The Wayback Machine - https://web.archive.org/web/20250616092056/https://thepalindrome.org/p/t...

The Anatomy of Logistic Regression

Basic model, invaluable insights

TIVADAR DANKA
JUN 08, 2025

♡ 29   💬 4   ⟳ 2   Share

In today's post, we'll talk about logistic regression.

Yes, I know. You are probably tired of hearing about it: it's basic, boring, and not that effective in practice.

Don't click away just yet! By dissecting how the logistic regression is constructed, we gain invaluable insight into how machine learning models are designed. Even the state-of-the-art ones.

So, why do I like the logistic regression so much?

Because it explains how we can turn *geometry* into *probability*. Let's see!

The Palindrome breaks down advanced math and machine learning concepts with visuals that make everything click.

Join the premium tier to get instant access to guided tracks on graph theory, foundations of mathematics, and neural networks from scratch.

You know that I always prefer to keep things simple. This way, we can understand any complex phenomenon by omitting irrelevant details and focusing only on the core components.

Because of this, let's focus on the most basic classification problem possible: one feature, two classes. Here's an example problem: given a student's number of hours spent studying ($x \in \mathbb{R}$), predict whether or not the exam is successful ($y \in \{0, 1\}$).

By plotting x against y, we can easily visualize the dataset.
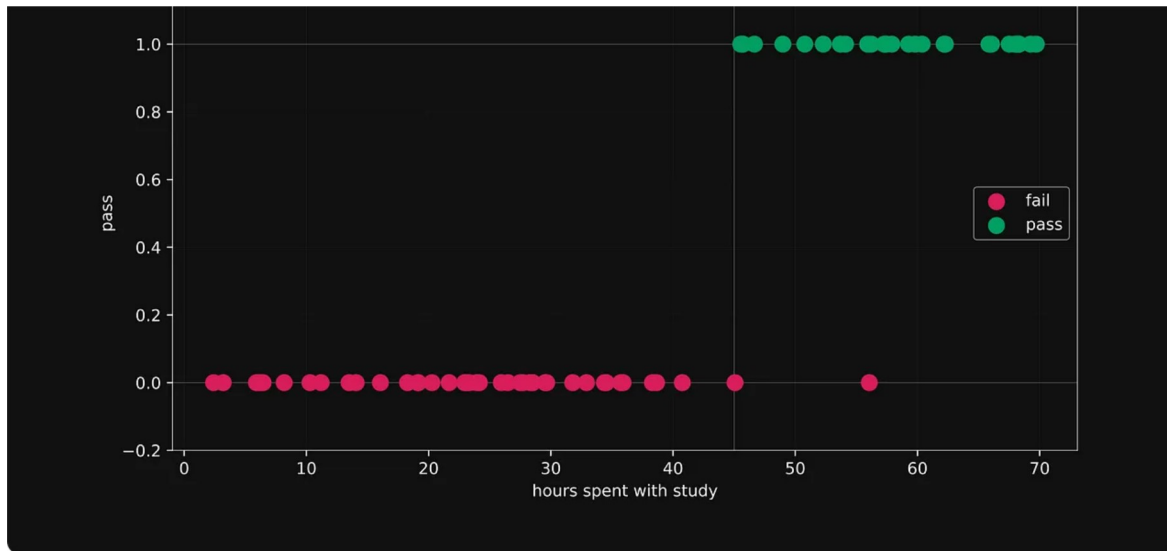
Let's build a predictive model!

Mathematically speaking, we are looking for a function that assumes values between zero and one:

- $\sigma(ax + b) > 1/2$ suggests *pass* (represented by the scalar value 1),
- while $\sigma(ax + b) < 1/2$ suggests *fail* (represented by the scalar value 0).

