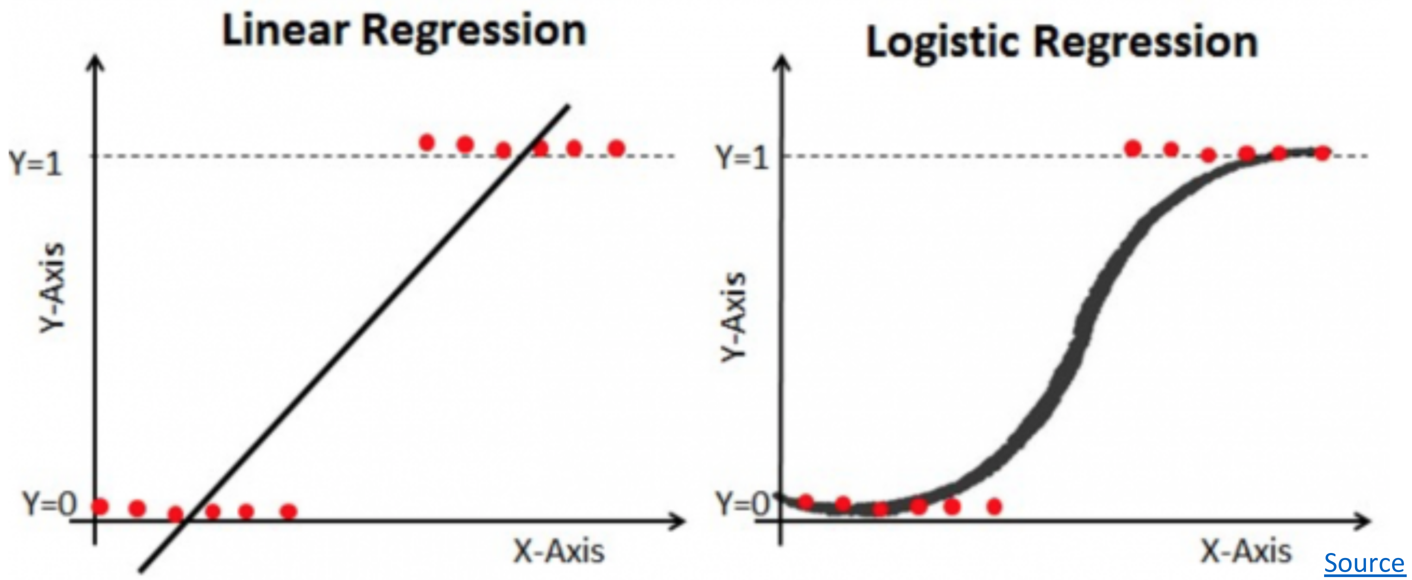
CSCI 3022: Intro to Data Science

Lecture 25: Logistic Regression

Rachel Cox

Department of Computer Science



Source

Announcements & Reminders

- ☐ Practicum 2 - Due Friday May 1 at 11:59pm
  - No LATE DAYS!
- ☐ Final Exam - May 5th
  - 24 hour window
  - info on Piazza

What will we learn today?

- ❑ Logistic Regression

- ❑ *Introduction to Statistical Learning, Chapter 4, Think Stats 11.6*

Regression as prediction

So far, we've learned about various forms of regression.

We've viewed regression in terms of learning a relationship between one or more features and a response:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ \dots + \beta_p x_{ip} + \epsilon_i$$

We've talked about using regression as a way to make predictions.

What about using regression as a classifier?

Regression as prediction

Example: Back to the Titanic data!

	age	outcome
0	25	survived
1	30	survived
2	35	survived
3	40	survived
4	45	died
5	50	died
6	55	died
7	60	died

