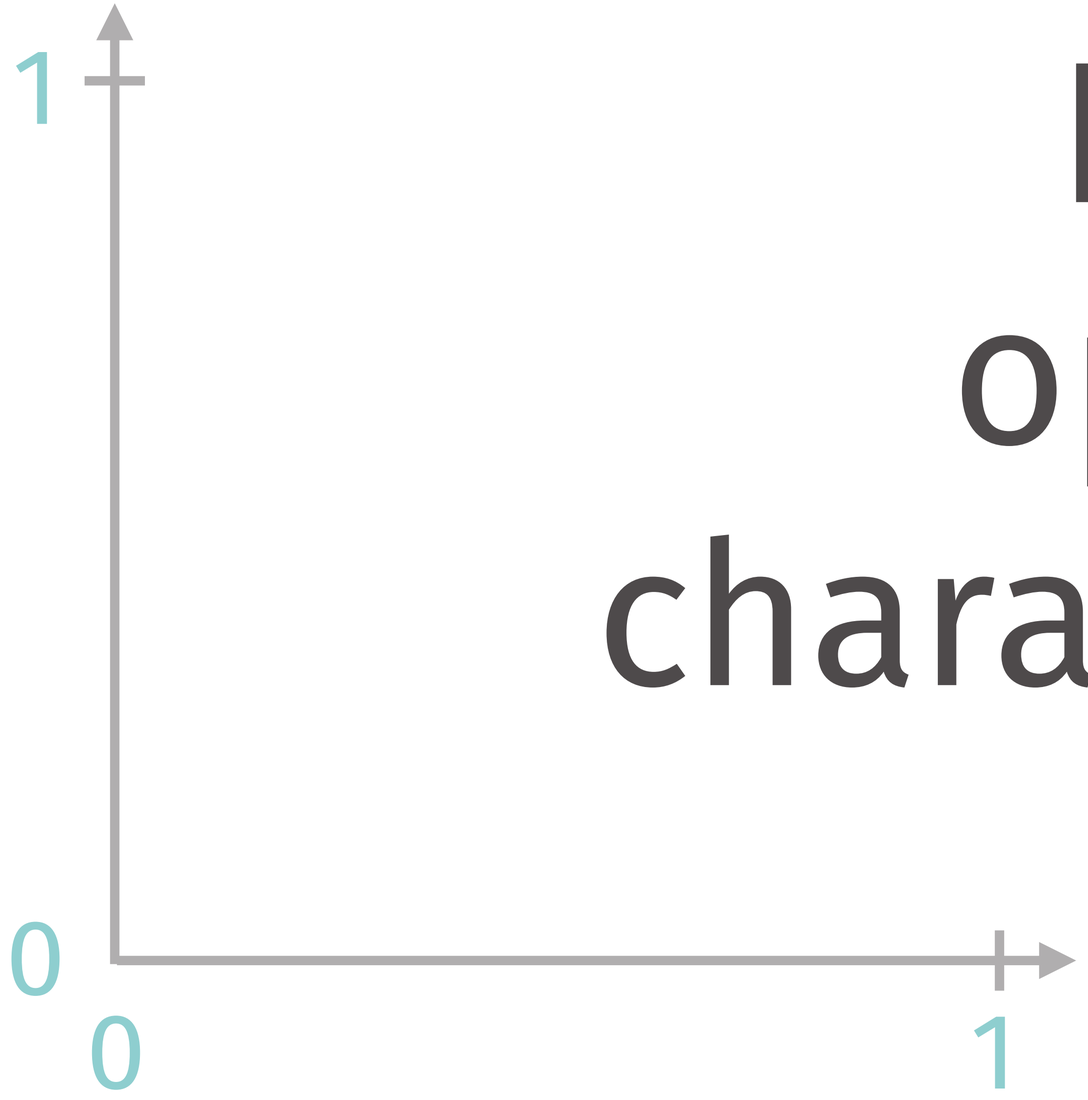
a	s	0.7
b	s	0.4
c	h	0.2

a	s	0.7	+	
b	s	0.4	+	$t = 0.1$
c	h	0.2	+	

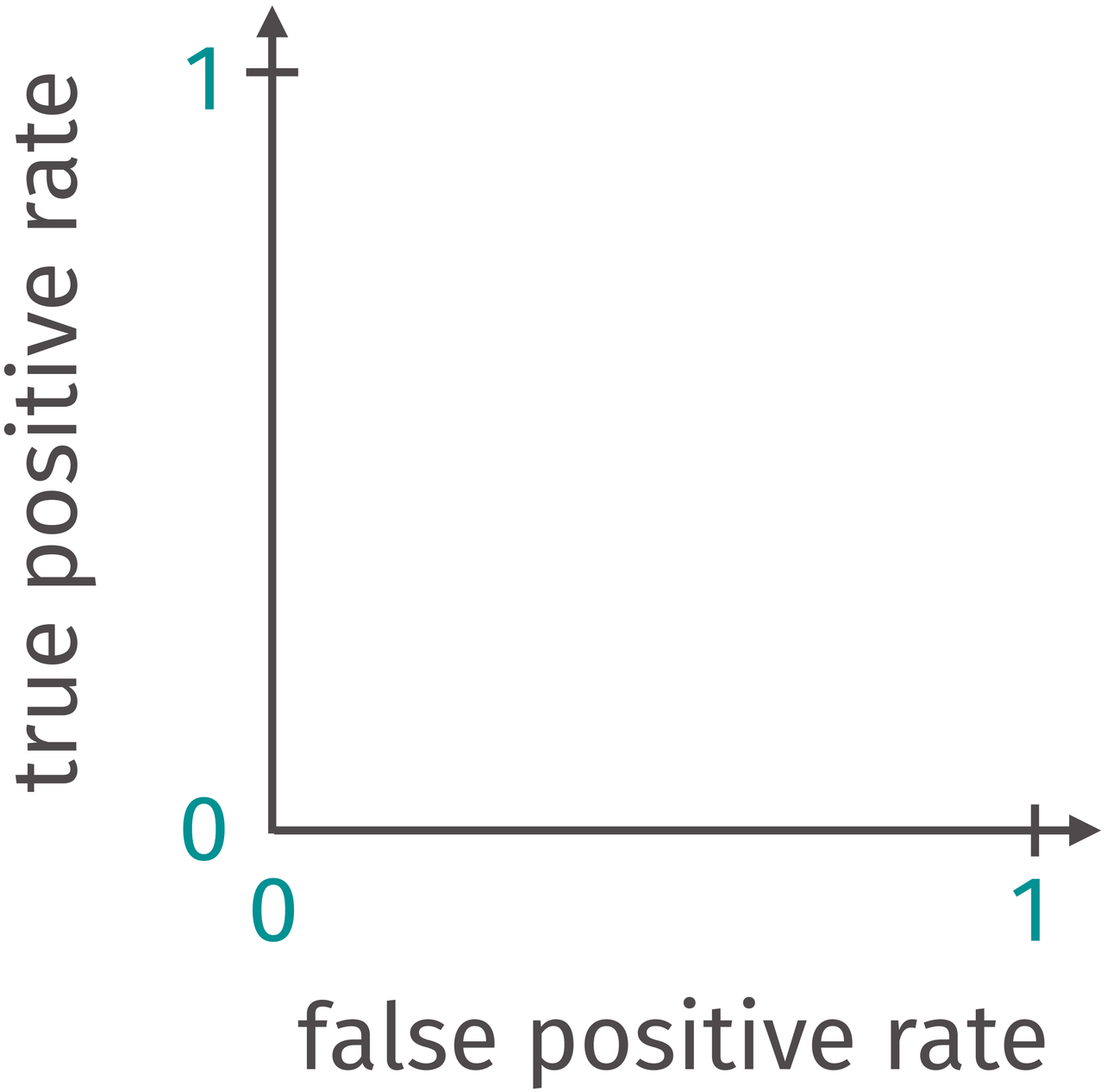
a	s	0.7	+	
b	s	0.4	+	$t = 0.3$
c	h	0.2	−	

