Logistic Regression

HPSI
Human Performance Systems, Inc.

Logistic Regression – a few points

- Logistic Regression is also very popular for establishing a baseline
- Here are a few of the traits that make it popular:
    - Like Naïve Bayes, Logistic Regression is very fast
    - It’s highly interpretable
    - Doesn’t require features to be scaled
    - Doesn’t require any tuning
    - Easy to regularize
    - Outputs predicted probabilities

Logistic Regression – a few points

- Although Logistic Regression is popular, it is often misunderstood.
- Here are a few common questions that people have when they think of logistic regression:
  - Why is it called "logistic regression" if it's used for **classification**?
  - Why is it called a li**near model**?
  - How do you interpret the **model coefficients**?

Probability Reminder

- Probability

HPSI
Human Performance Systems, Inc.

Probability Reminder

- Probability
  - Measure of **Likelihood**, meaning how **likely** is something going to happen, how likely is something true

Probability Reminder

- Probability
  - Measure of **Likelihood**, meaning how **likely** is something going to happen, how likely is something true
  - Expressed as a number between 0-1
  - 0 ……………… 1
  - 0……………… 100%

Probability Reminder

- Probability
  - Measure of **Likelihood**, meaning how **likely** is something going to happen, how likely is something true
  - Expressed as a number between 0-1
  - 0 ………………. 1
  - 0………………. 100%

Definitely going to happen

HPSI
Human Performance Systems, Inc.

Probability Reminder

- Probability
  - Measure of **Likelihood**, meaning how **likely** is something going to happen, how likely is something true
  - Expressed as a number between 0-1
  - 0 ……………… 1
  - 0……………….. 100%

Definitely **NOT** going to happen

HPSI
Human Performance Systems, Inc.

Probability Reminder

- Probability
  - Measure of **Likelihood**, meaning how **likely** is something going to happen, how likely is something true
  - Expressed as a number between 0-1
  - 0 ………5……… 1
  - 0……….5……… 100%

In the case of a **binary outcome** we have some midpoint where it's equally probable/likely

HPSI
Human Performance Systems, Inc.

ODDS

- Probability
    - Another measure of probability is **ODDS**

HPSI
Human Performance Systems, Inc.

ODDS

- Probability
  - Another measure of probability is **ODDS**
  - $$\boldsymbol{Probability} = \frac{one\ outcome}{\boldsymbol{ALL}\ \boldsymbol{possible}\ outcome}$$

HPSI Human Performance Systems, Inc.

ODDS

- Probability
  - Another measure of probability is **ODDS**
  - $\boldsymbol{Probability} = \dfrac{one\ outcome}{\boldsymbol{ALL\ possible}\ outcome}$
  - $\boldsymbol{Odds} = \dfrac{one\ outcome}{\boldsymbol{ALL\ OTHER}\ outcomes}$

HPSI
Human Performance Systems, Inc.

Back to the coinflip

- Flip a coin
  - $\boldsymbol{P\ (Heads)} = \frac{one\ outcome}{ALL\ possible\ outcome} = \frac{1}{1+1} = \frac{H}{H+T} = 0.5$
  - $\boldsymbol{O(Heads)} = \frac{one\ outcome}{ALL\ \boldsymbol{OTHER}\ outcome} = \frac{1}{1} = \frac{H}{T} = 1$
  - Ratio Representation

HPSI
Human Performance Systems, Inc.