Recode outcomes as y = {0, 1}

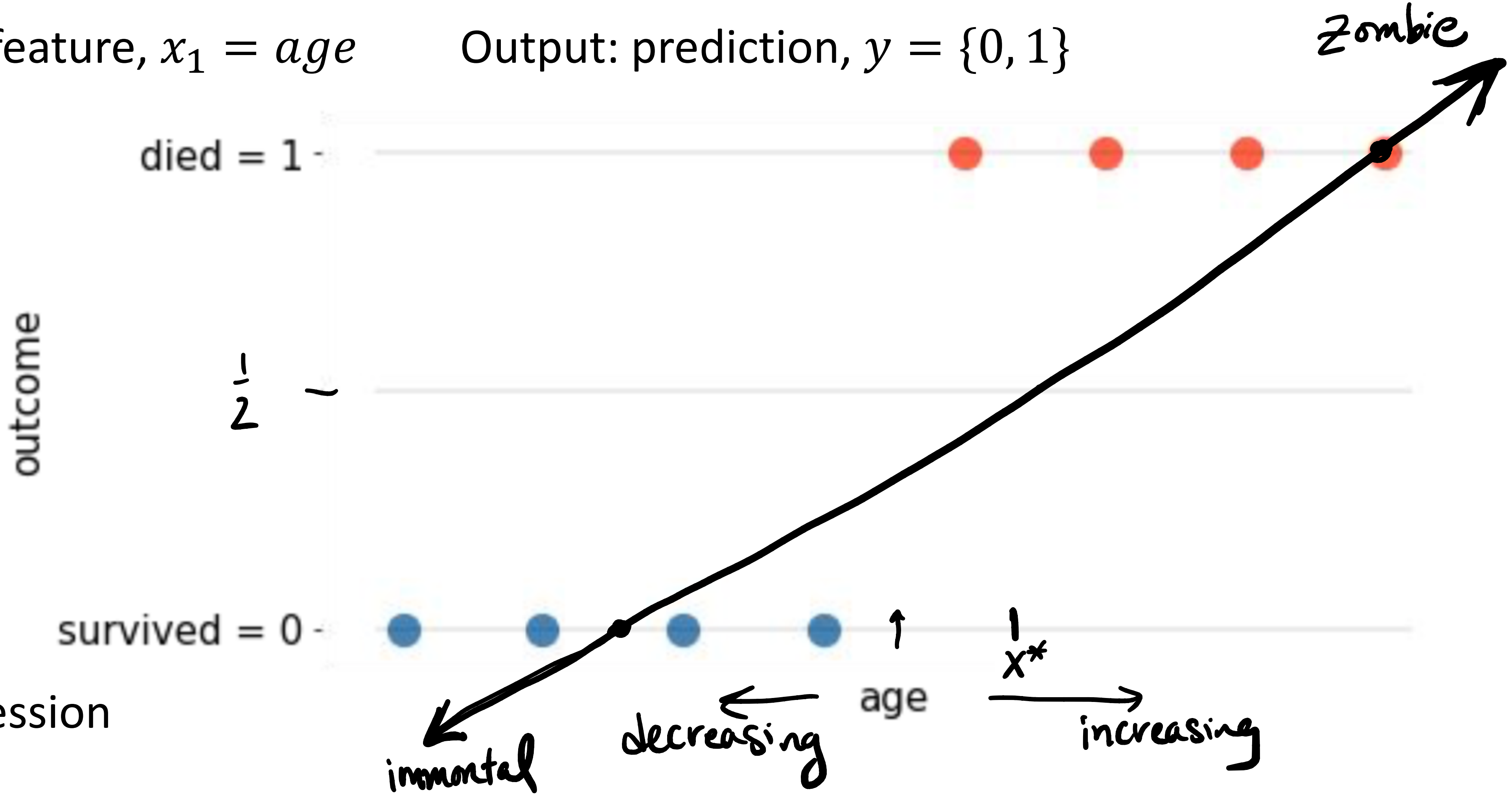
	age	outcome
0	25	0
1	30	0
2	35	0
3	40	0
4	45	1
5	50	1
6	55	1
7	60	1

- Let's try using linear regression to take the feature x = Age and predict the response y = Outcome

Regression as prediction

Example: Suppose you want to predict whether a passenger on the Titanic survived or not, based on passenger Age as the sole feature.

Model input: single feature, $x_1 = age$ Output: prediction, $y = \{0, 1\}$

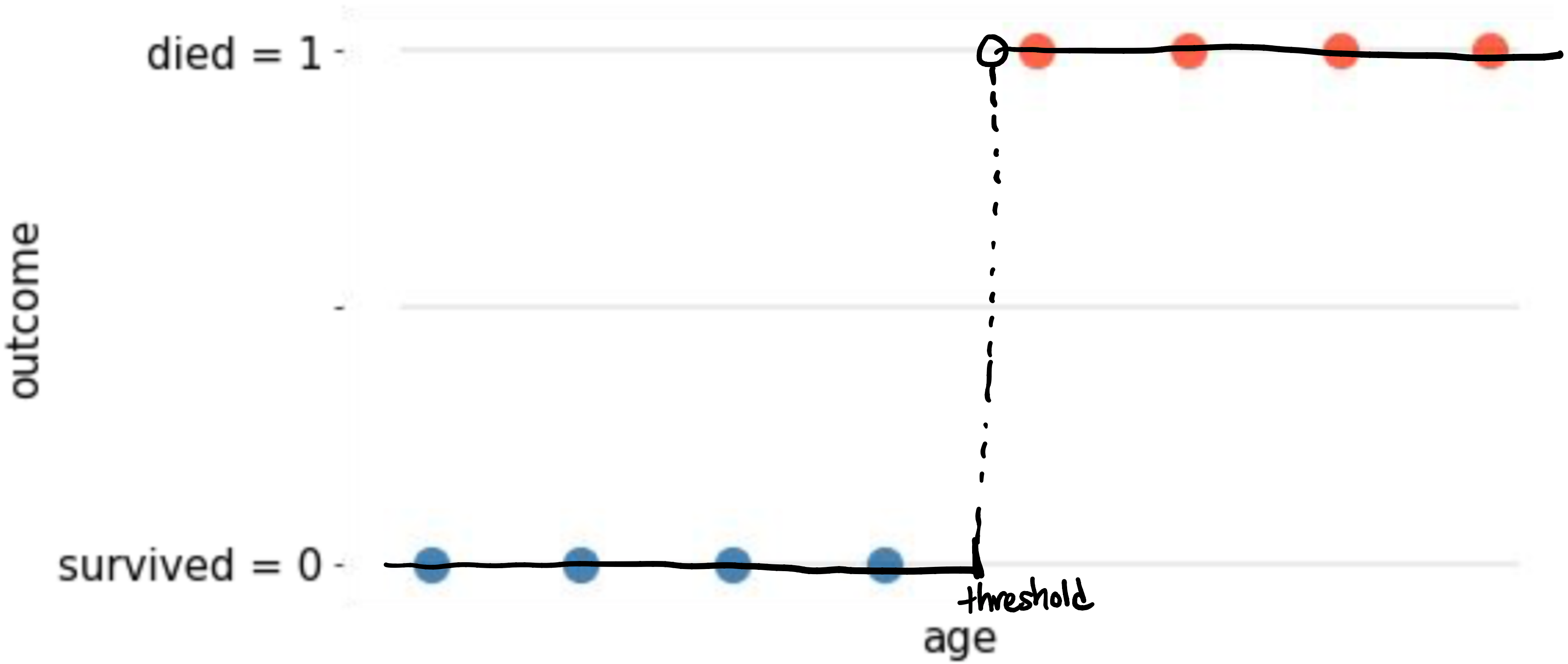


First Idea: Linear regression

$$y = \beta_0 + \beta_1 x_1$$

Regression as prediction

Example: Suppose you want to predict whether a passenger on the Titanic survived or not, based on passenger Age as the sole feature.