Following our principle of simplicity, (one of) the most basic classification models is logistic regression, or in other words, a linear regression plus a Sigmoid:
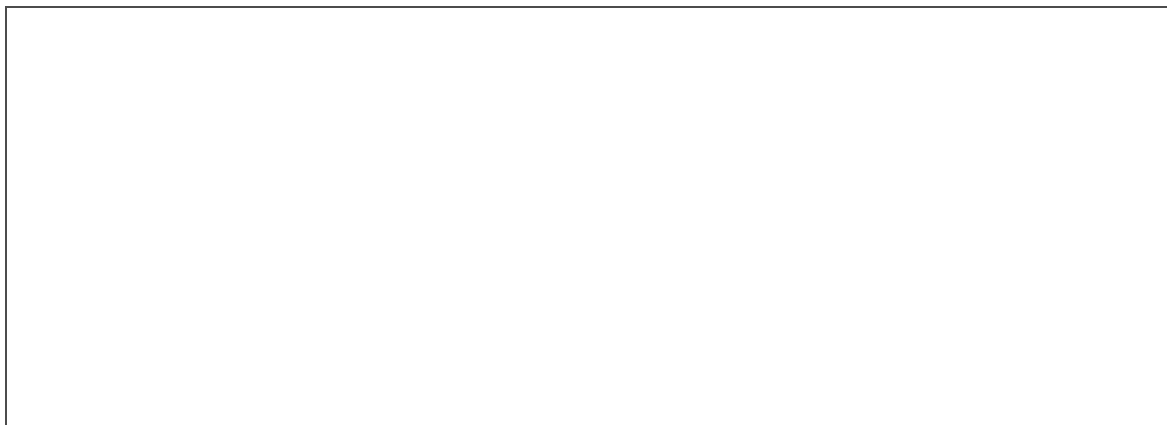
Upon training — that is, finding a and b that minimizes the error on the training dataset — we obtain a function that takes a scalar input x (the number of hours spent with study) and returns the *estimated probability* of the student with x hours of study to pass the exam.

Here's how such a model looks.
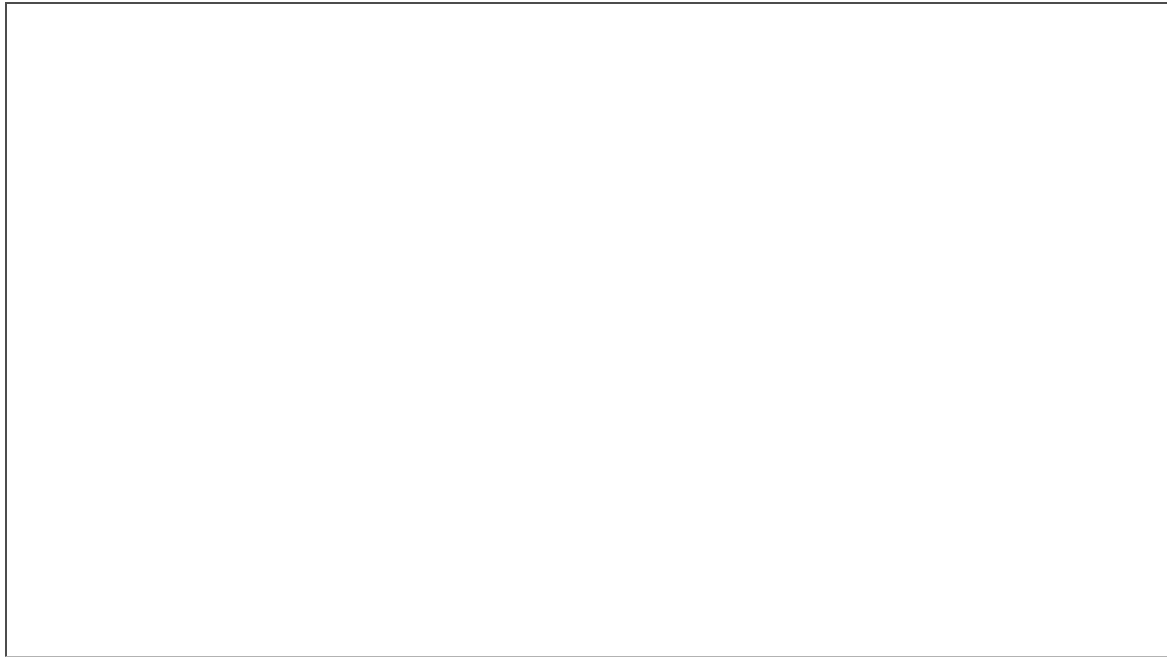
But what does the expression *σ(ax + b)*, involving fractions and exponentials, have to do with probability?

Let's deconstruct the model step by step.

*Struggling to learn math and machine learning on your own?*

The Palindrome makes complex ideas click — with visual explainers, step-by-step guides, and clear learning paths.