Back to the coinflip

- Flip a coin
  - $\boldsymbol{P\ (Heads)} = \frac{one\ outcome}{ALL\ possible\ outcome} = \frac{1}{1+1} = \frac{H}{H+T} = 0.5$
  - $\boldsymbol{O(Heads)} = \frac{one\ outcome}{ALL\ \boldsymbol{OTHER}\ outcome} = \frac{1}{1} = \frac{H}{T} = 1$
  - Ratio Representation
    - Odds can be represented as (1:1)

Back to the coinflip

- Flip a coin
  - $\boldsymbol{P\ (Heads)} = \frac{one\ outcome}{ALL\ possible\ outcome} = \frac{1}{1+1} = \frac{H}{H+T} = 0.5$
  - $\boldsymbol{O(Heads)} = \frac{one\ outcome}{ALL\ \boldsymbol{OTHER}\ outcome} = \frac{1}{1} = \frac{H}{T} = 1$
  - Ratio Representation
    - Odds can be represented as (1:1)
    - We say the odds are **even** because these numbers are the same

HPSI Human Performance Systems, Inc.

Rolling a Die

- Roll a Die
  - We are interested in rolling a 3
  - $\boldsymbol{P}\,(\text{"}\mathbf{3}\text{"}) = \dfrac{one\ outcome}{ALL\ possible\ outcome} = \dfrac{\{3\}}{\{1,2,3,4,5,6\}} = \dfrac{1}{6} = 0.1667$
  - $\boldsymbol{O}(\text{"}\mathbf{3}\text{"}) = \dfrac{one\ outcome}{ALL\ \boldsymbol{OTHER}\ outcome} = \dfrac{\{3\}}{\{1,2,4,5,6\}} = \dfrac{1}{5} = 0.2$

HPSI
Human Performance Systems, Inc.

Rolling a Die

- Roll a Die
  - We are interested in rolling a 3
  - $\boldsymbol{P}\ (\text{"}\mathbf{3}\text{"}) = \dfrac{one\ outcome}{ALL\ possible\ outcome} = \dfrac{\{3\}}{\{1,2,3,4,5,6\}} = \dfrac{1}{6} = 0.1667$
  - $\boldsymbol{O}(\text{"}\mathbf{3}\text{"}) = \dfrac{one\ outcome}{ALL\ \boldsymbol{OTHER}\ outcome} = \dfrac{\{3\}}{\{1,2,4,5,6\}} = \dfrac{1}{5} = 0.2$
  - Ratio Representation
    - Odds can be represented as (0.2:1)

Rolling a Die

- Roll a Die
  - We are interested in rolling a 3
  - $$\boldsymbol{P}\ (\text{"}\mathbf{3}\text{"}) = \frac{one\ outcome}{ALL\ possible\ outcome} = \frac{\{3\}}{\{1,2,3,4,5,6\}} = \frac{1}{6} = 0.1667$$
  - $$\boldsymbol{O}(\text{"}\mathbf{3}\text{"}) = \frac{one\ outcome}{ALL\ \boldsymbol{OTHER}\ outcome} = \frac{\{3\}}{\{1,2,4,5,6\}} = \frac{1}{5} = 0.2$$
  - Ratio Representation
    - Odds can be represented as (0.2:1)
    - We say the odds are **less than even**

HPSI
Human Performance Systems, Inc.

Rolling a Die

- Roll a Die
  - Now say we are interested in rolling <=5
  - $\boldsymbol{P}\ (\leq \mathbf{5}) = \dfrac{one\ outcome}{ALL\ possible\ outcome} = \dfrac{\{5,4,3,2,1\}}{\{6,5,4,3,2,1\}} = \dfrac{5}{6} = 0.83$
  - $\boldsymbol{O}(\leq \mathbf{5}) = \dfrac{one\ outcome}{ALL\ \boldsymbol{OTHER}\ outcome} = \dfrac{\{5,4,3,2,1\}}{\{6\}} = \dfrac{5}{1} = 5$

Rolling a Die

- Roll a Die
  - Now say we are interested in rolling <=5
  - $\boldsymbol{P}\ (\leq \mathbf{5}) = \frac{one\ outcome}{ALL\ possible\ outcome} = \frac{\{5,4,3,2,1\}}{\{6,5,4,3,2,1\}} = \frac{5}{6} = 0.83$
  - $\boldsymbol{O}(\leq \mathbf{5}) = \frac{one\ outcome}{ALL\ \boldsymbol{OTHER}\ outcome} = \frac{\{5,4,3,2,1\}}{\{6\}} = \frac{5}{1} = 5$
  - Ratio Representation
    - Odds can be represented as (5:1)

HPSI Human Performance Systems, Inc.