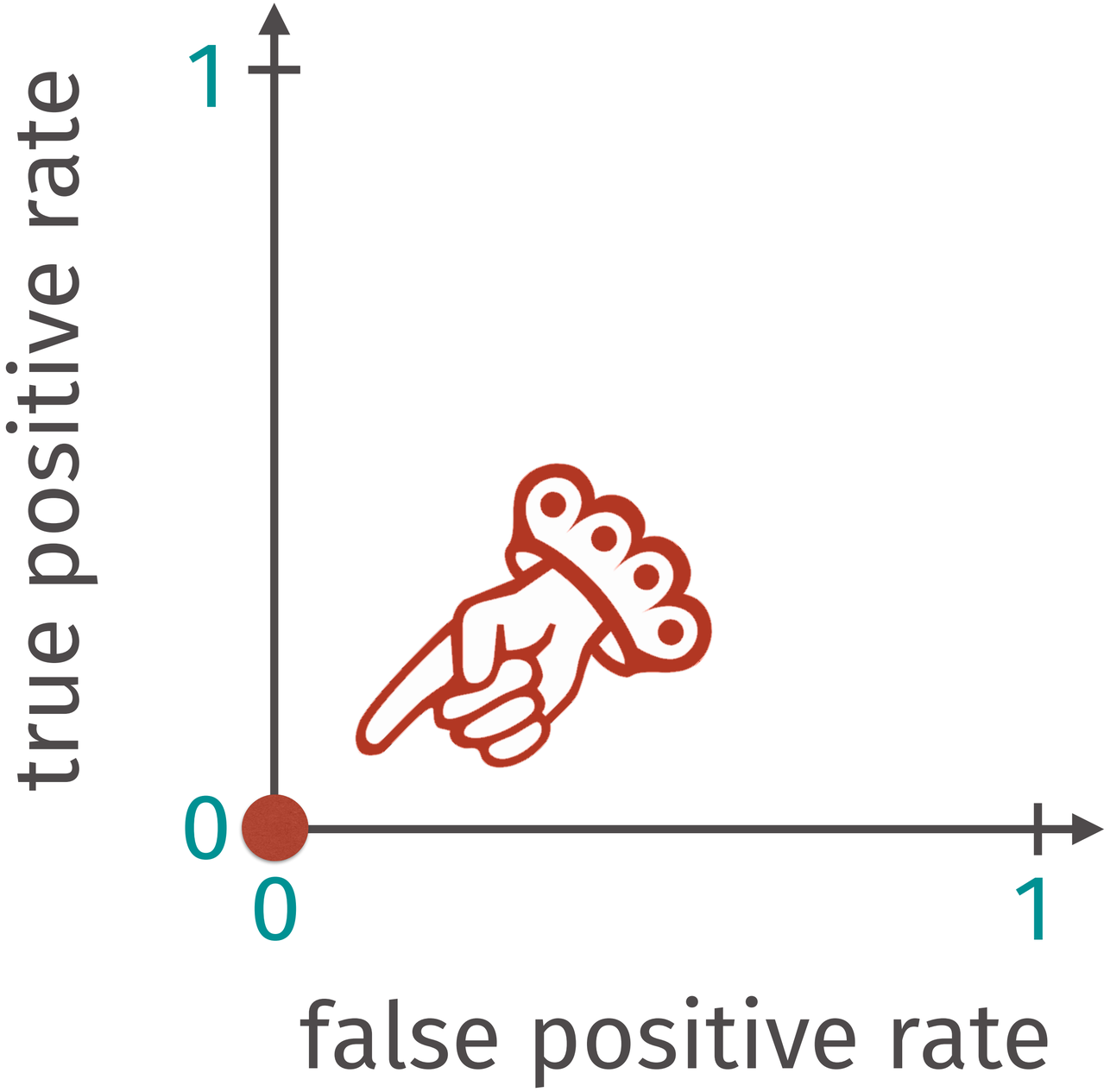
a	s	0.7	+	
b	s	0.4	−	$t = 0.5$
c	h	0.2	−	