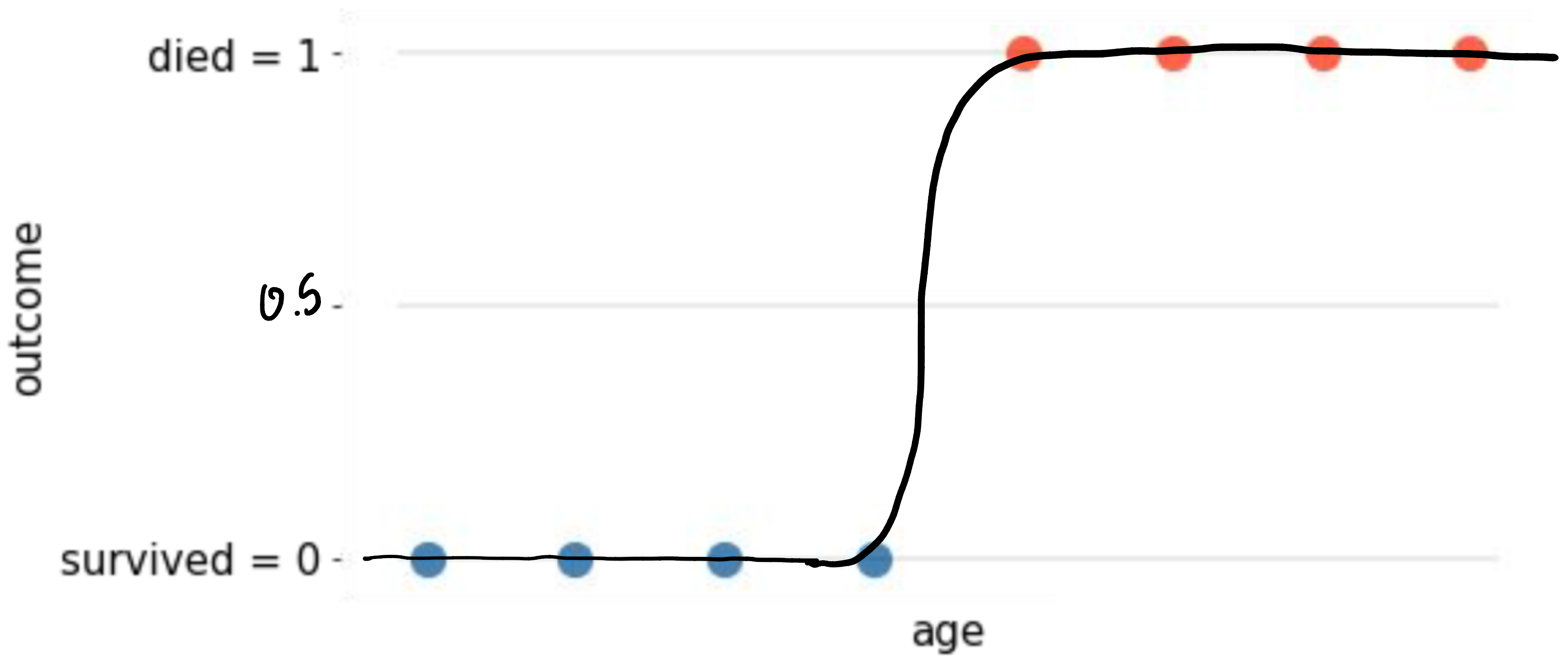
Second Idea:
Piecewise function

$$y = \begin{cases} 1 & \text{if } x_1 > \text{some threshold} \\ 0 & \text{otherwise} \end{cases}$$