Rolling a Die

- Roll a Die
  - Now say we are interested in rolling <=5
  - $\boldsymbol{P}\ (\leq \mathbf{5}) = \dfrac{one\ outcome}{ALL\ possible\ outcome} = \dfrac{\{5,4,3,2,1\}}{\{6,5,4,3,2,1\}} = \dfrac{5}{6} = 0.83$
  - $\boldsymbol{O}(\leq \mathbf{5}) = \dfrac{one\ outcome}{ALL\ \boldsymbol{OTHER}\ outcome} = \dfrac{\{5,4,3,2,1\}}{\{6\}} = \dfrac{5}{1} = 5$
  - Ratio Representation
    - Odds can be represented as (5:1)
    - We say the odds are **better than even** (in your favor)

HPSI
Human Performance Systems, Inc.

Ranges for Odds and Probability

$$0 \le Odds \le \infty$$

$$0 \le Probability \le 1$$

Odds Ratio Example

- Smokers, non-smokers study
- Interested in studying smoking/lung cancer, 2 groups
- Group 1 (Case Group) – Have lung cancer
  - 8 Smokers
  - 2 Non-smokers
- Group 2 (Control Group) – Don't have lung cancer
  - 4 Smokers
  - 6 Non-smokers

Odds Ratio Example

- Smokers, non-smokers study
- Interested in studying smoking/lung cancer, 2 groups
- Group 1 (Case Group) – Have lung cancer
  - 8 Smokers
  - 2 Non-smokers
- Group 2 (Control Group) – Don’t have lung cancer
  - 4 Smokers
  - 6 Non-smokers

	Case	Control
Smokers	8	4
Non-Smokers	2	6

HPSI
Human Performance Systems, Inc.

Odds Ratio Example

- Smokers, non-smokers study
- Interested in studying smoking/lung cancer, 2 groups
- Group 1 (Case Group) – Have lung cancer
  - 8 Smokers
  - 2 Non-smokers
- Group 2 (Control Group) – Don’t have lung cancer
  - 4 Smokers
  - 6 Non-smokers

	Case	Control
Smokers	8	4
Non-Smokers	2	6

$$Odds\ of\ Smoking, Case\ Group = \frac{8}{2} = 4$$

Odds Ratio Example

- Smokers, non-smokers study
- Interested in studying smoking/lung cancer, 2 groups
- Group 1 (Case Group) – Have lung cancer
    - 8 Smokers
    - 2 Non-smokers
- Group 2 (Control Group) – Don't have lung cancer
    - 4 Smokers
    - 6 Non-smokers

	Case	Control
Smokers	8	4
Non-Smokers	2	6

$$Odds\ of\ Smoking, Case\ Group = \frac{8}{2} = 4$$

$$Odds\ of\ Smoking, Control\ Group = \frac{4}{6} = 0.66$$

HPSI
Human Performance Systems, Inc.

Odds Ratio Example

- Smokers, non-smokers study
- Interested in studying smoking/lung cancer, 2 groups
- Group 1 (Case Group) – Have lung cancer
  - 8 Smokers
  - 2 Non-smokers
- Group 2 (Control Group) – Don't have lung cancer
  - 4 Smokers
  - 6 Non-smokers

	Case	Control
Smokers	8	4
Non-Smokers	2	6

$$Odds\ of\ Smoking, Case\ Group = \frac{8}{2} = 4$$

$$Odds\ of\ Smoking, Control\ Group = \frac{4}{6} = 0.66$$

$$Odds\ Ratio = \frac{Odds\ Case\ Group}{Odds\ Control\ Group} = \frac{4}{0.66} = 6$$

HPSI
Human Performance Systems, Inc.