a	s	0.7	–	$t = 0.9$
b	s	0.4	–	
c	h	0.2	–	

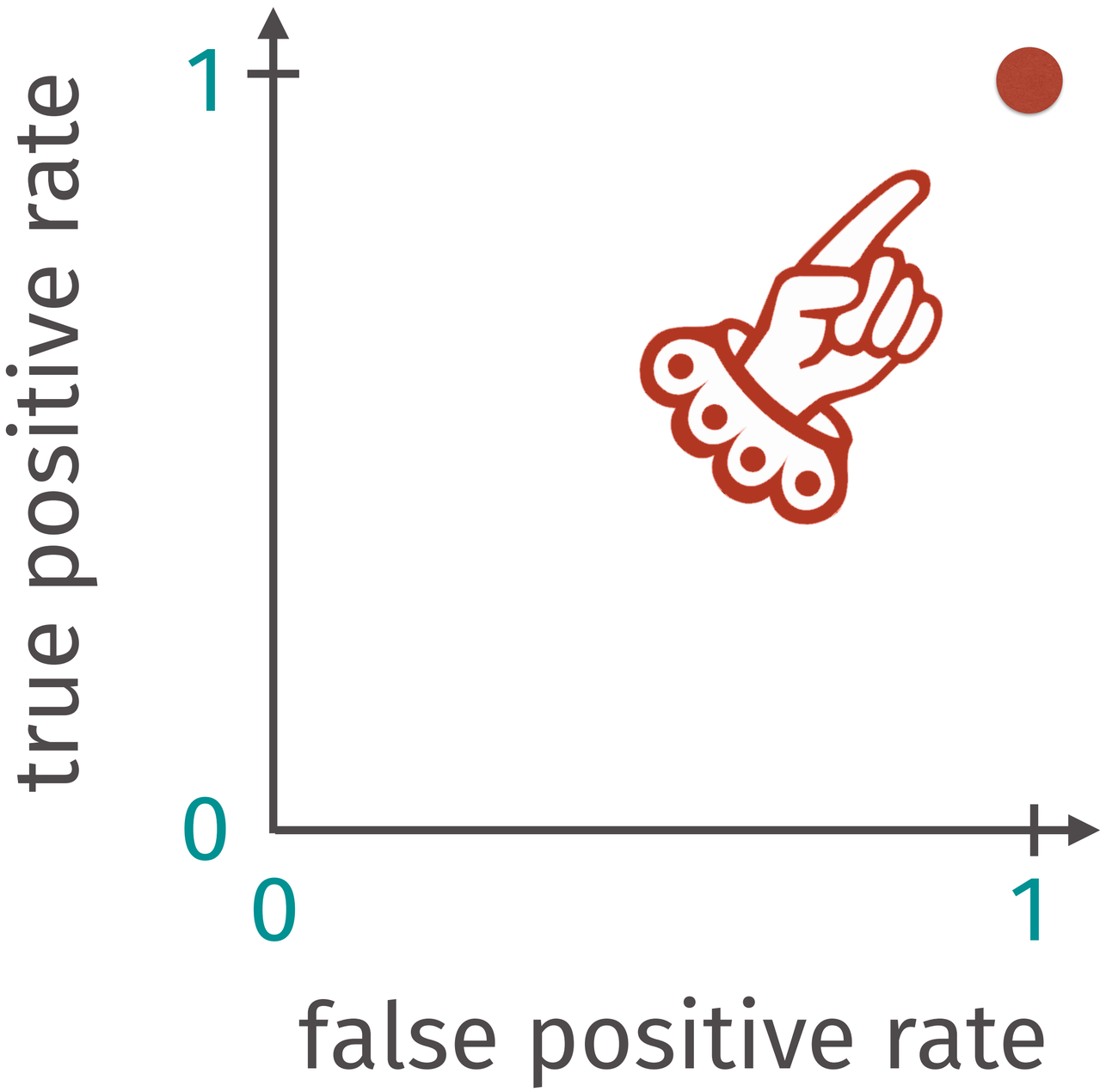


Receiver operating characteristic curve

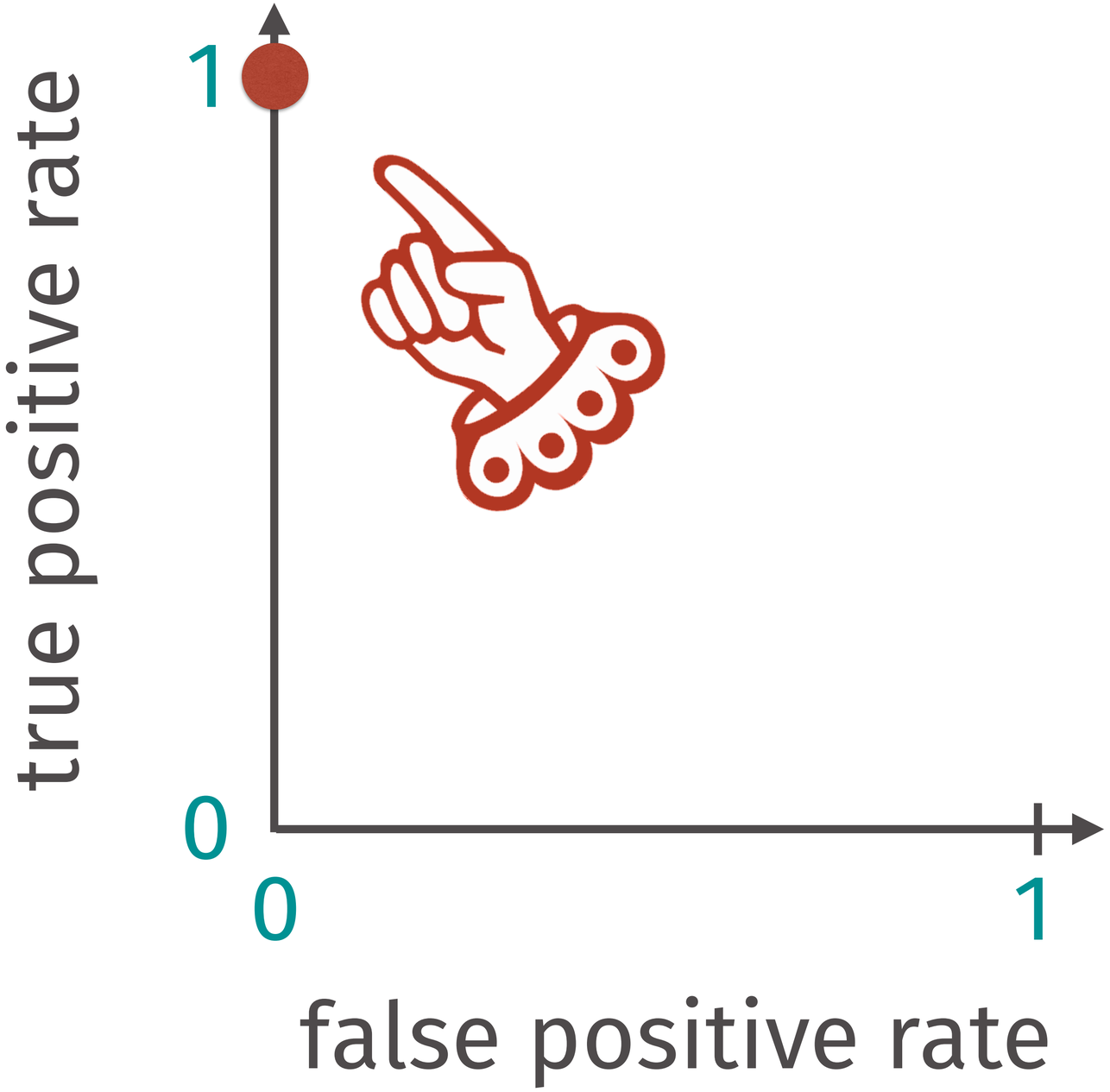

true positive rate
1
0
0
1
false positive rate



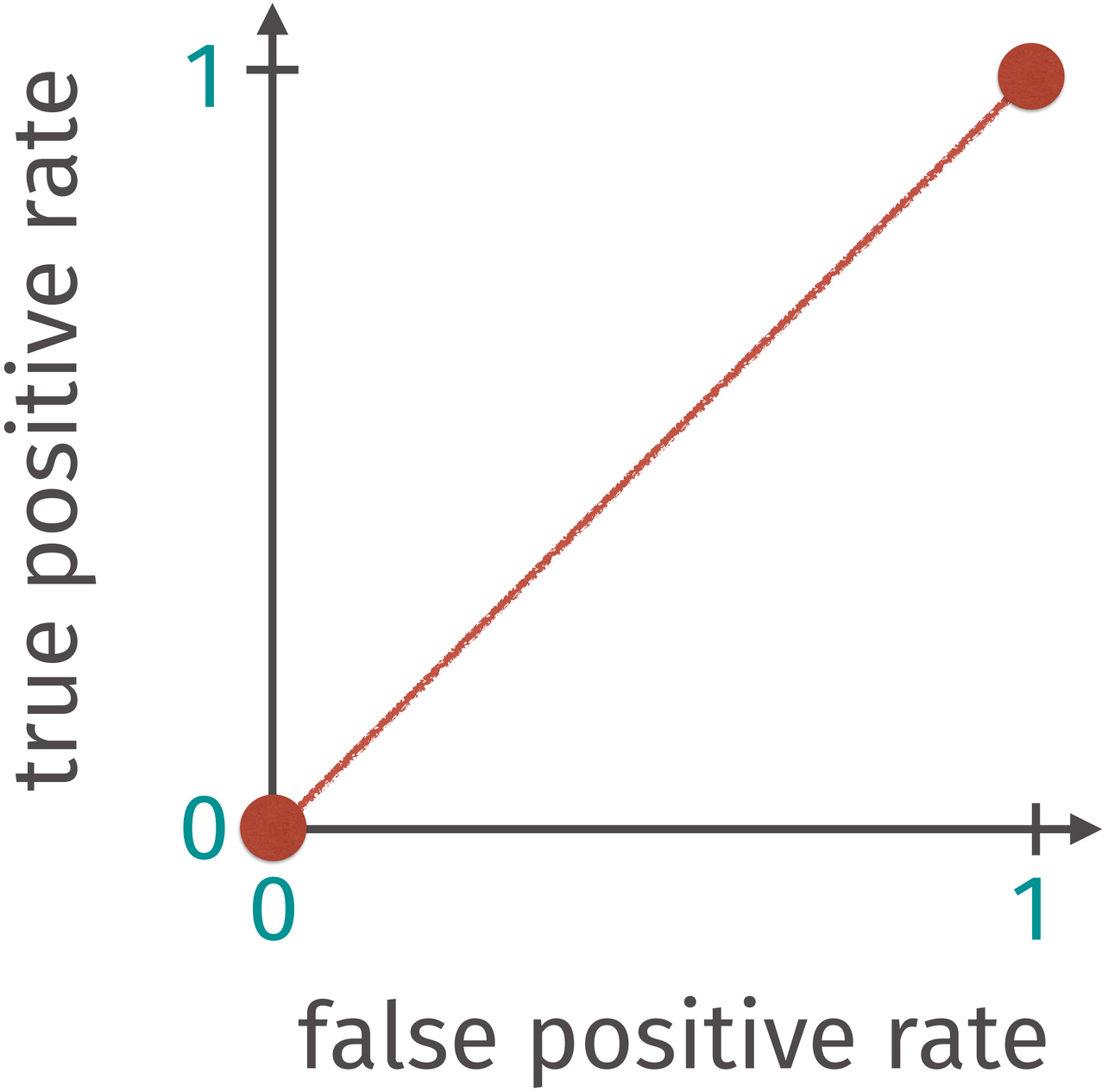
$t = 1.0$



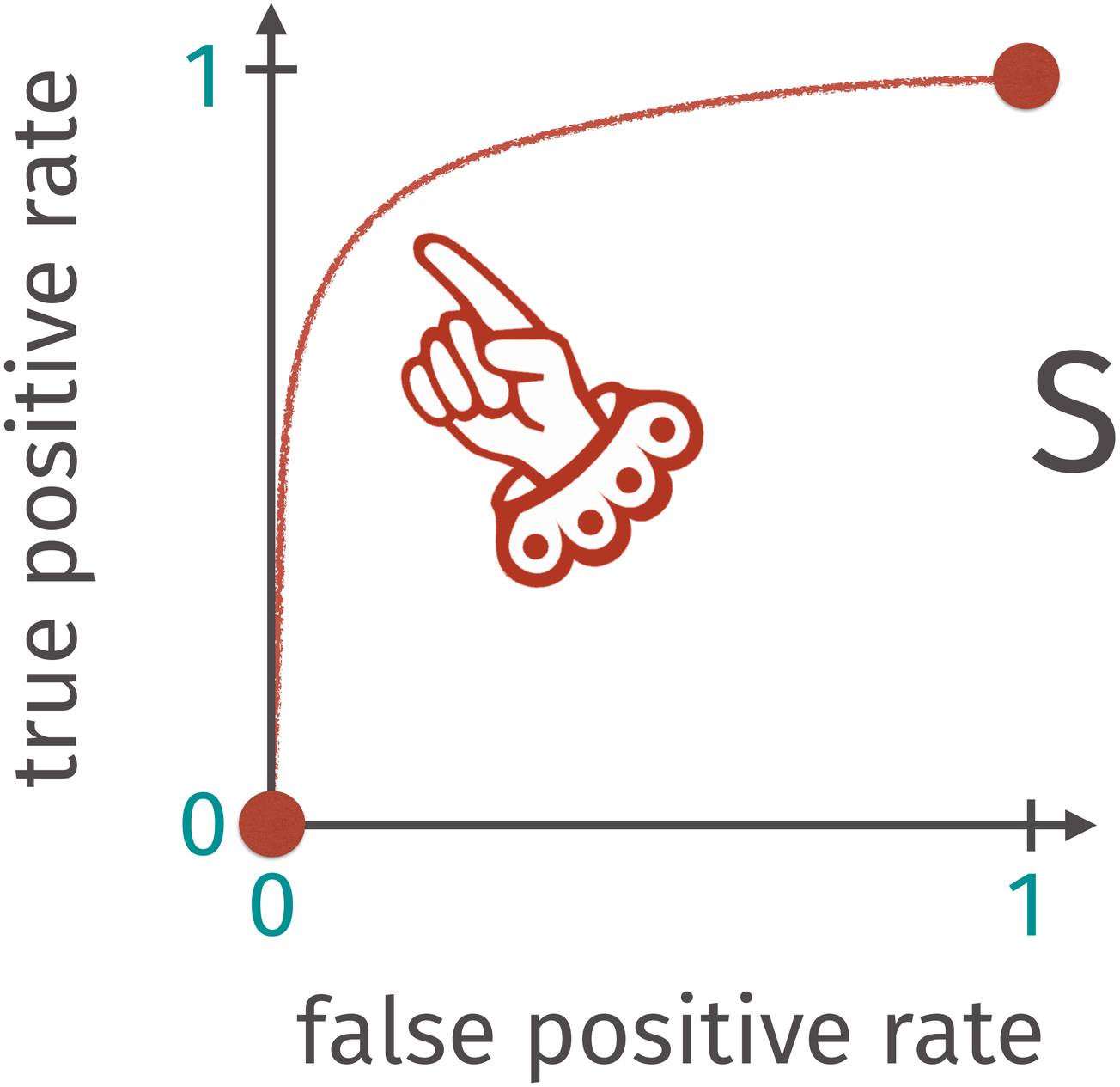
$t = 0.0$



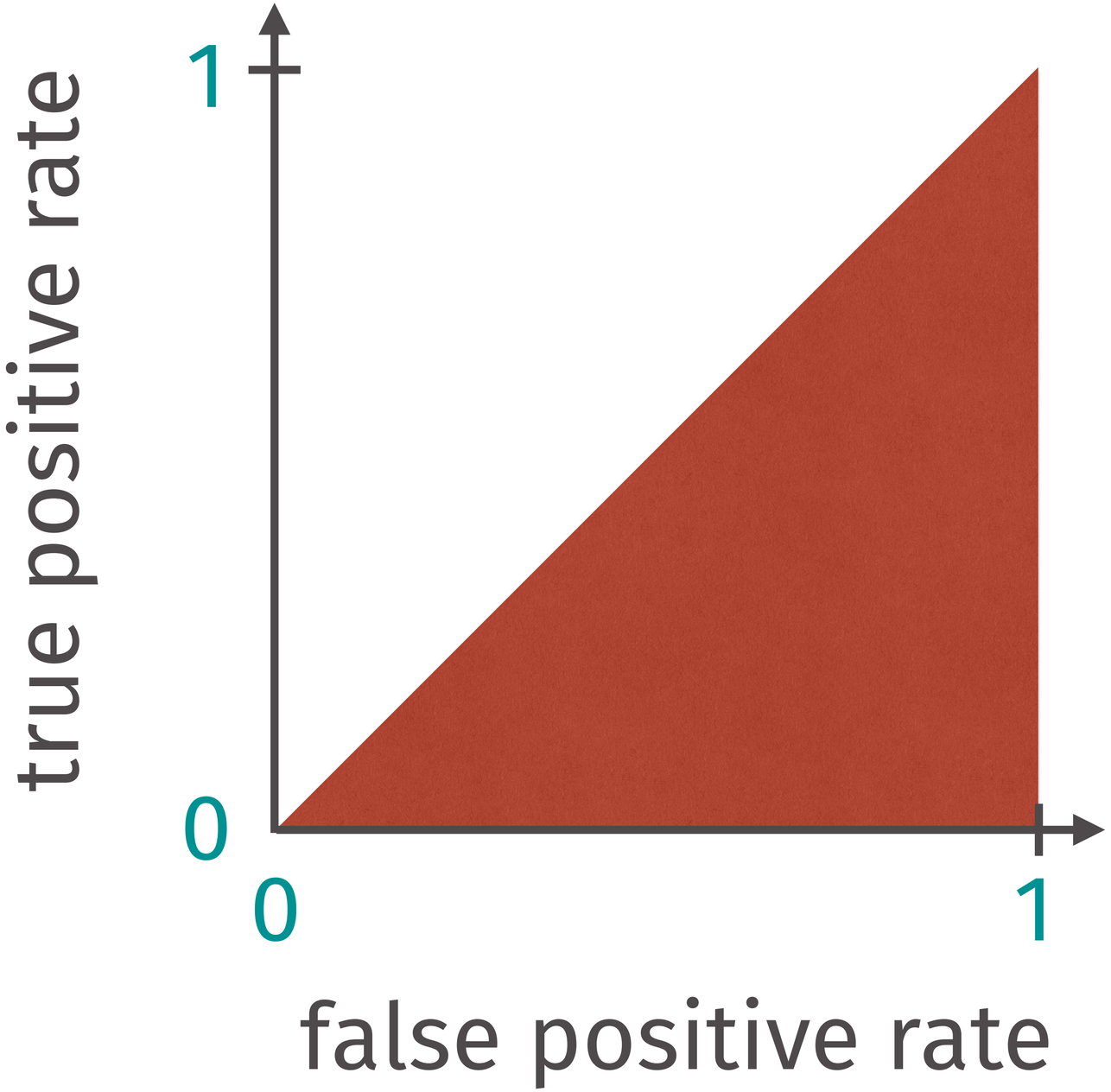
best *t*



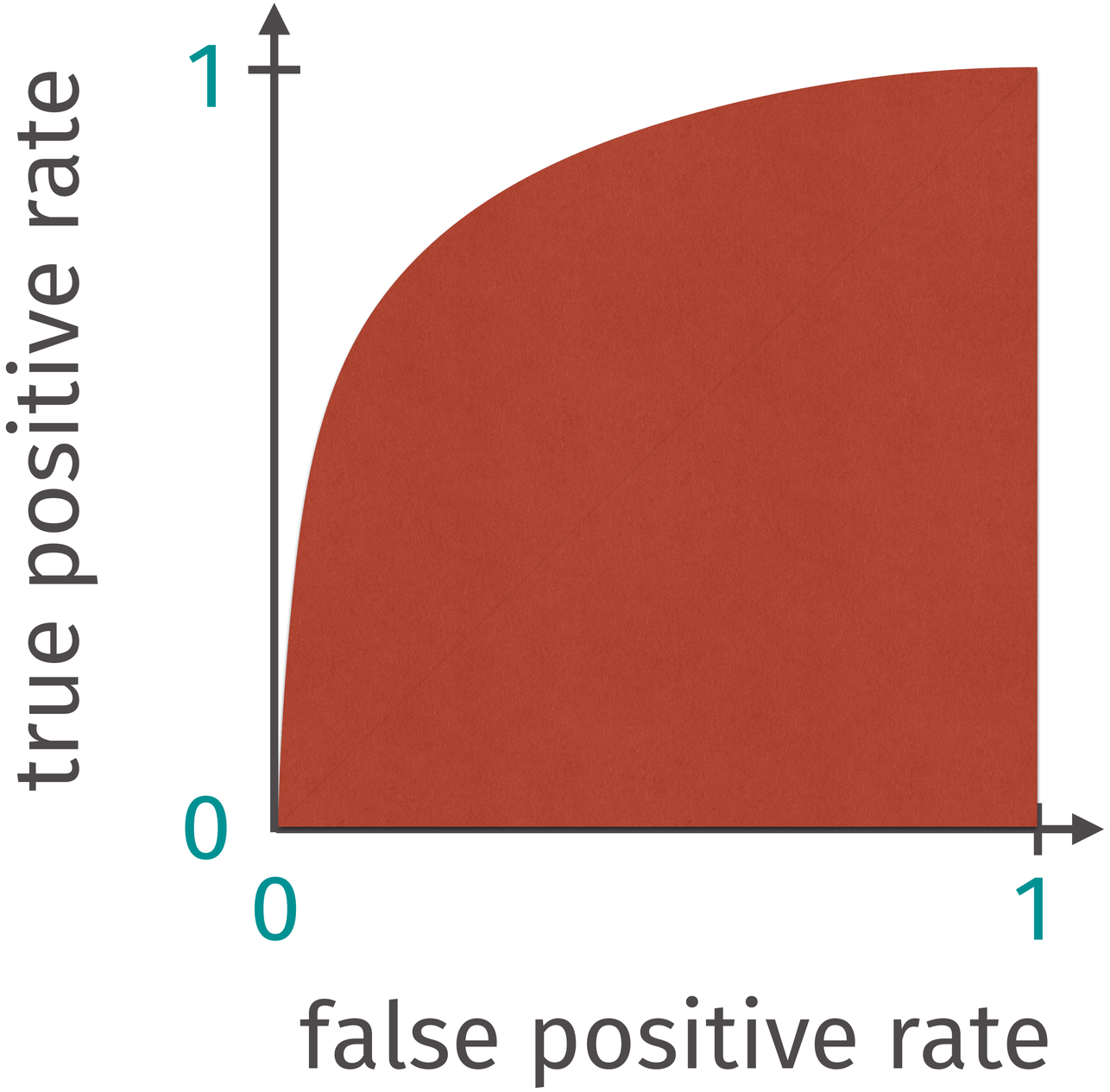
Random score



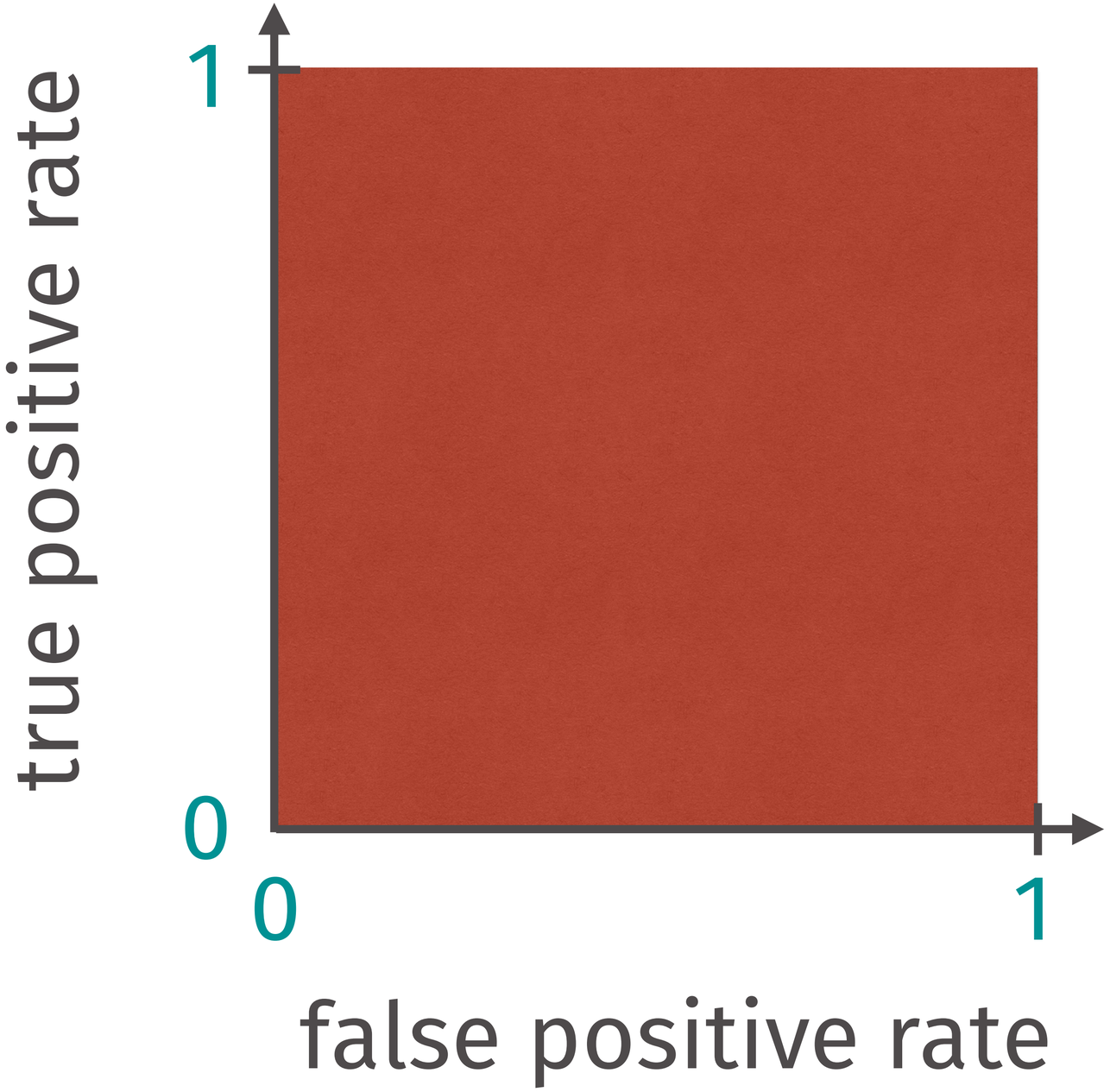
Something better



AUC = 0.5



AUC > 0.5



AUC = 1.0