- ok fit of data
- not differentiable

Regression as prediction

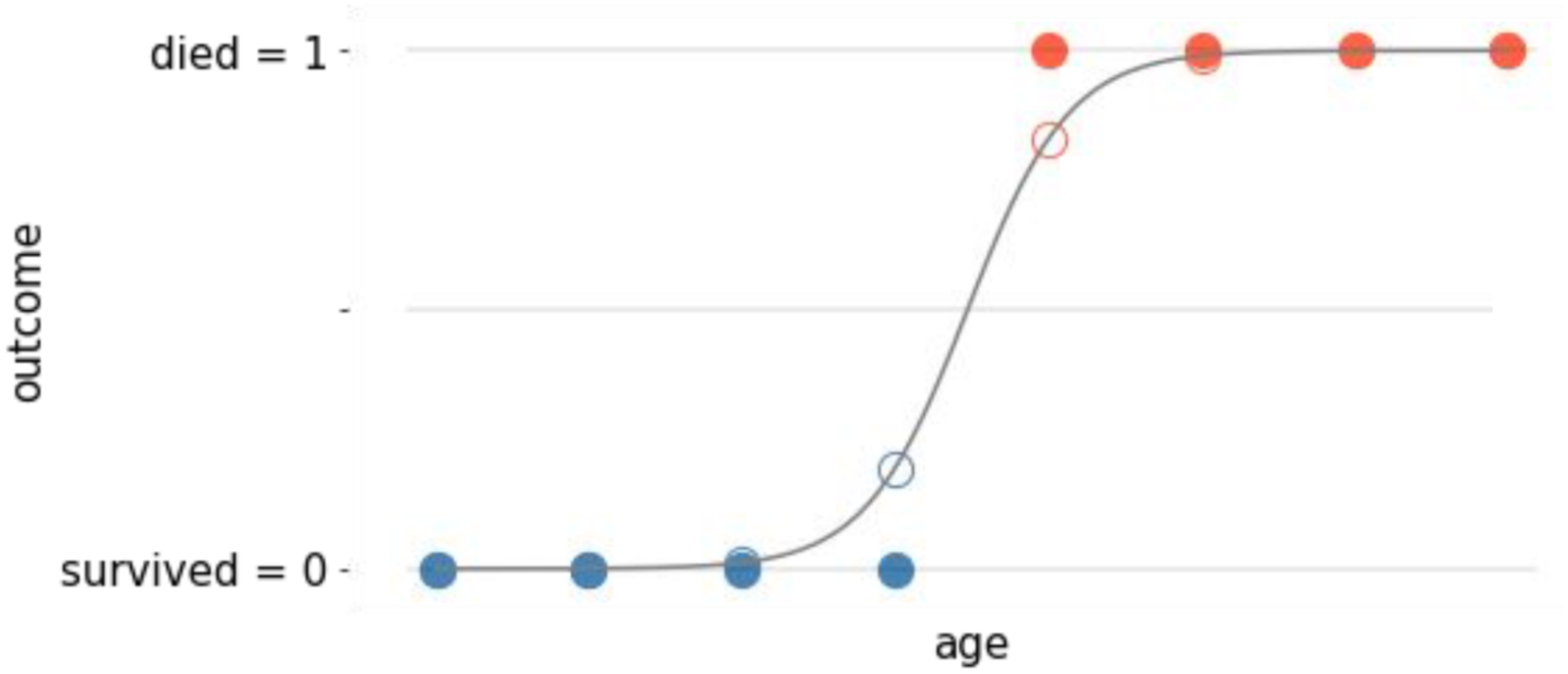
Example: Suppose you want to predict whether a passenger on the Titanic survived or not, based on passenger Age as the sole feature.



Idea:
Need something that behaves more like a probability $p(y=1 \mid age)$

Regression as prediction

Example: Suppose you want to predict whether a passenger on the Titanic survived or not, based on passenger Age as the sole feature.



This curve looks nice. What is it?

The sigmoid function

$$\mathrm{sigm}(z) = \frac{1}{1+e^{-z}}$$

If $z=0$

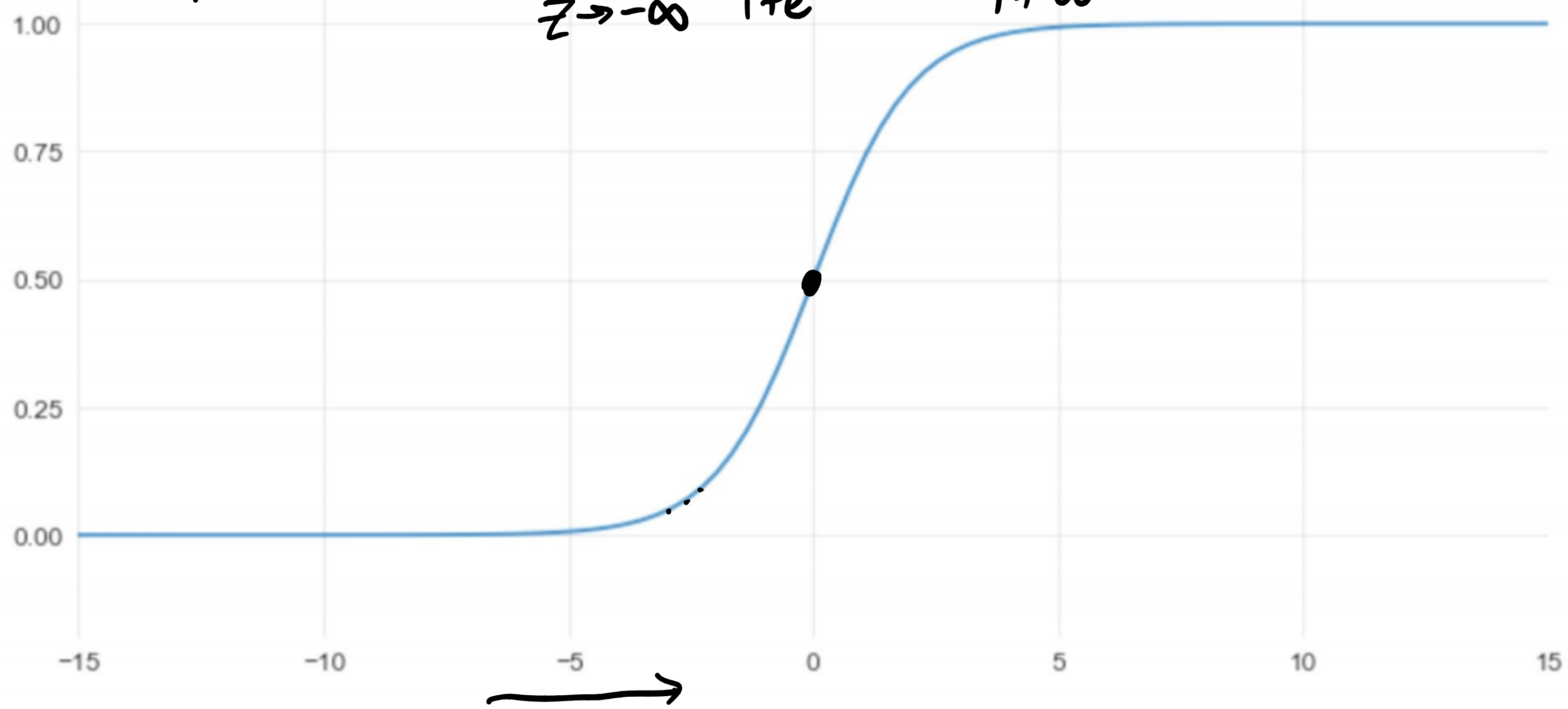
$$\mathrm{sigm}(0) = \frac{1}{1+e^{-0}} = \frac{1}{1+1} = \frac{1}{2}$$