Odds Ratio Example

- Smokers, non-smokers study
- Interested in studying smoking/lung cancer, 2 groups
- Group 1 (Case Group) – Have lung cancer
  - 8 Smokers
  - 2 Non-smokers
- Group 2 (Control Group) – Don’t have lung cancer
  - 4 Smokers
  - 6 Non-smokers
- How can we compare the odds?

	Case	Control
Smokers	8	4
Non-Smokers	2	6

$$Odds\ of\ Smoking, Case\ Group = \frac{8}{2} = 4$$

$$Odds\ of\ Smoking, Control\ Group = \frac{4}{6} = 0.66$$

$$Odds\ Ratio = \frac{Odds\ Case\ Group}{Odds\ Control\ Group} = \frac{4}{0.66} = 6$$

HPSI
Human Performance Systems, Inc.

Odds Ratio Example

- Smokers, non-smokers study
- Interested in studying smoking/lung cancer, 2 groups
- Group 1 (Case Group) – Have lung cancer
  - 8 Smokers
  - 2 Non-smokers
- Group 2 (Control Group) – Don't have lung cancer
  - 4 Smokers
  - 6 Non-smokers
- How can we compare the odds?
- Conclusion:
  - The Case group was 6 times more likely to be a smoker
  - Making no claims about causation
  - Simply that people who have lung cancer are 6 times more likely to have smoked

	Case	Control
Smokers	8	4
Non-Smokers	2	6

$$Odds\ of\ Smoking, Case\ Group = \frac{8}{2} = 4$$

$$Odds\ of\ Smoking, Control\ Group = \frac{4}{6} = 0.66$$

$$Odds\ Ratio = \frac{Odds\ Case\ Group}{Odds\ Control\ Group} = \frac{4}{0.66} = 6$$

- From Wikipedia:
  - *"The natural logarithm, formerly known as the hyperbolic logarithm, is the logarithm to the base e, where e is an irrational constant approximately equal to 2.718281828459."*

- From Wikipedia:
  - *"The natural logarithm, formerly known as the hyperbolic logarithm, is the logarithm to the base e, where e is an irrational constant approximately equal to 2.718281828459."*
  - Describing ***e*** as "a constant approximately 2.71828…" is like calling **pi** "an irrational number, approximately equal to 3.1415…". Sure, it's true, but you completely missed the point.

- A Better way to think about π/ e
  - **π is the ratio between circumference and diameter shared by all circles**. It is a fundamental ratio inherent in all circles and therefore impacts any calculation of circumference, area, volume, and surface area for circles, spheres, cylinders, and so on. (not to mention the trigonometric functions derived from circles (sin, cos, tan).

- A Better way to think about π/ e
  - **π is the ratio between circumference and diameter shared by all circles**. It is a fundamental ratio inherent in all circles and therefore impacts any calculation of circumference, area, volume, and surface area for circles, spheres, cylinders, and so on. (not to mention the trigonometric functions derived from circles (sin, cos, tan).
  - **e is the base rate of growth shared by all continually growing processes.** e lets you take a simple growth rate (where all change happens at the end of the year) and find the impact of compound, continuous growth

e

- A Better way to think about π/ e
  - **π is the ratio between circumference and diameter shared by all circles**. It is a fundamental ratio inherent in all circles and therefore impacts any calculation of circumference, area, volume, and surface area for circles, spheres, cylinders, and so on. (not to mention the trigonometric functions derived from circles (sin, cos, tan).
  - **e is the base rate of growth shared by all continually growing processes.** e lets you take a simple growth rate (where all change happens at the end of the year) and find the impact of compound, continuous growth
  - e shows up whenever systems grow exponentially and continuously: population, radioactive decay, interest calculations, and more.