```
Return-Path: <FlirtLife@happysillyfeetman.net>
X-Original-To: jan@stepien.cc
Delivered-To: jan@stepien.cc
Received: from ktxd8z.happysillyfeetman.net
    by r245-52.iq.pl (Postfix) with ESMTP id DCE364A61FC4
    for <jan@stepien.cc>; Wed, 14 Oct 2015 08:46:12 +0200 (CEST)
To: <jan@stepien.cc>
Date: Tue, 13 Oct 2015 23:46:07 -0700
Message-ID: <530723999953959530767719911262707@...>
From: "FlirtLife" <FlirtLife@happysillyfeetman.net>
Subject: ATTN: You have (1) New Message
Content-Language: en-us
MIME-Version: 1.0
Content-Transfer-Encoding: 8bit
Content-Type: multipart/alternative;
    boundary="-----=Part.592.3860.1444805167"

------=Part.592.3860.1444805167
Content-Transfer-Encoding: 8bit
Content-Type: text/plain; charset="UTF-8"

You received a new message on Flirtlocal.com

From: adorableAira

I need a spark in my sexual life...i want someone who can...
```

bogolexer

rtrn:FlirtLife
rtrn:happysilly.net
head:X-Original-To
head:jan
head:stepien.cc
head:Delivered-To
head:jan
head:stepien.cc
rcvd:from
rcvd:r245-52.iq.pl
rcvd:Postfix
rcvd:with
rcvd:ESMTP
rcvd:for
rcvd:jan
rcvd:stepien.cc
rcvd:Wed
rcvd:Oct
rcvd:CEST
to:jan
to:stepien.cc
head:Date
head:Message-ID
from:FlirtLife
from:FlirtLife
from:happysilly.net
subj:ATTN
subj:You
subj:have
subj:New
subj:Message
head:Content-Language
head:en-us
head:MIME-Version
head:Content-Transfer
head:bit
head:Content-Type
head:multipart
head:alternative
mime:Content-Transfer
mime:bit
mime:Content-Type
mime:text
mime:plain
mime:charset
mime:UTF-8
You
received
new
message
Flirtlocal.com
From
adorableAira
need
spark
sexual
life...i