As $z \to \infty$ $\lim_{z\to\infty} \frac{1}{1+e^{-z}} = \frac{1}{1+0} = 1$

As $z \to -\infty$ $\lim_{z\to-\infty} \frac{1}{1+e^{-z}} = \frac{1}{1+\text{"}\infty\text{"}} = 0$

Has nice properties:

- Behaves like a probability
  0, 1
- Distinguishes between points
- Really smooth
  ⇒ differentiable



Logistic regression

linear regression: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

The model: $p(y = 1 \mid x) = \text{sigm}(\beta_0 + \beta_1 x)$

$z = \beta_0 + \beta_1 x$

Learn weights β_0 and β_1 from the data

Classify data point x according to: $\hat{y} = \begin{cases} 1 & \text{if sigm}\left(\hat{\beta}_0 + \hat{\beta}_1 x\right) \geq 0.5 \\ 0 & \text{if sigm}\left(\hat{\beta}_0 + \hat{\beta}_1 x\right) < 0.5 \end{cases}$ ✓

How do we find the decision boundary?

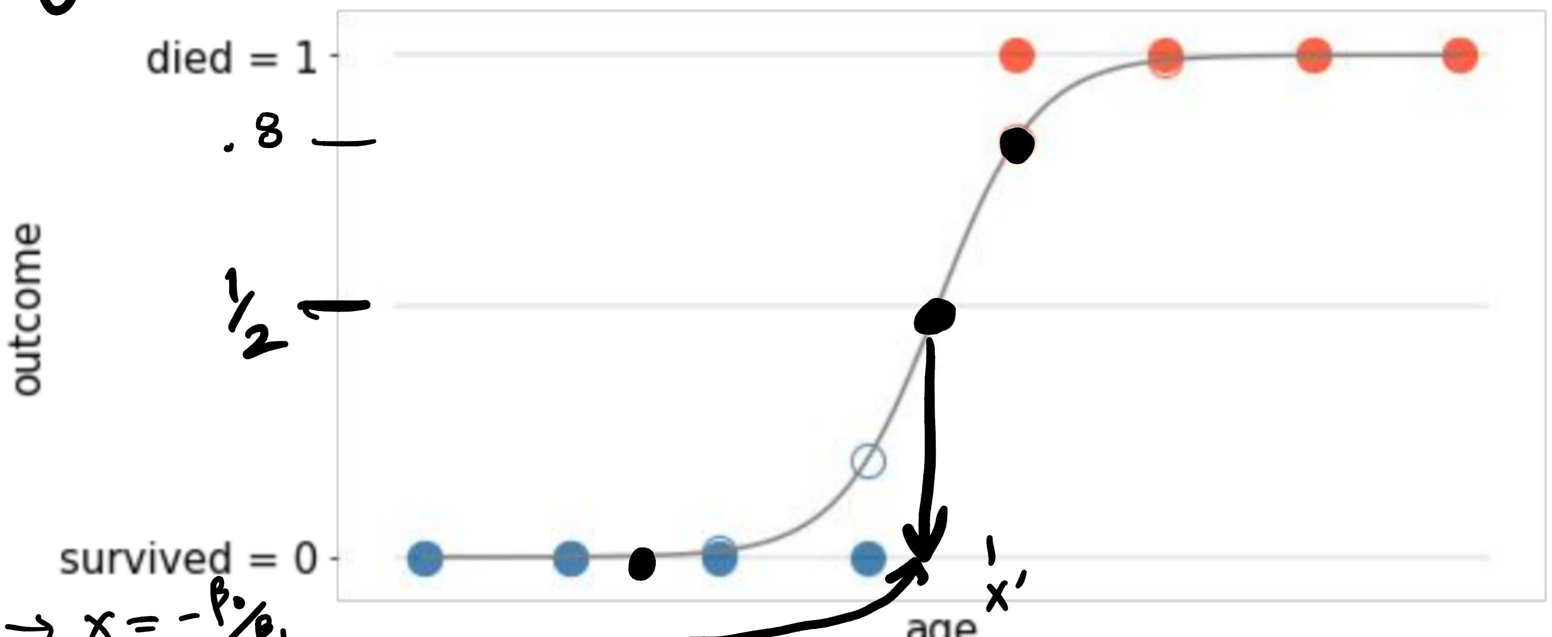
$$\frac{1}{1+e^{-(\beta_0+\beta_1 x)}} = \frac{1}{2}$$

$$1+e^{-(\beta_0+\beta_1 x)} = 2$$

$$e^{-(\beta_0+\beta_1 x)} = 1$$

$$-(\beta_0+\beta_1 x) = 0$$

$$\beta_1 x = -\beta_0 \rightarrow x = -\beta_0/\beta_1$$



Logistic regression

Our inevitable path to logistic regression and the sigmoid function began with our insistence on modeling the relationship between features and the response as a legitimate probability.

With some basic algebra, we can arrive at an interpretation of logistic regression that is very regression-like.

First we need to talk about odds.