- A Better way to think about π / e
  - **π is the ratio between circumference and diameter shared by all circles**. It is a fundamental ratio inherent in all circles and therefore impacts any calculation of circumference, area, volume, and surface area for circles, spheres, cylinders, and so on. (not to mention the trigonometric functions derived from circles (sin, cos, tan).
  - **e is the base rate of growth shared by all continually growing processes.** e lets you take a simple growth rate (where all change happens at the end of the year) and find the impact of compound, continuous growth
  - e shows up whenever systems grow exponentially and continuously: population, radioactive decay, interest calculations, and more.
  - Just like every number can be considered a scaled version of 1 (the base unit), every circle can be considered a scaled version of the unit circle (radius 1), and every rate of growth can be considered a scaled version of e (unit growth, perfectly compounded).

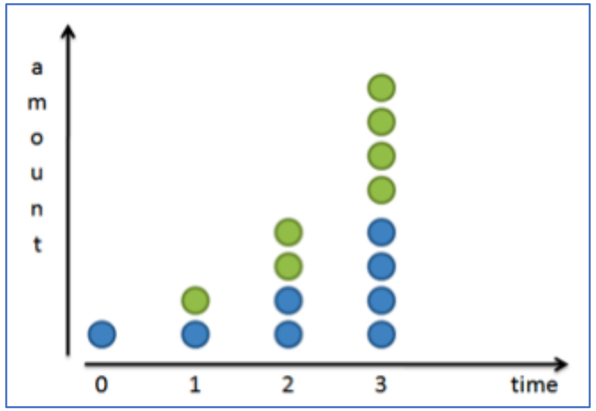
Exponential Growth

➢ *Understanding Exponential Growth*

- ❖ Let's look at a basic system that **doubles** after an amount of time. For example,
- ❖ Bacteria can split and "doubles" every 24 hours
- ❖ We get twice as many noodles when we fold them in half.
- ❖ Your money doubles every year if you get 100% return

Exponential Growth

- *Understanding Exponential Growth*
  - Let's look at a basic system that **doubles** after an amount of time. For example,
  - Bacteria can split and "doubles" every 24 hours
  - We get twice as many noodles when we fold them in half.
  - Your money doubles every year if you get 100% return
- **It looks like this:**