Join the premium tier to finally understand graph theory, neural networks, and the foundations of mathematics in a structured way.

$$\sigma(ax + b) = \frac{1}{1 + e^{-(ax+b)}}$$

First, logistic regression transforms the score via *y = ax + b*, which describes a *linear*

function, visualized by a *line* on the plane. Take a look.
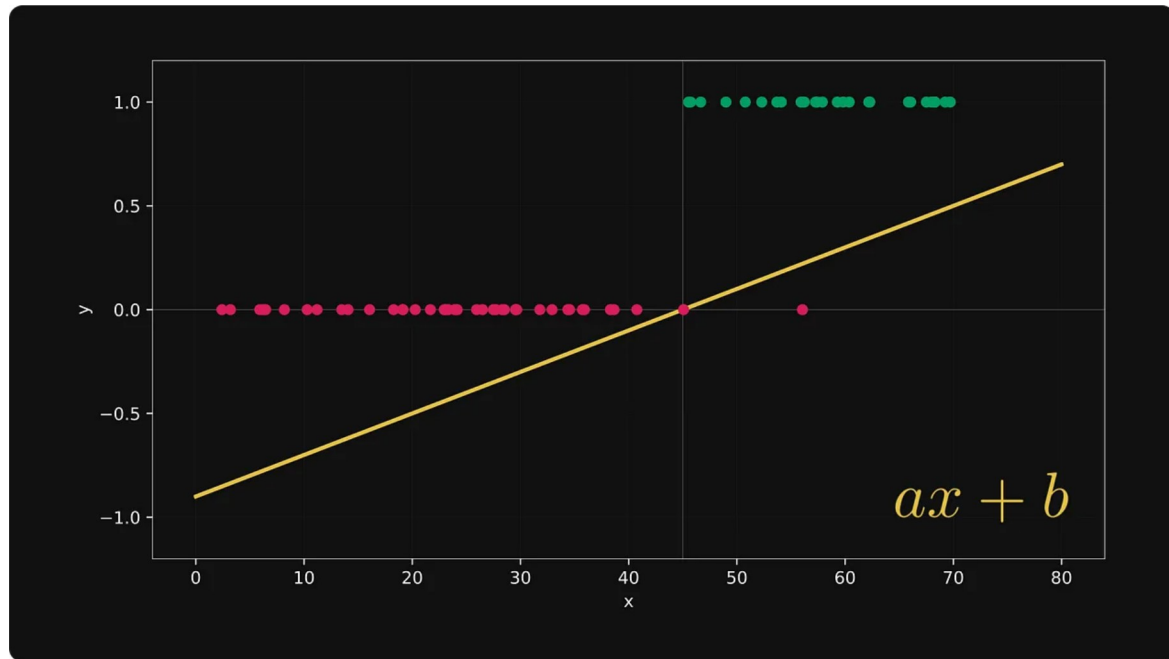
The function $y = ax + b$ maps the unit of measurement (hours in our case) into *logits*, short for *logistic unit.*

If you have a sharp eye, you probably noticed that a positive logit score indicates a pass, while a negative one indicates a fail.