Logistic regression

In statistics, the **odds** of an event is the ratio of the probability that the event occurs, divided by the probability that the event does not occur, and then generally flipped to get a value bigger than 1

$$\text{odds} = \frac{p}{1-p}$$

Example: If p = 0.75, then odds = $\frac{.75}{1-.75} = \frac{.75}{.25} = 3$

We would say that the odds are 3 to 1 in favor

Example: If p=0.1, then odds = $\frac{0.1}{1-0.1} = \frac{0.1}{0.9} = 1/9$

We would say that the odds are 9 to 1 against

Logistic regression

In logistic regression, we model $p = \underbrace{p(y = 1 \mid x)}_{} = \underbrace{\mathrm{sigm}(\beta_0 + \beta_1 x)}_{}$

What if we calculate the odds that y = 1, given the data x?

$$odds = \frac{p}{1-p} = \frac{\frac{1}{1+e^{-(\beta_0+\beta_1 x)}}}{1 - \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}}$$

$$= \frac{\frac{1}{1+e^{-(\beta_0+\beta_1 x)}}}{\frac{1+e^{-(\beta_0+\beta_1 x)}-1}{1+e^{-(\beta_0+\beta_1 x)}}}$$

$$= \frac{\frac{1}{1+e^{-(\beta_0+\beta_1 x)}}}{\frac{e^{-(\beta_0+\beta_1 x)}}{1+e^{-(\beta_0+\beta_1 x)}}}$$

$$= \frac{1}{\cancel{(1+e^{-(\beta_0+\beta_1 x)})}} \cdot \frac{\cancel{(1+e^{-(\beta_0+\beta_1 x)})}}{e^{-(\beta_0+\beta_1 x)}}$$

$$= \frac{1}{e^{-(\beta_0+\beta_1 x)}}$$

$$= e^{\beta_0+\beta_1 x}$$

$$odds = e^{\beta_0+\beta_1 x}$$

$$\ln(odds) = \beta_0+\beta_1 x$$

Logistic regression

Taking the natural log of both sides, we get:

$$\log(\text{odds}) = \beta_0 + \beta_1 x$$

➢ We have been doing linear regression all along, but for the log-odds instead of probability.

Let's look at the coefficient β_1: odds = $\exp(\beta_0 + \beta_1 x)$ $= e^{\beta_0 + \beta_1 x}$

With a unit increase in x, we get: odds = $\exp\big(\beta_0 + \beta_1 (x+1)\big)$

$= e^{\beta_0 + \beta_1 (x+1)} = e^{\beta_0 + \beta_1 x + \beta_1} = \underbrace{e^{\beta_1}}\, e^{\beta_0 + \beta_1 x}$

So we have a new interpretation of the Logistic Regression weight β_1:

For a unit increase of 1, the odds change by e^{β_1}

Logistic regression

The Logistic Regression model with a single feature looks like:
$p(y = 1 \mid x) = sigm(\beta_0 + \beta_1 x)$

But in real life we typically have many features

<u>**Example**</u>:
Predict the probability of precipitation
Features: temperature, pressure, humidity, wind speed, whether it rained yesterday …

Multiple features Logistic Regression model:

$$p(y = 1 \mid x) = sigm\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \ldots + \beta_p x_p\right)$$

Logistic regression

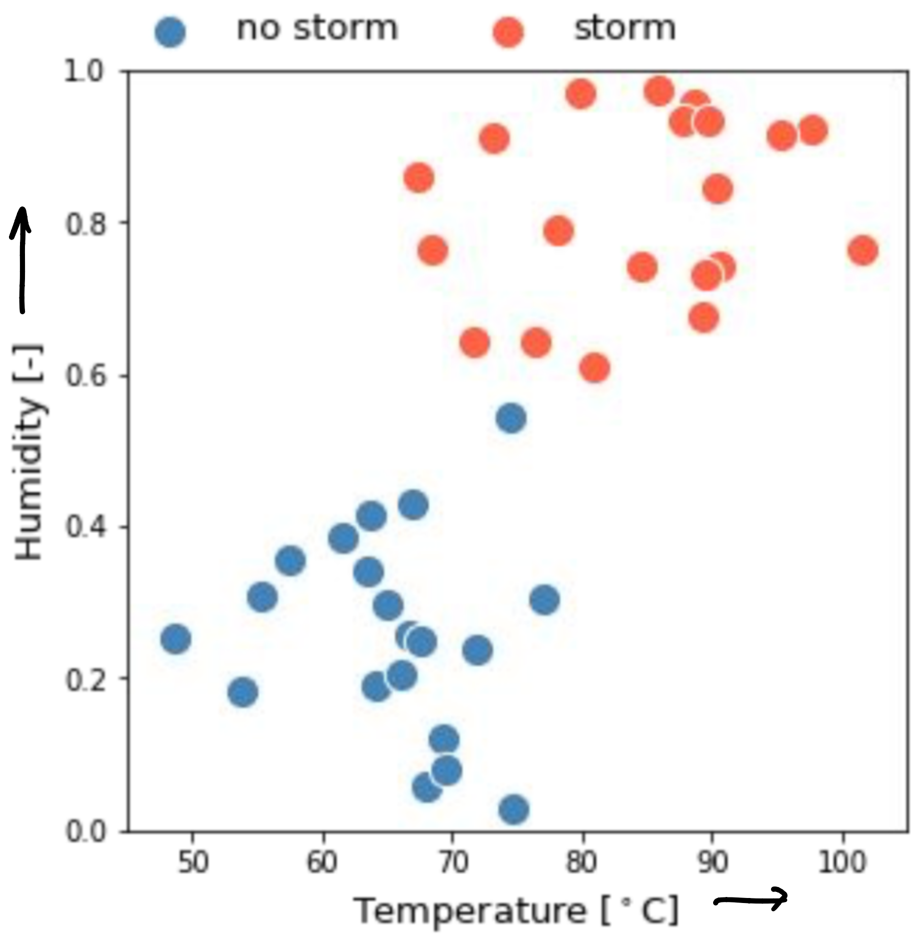
Example:
Predict the probability of precipitation
Features: temperature, pressure, humidity, wind speed, whether it rained yesterday …