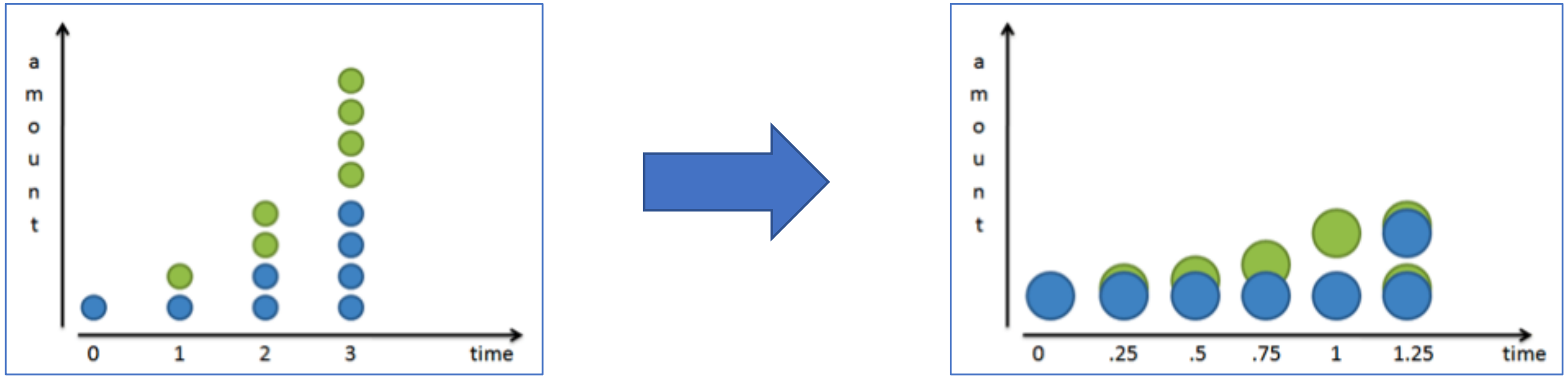


Exponential Growth

➢ *Understanding Exponential Growth*

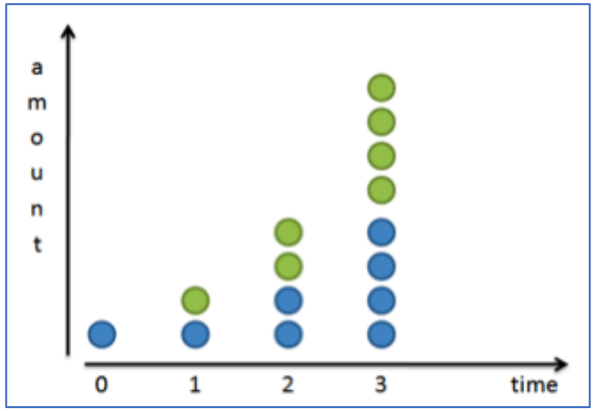
- ❖Let's look at a basic system that **doubles** after an amount of time. For example,
- ❖Bacteria can split and "doubles" every 24 hours
- ❖We get twice as many noodles when we fold them in half.
- ❖Your money doubles every year if you get 100% return

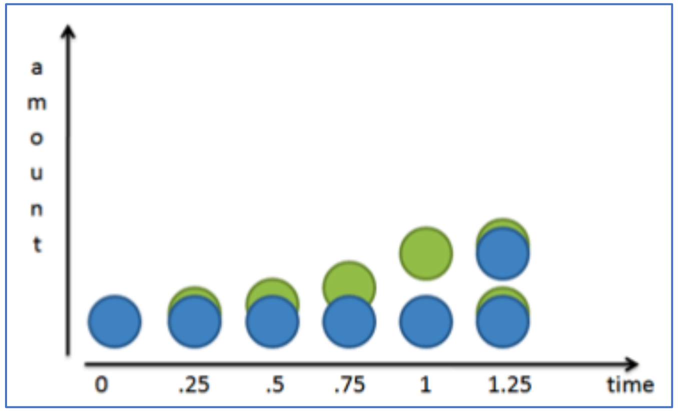
➢ **It looks like this:**



The secret to *e*

- ***The big secret: e merges rate and time***
  - This is wild, e^x can mean two things:
  - x is the number of times we multiply a growth rate: 100% growth for 3 years is e^3
  - x is the growth rate itself: 300% growth for one year is e^3





HPSI
Human Performance Systems, Inc.

The secret to *e*

- *The big secret: e merges rate and time*
  - This is wild, e^x can mean two things:
  - x is the number of times we multiply a growth rate: 100% growth for 3 years is e^3
  - x is the growth rate itself: 300% growth for one year is e^3
- And now you know why it's "e", and not pi or some other number: e raised to "r*t" gives you the growth impact of rate r and time t.

$$growth = e^x = e^{rt}$$

Demystifying the Natural Logarithm (ln)

- After understanding the exponential function, the next component required for logistic regression is the natural log
- If you look-up the Wikipedia definition, it's "the inverse of e^x"
- e and the Natural Log are twins:
  - e^x is the **amount** of continuous growth after a certain amount of time.
  - Natural Log (ln) is the amount of **time** needed to reach a certain level of continuous growth