Next, the logit is exponentiated, transforming the logits into positive real numbers by e^{ax+b}.
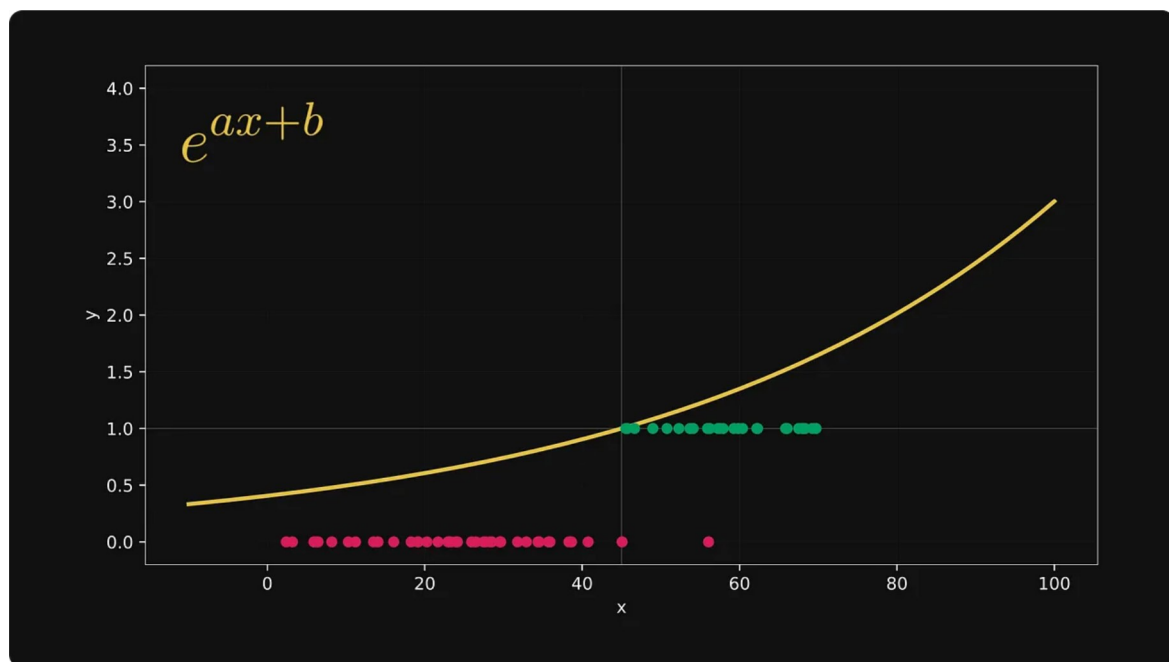
This time, the "cutoff value" moves from 0 to 1.

As 1) more hours of study make a successful exam more likely, and 2) the expression $\sigma(ax + b)$ involves a reciprocal, we flip the values along the x-axis, yielding $e^{-(ax+b)}$.
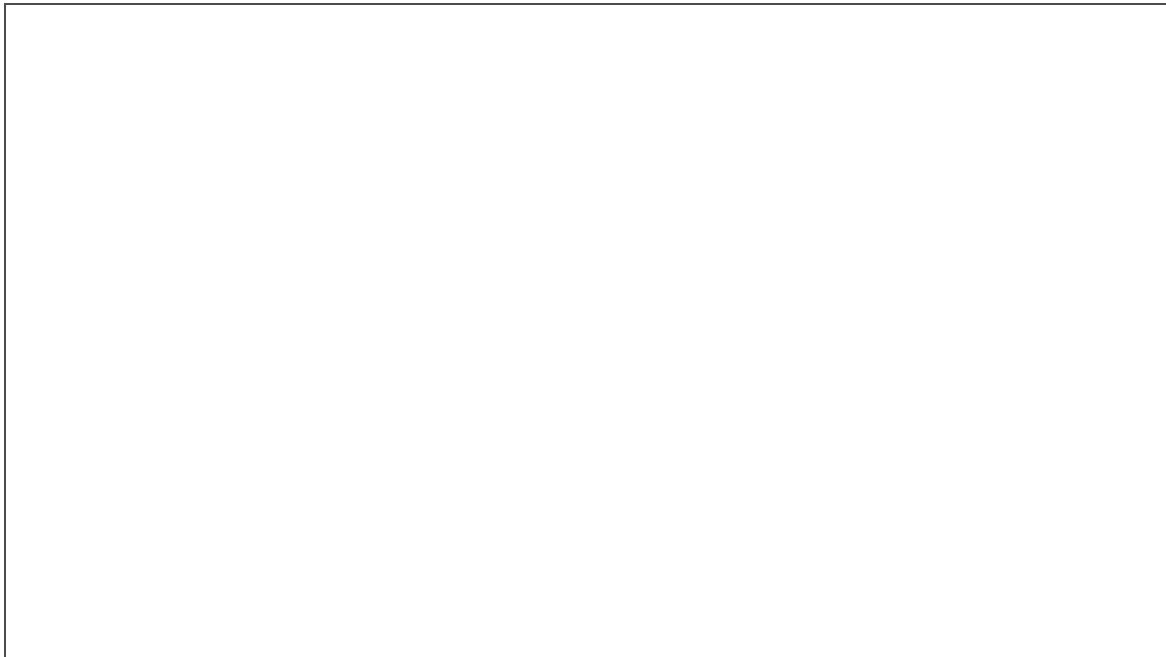
We are almost there. Before taking the reciprocal of the *exponentiated negative logit score* (rolls right off your tongue, doesn't it?), think about the properties of probability.