Check for presence of tokens

to:stępień	0.52
…and to:jan	0.63
rcvd:tlsv1.2	0.66
…and to:stępień	0.84
…and 3 more tokens	0.89

Let’s take
old correspondence
into account

2499 innocent emails

602 spammy emails

Let’s talk about data compression

@janstepien

50 lines of each email
concatenate them together
and compress with lz4
...for both spam and ham

spam corpus ++ new email

ham corpus ++ new email

lz4

AUC 0.84

preprocess with bogolexer
50 tokens instead of lines

AUC 0.89

preprocess with bogolexer
50 tokens instead of lines

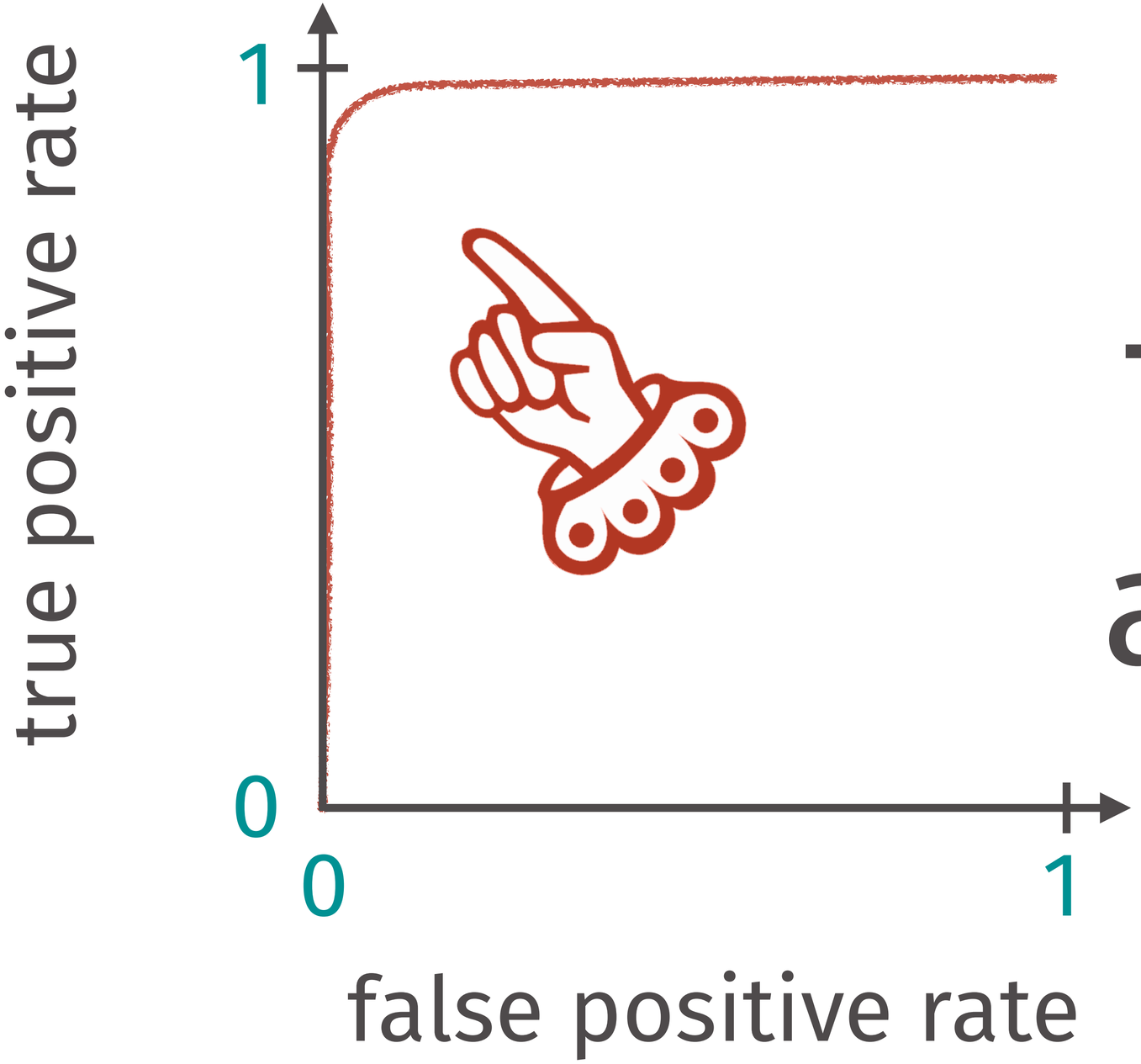
XZ

AUC 0.977

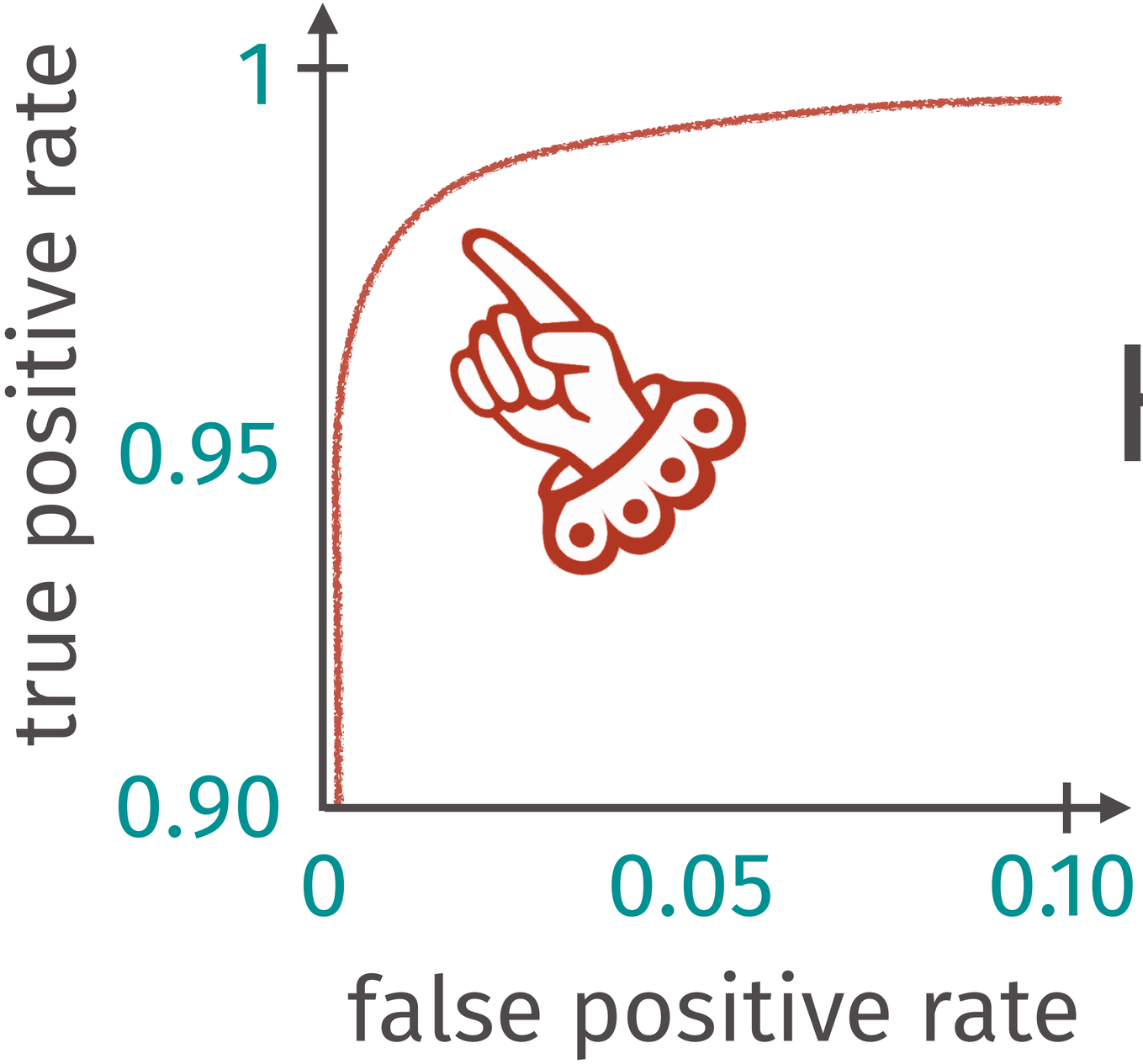
preprocess with bogolexer
150 tokens instead of lines