Multiple features Logistic Regression model:

$$p(y = 1 \mid x) = sigm(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$

Predict: y = 1 = storm
y = 0 = no storm

Features: x_1 = temperature
x_2 = humidity



Logistic regression

Example:

Predict the probability of precipitation

Features: temperature, pressure, humidity, wind speed, whether it rained yesterday …

Multiple features Logistic Regression model:

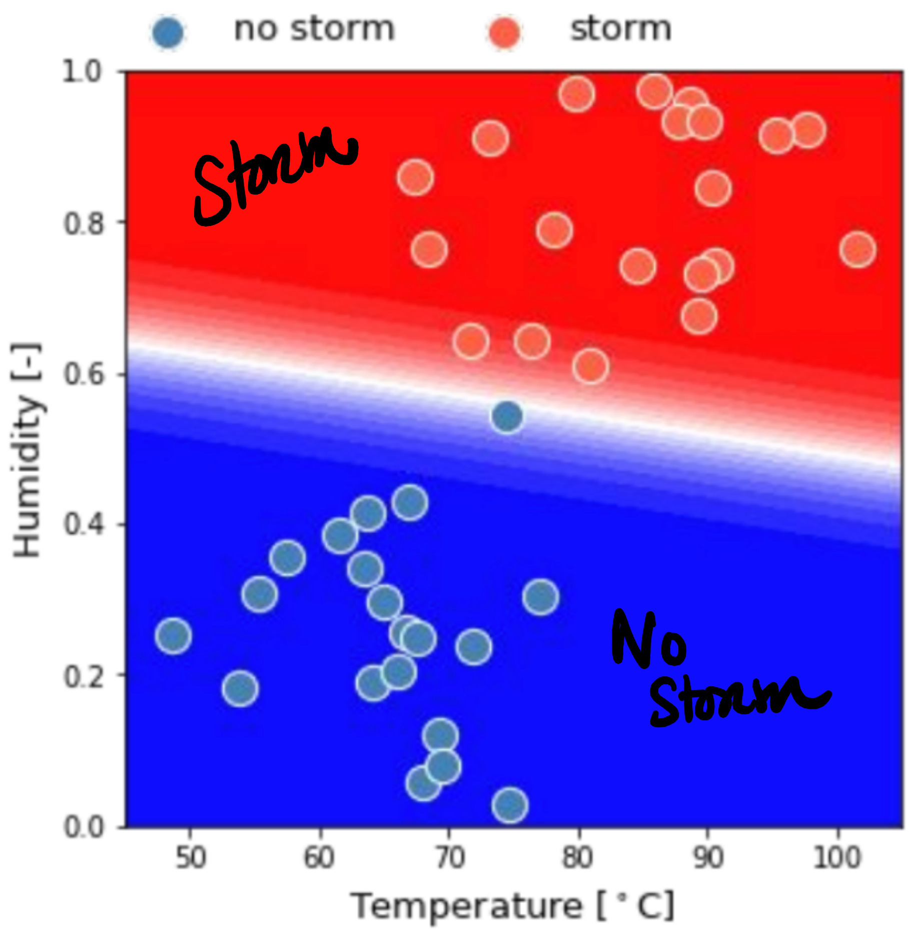
$$p(y = 1 \mid x) = \underbrace{sigm}(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$

Predict: y = 1 = storm

y = 0 = no storm

Features: x_1 = temperature

x_2 = humidity



Decision Boundary is a line

Above the Decision Boundary, predict class 1

Below the decision Boundary, predict class 2

Logistic regression

Titantic $\frac{1}{1+e^{-z}}$ At $z = 0$ $\frac{1}{1+e^{-0}} = \frac{1}{2}$

The decision boundary is the line/surface that separates predictions into Class 0 and Class 1

For a 2-feature model, it is described by:

$$p(y = 1 \mid x) = \underbrace{sigm(\beta_0 + \beta_1 x_1 + \beta_2 x_2)} = 0.5$$

Which is just a line in 2D space:

$$\frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2)}} = \frac{1}{2}$$

$$1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2)} = 2$$

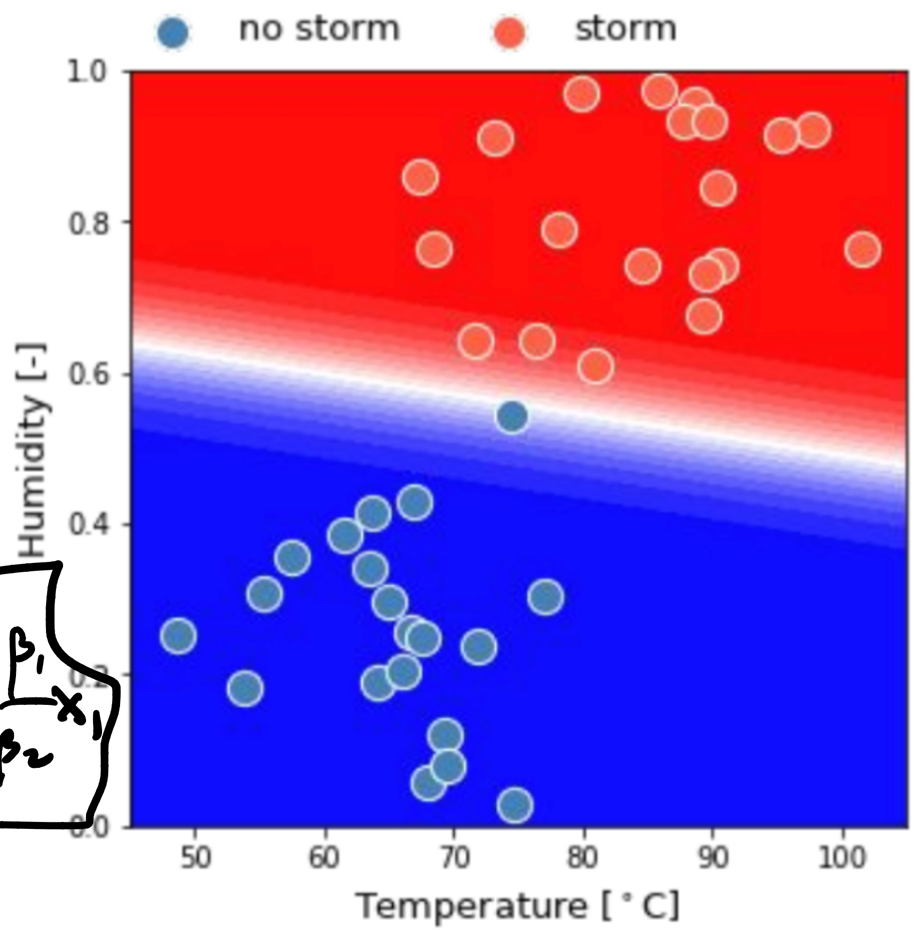
$$\ln\left(e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2)}\right) = \ln(1)$$

$$-(\beta_0+\beta_1 x_1+\beta_2 x_2) = 0$$

$$-\beta_0-\beta_1 x_1-\beta_2 x_2 = 0$$

$$-\beta_2 x_2 = \beta_0+\beta_1 x_1$$

$$\boxed{x_2 = -\frac{\beta_0}{\beta_2} - \frac{\beta_1}{\beta_2}x_1}$$



Logistic regression

The Sigmoid function has some nice differential properties that ~~we'll explore next time~~.

The most important of these is that:

$f' = p(1-p)$

If $f(z) = \mathrm{sigm}(z)$,

then $f'(z) = \mathrm{sigm}(z)(1 - \mathrm{sigm}(z))$

$f(z) = \frac{1}{1+e^{-z}}$

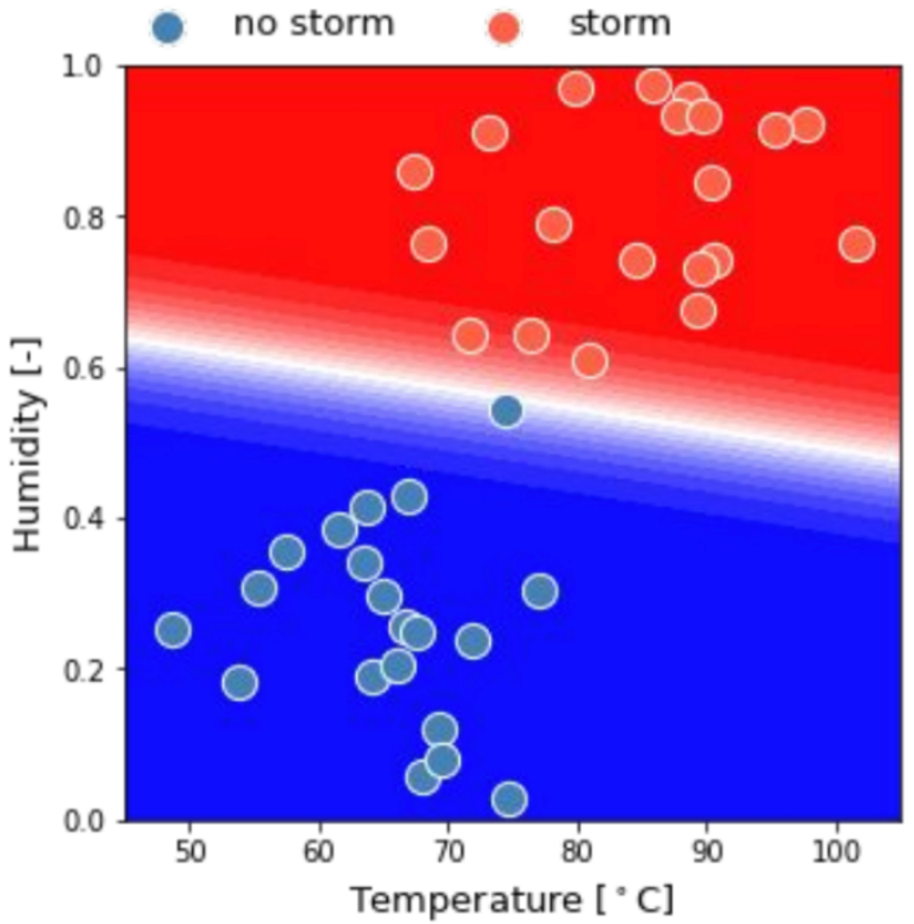
$f'(z) = \frac{-(-e^{-z})}{(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2}$

$= \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}}$

$= \mathrm{sign}(z) \cdot \frac{(1+e^{-z}-1)}{1+e^{-z}}$

$= \mathrm{sigm}(z) . \frac{1+e^{-z}}{1+e^{-z}} - \frac{1}{1+e^{-z}}$

$= \mathrm{sigm}(z)(1 - \mathrm{sigm}(z))$



Next Time:

- Review for Final Exam