Probabilities are always between 0 and 1. To achieve this, we have to increase the values, ensuring they are always larger or equal to 1. This gives *$1 + e^{-(ax+b)}$* below.
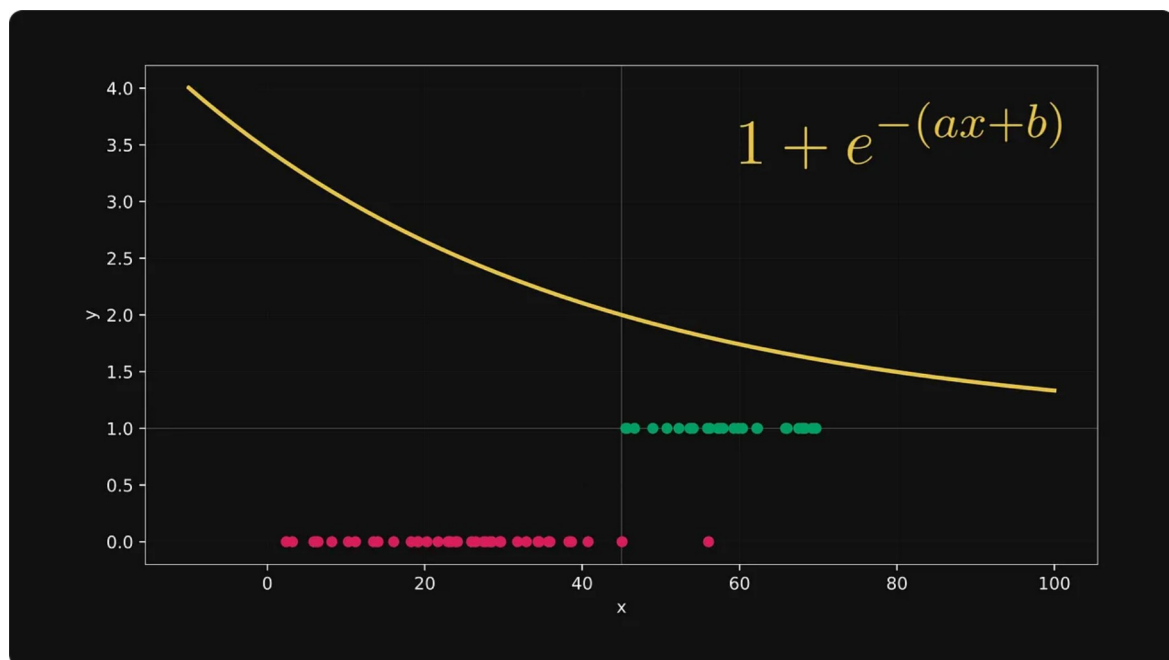
After all this setup, we are ready to take the last step, turning the scores into probabilities, done by a simple reciprocation.
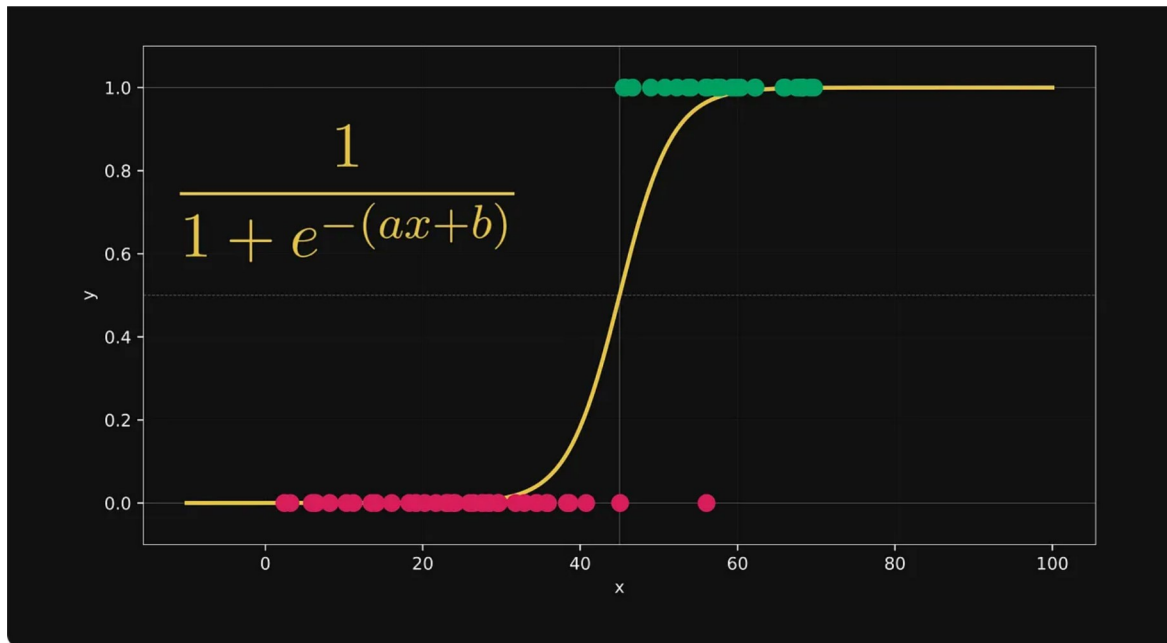
And we are done!