XZ

AUC 0.993



That’s not a problem



Here's the tricky bit

And now for something completely different

Apache SpamAssassin

Rule-based, statistical, and online modes of operation

Apache SpamAssassin

Rule-based, statistical, and online modes of operation

AUC 0.798

Apache SpamAssassin

Rule-based, statistical, and
online modes of operation

AUC 0.987

Bogofilter

Of bogolexer fame

AUC 0.990

vorpal

adj. Sharp or deadly.

adj. Having a special power making decapitation likely.

Home

Kyle Willett edited this page on Apr 27 · 22 revisions

Vowpal Wabbit

The Vowpal Wabbit (VW) project is a fast out-of-core learning system sponsored by Microsoft Research and (previously) Yahoo! Research. Support is available through the mailing list.

There are two ways to have a fast learning algorithm: (a) start with a slow algorithm and speed it up, or (b) build an intrinsically fast learning algorithm. This project is about approach (b), and it's reached a state where it may be useful to others as a platform for research and experimentation.

There are several optimization algorithms available with the baseline being sparse gradient descent (GD) on a loss function (several are available), The code should be easily usable. Its only external dependence is on the boost library, which is often installed by default.

To build vw from source, in various environments, please follow the instructions in the README.md file.

- Download
- Tutorial
- Command line arguments
- Algorithm details (e.g., input format, loss functions)

What examples of production deployments of Vowpal Wabbit are there?

This is a follow-up question to Which data stream mining tools can handle Big Data?.

John Langford, Machine Learning Researcher and Vowpal Wabbit author

Answered Dec 8, 2016

I'm aware of use at Amazon, Facebook, Yahoo, American Express, AOL, Baidu, Graphlab/Turi/Dato/Apple, FTI Consulting, IBM, Twitter, Yandex, and Microsoft.

I'm sure there are many others as well, as this is not closely tracked.

ABOUT THE AUTHOR

John Langford

Machine Learning Researcher and Vowpal Wabbit author

139.5k answer views 1.9k this month

Session Host Apr 2016

Feature Hashing for Large Scale Multitask Learning

Kilian Weinberger KILIAN@YAHOO-INC.COM
Anirban Dasgupta ANIRBAN@YAHOO-INC.COM
Josh Attenberg JOSH@CIS.POLY.EDU
John Langford JL@HUNCH.NET
Alex Smola ALEX@SMOLA.ORG
Yahoo! Research, 2821 Mission College Blvd., Santa Clara, CA 95051 USA

Keywords: kernels, concentration inequalities, document classification, classifier personalization, multitask learning

Abstract

Empirical evidence suggests that hashing is an effective strategy for dimensionality reduction and practical nonparametric estimation. In this paper we provide exponential tail bounds for feature hashing and show that the interaction between random subspaces is negligible with high probability. We demonstrate the feasibility of this approach with experimental results for a new use case — multitask learning with hundreds of thousands of tasks.

1. Introduction

Kernel methods use inner products as the basic tool for comparisons between objects. That is, given objects

frequently encounter the opposite problem: the original input space is almost linearly separable (often because of the existence of handcrafted non-linear features), yet, the training set may be prohibitively large in size and very high dimensional. In such a case, there is no need to map the input vectors into a higher dimensional feature space. Instead, limited memory makes storing a kernel matrix infeasible.

For this common scenario several authors have recently proposed an alternative, but highly complimentary variation of the kernel-trick, which we refer to as the *hashing-trick*: one *hashes* the high dimensional input vectors x into a *lower* dimensional feature space $\mathbb{R}^m$ with $\phi : \mathcal{X} \to \mathbb{R}^m$ (Langford et al., 2007; Shi et al., 2009). The parameter vector of a classifier can therefore live in $\mathbb{R}^m$ instead of in $\mathbb{R}^n$ with kernel matrices or $\mathbb{R}^d$ in the original input space, where $m \ll n$ and $m \ll d$. Different

tokenise with bogolexer
sort and deduplicate
and build a VW model

VW

AUC 0.998

Technical details

Postfix as my MTA
procmail to call the filter
bogolexer and VW to score
and procmail to deliver

```
for message in $(find archive/ -type f); do
  bogolexer -p < $message \
    | tr '[:upper:]' '[:lower:]' \
    | egrep -vi '(bogosity|bogofilter|vowpal-wabbit)' \
    | sort | uniq | tr "\n" " "
done
```

```
vw bag.gz -b 15 --passes 8 --cache_file m-cache -f model
```

```
bogolexer -p \
  | tr '[:upper:]' '[:lower:]' \
  | sort | uniq | tr "\n" " " \
  | vw -t -i model -p /dev/stdout
```

Vowpal-Wabbit-Says: 0.964

```
cmd="< I filter.sh | grep ^Vowpal-Wabbit-Says \
     | sed 's|^Vowpal-Wabbit-Says: ||; s|$| I|;'"

find archive/ -type f -print0 \
     | xargs -0 -P 4 -I I sh -c "$cmd" | tee results

grep -v archive/Junk/ results | sort -r | head

grep archive/Junk/ results | sort | head
```

hunch.net/~vw

Spam filtering
is an arms race

jan@stepien.cc

Combating spam

 or 

how I befriended the Killer Rabbit of Caerbannog

with **Jan Stępień**

@janstepien