One last thing. In the introduction, I mentioned "turning *geometry* into *probability*". We have seen the probability part, but where's the geometry?

This becomes apparent when we move to higher dimensions. Check out this two-dimensional classification problem.

There, the logits given by $y = a_1x_1 + a_2x_2 + b$ forms a plane, and the decision boundary $a_1x_1 + a_2x_2 + b = 0$ forms a line.

Check out the illustration below, with 1) the white line indicating the decision boundary, and 2) the colormap indicating the predicted probabilities.
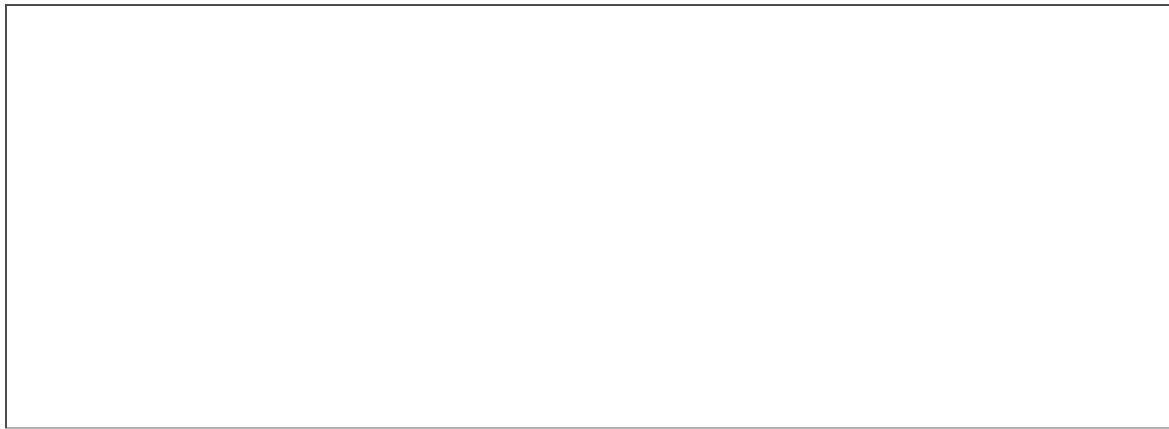
In essence, the logit $a_1x_1 + a_2x_2 + b$ is an indicator of the signed distance from the decision boundary, bringing the aforementioned geometry to the table.

In fact, $a_1x_1 + a_2x_2$ represents some kind of distance. But that's a topic for another day!

Subscribe to The Palindrome

By Tivadar Danka · Launched 3 years ago

mathematics ∪ machine learning

Type your email...	Subscribe

← Previous   Next →

Discussion about this post

Comments   Restacks

Write a comment...

David Sutherland 5d   ...

💜 Liked by Tivadar Danka

Thanks for writing posts set to this size. I like having bite sized posts for me to digest.

♡ LIKE (1)   💬 REPLY   ↑ SHARE

1 reply by Tivadar Danka

San Sani Nassif   8d   ...

Something is odd about the plots pre/post the application of the reciprocal. The yellow lines are not such that 1 / first = second.

♡ LIKE   ◯ REPLY   ⬆ SHARE

1 reply by Tivadar Danka

2 more comments...

San Sani Nassif   8d

Something is odd about the plots pre/post the application of the reciprocal. The yellow lines are not