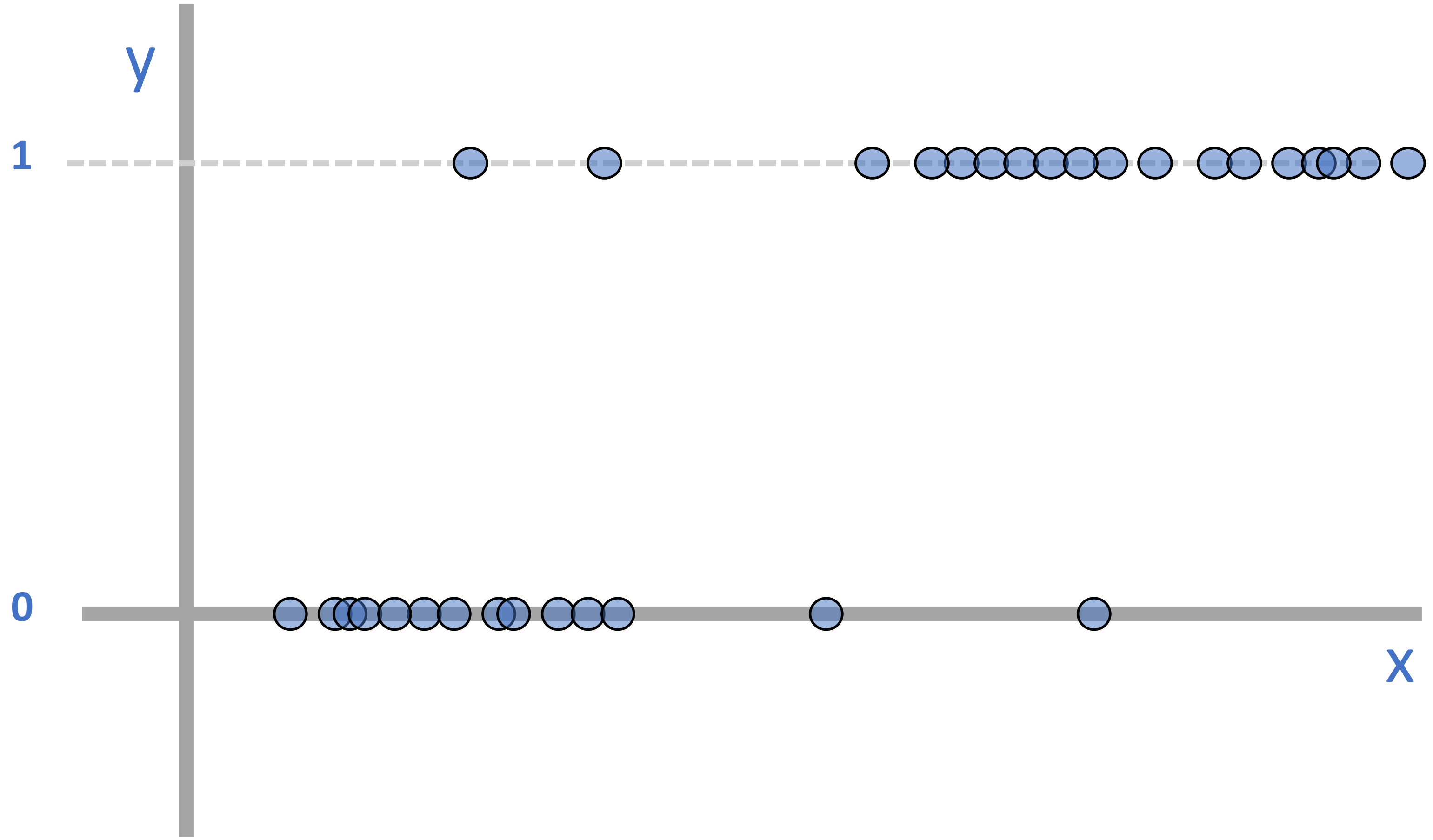
HPSI
Human Performance Systems, Inc.

Logistic Regression

- When we think of linear regression, we usually think of predicting a numeric quantity
- Logistic Regression is somewhat misleading in that it serves primarily as a binary classification model
- Categorical response (y) with 2 levels (binary: 0 and 1)
  - Passing or failing a test
  - Surviving a plane crash or not
  - Hospitalisation required or not
  - Diagnosis of diabetes (yes / no)
  - Labelling (over/under some threshold)
- Predictor variables (x_i) can take on any form: binary, categorical, and/or continuous

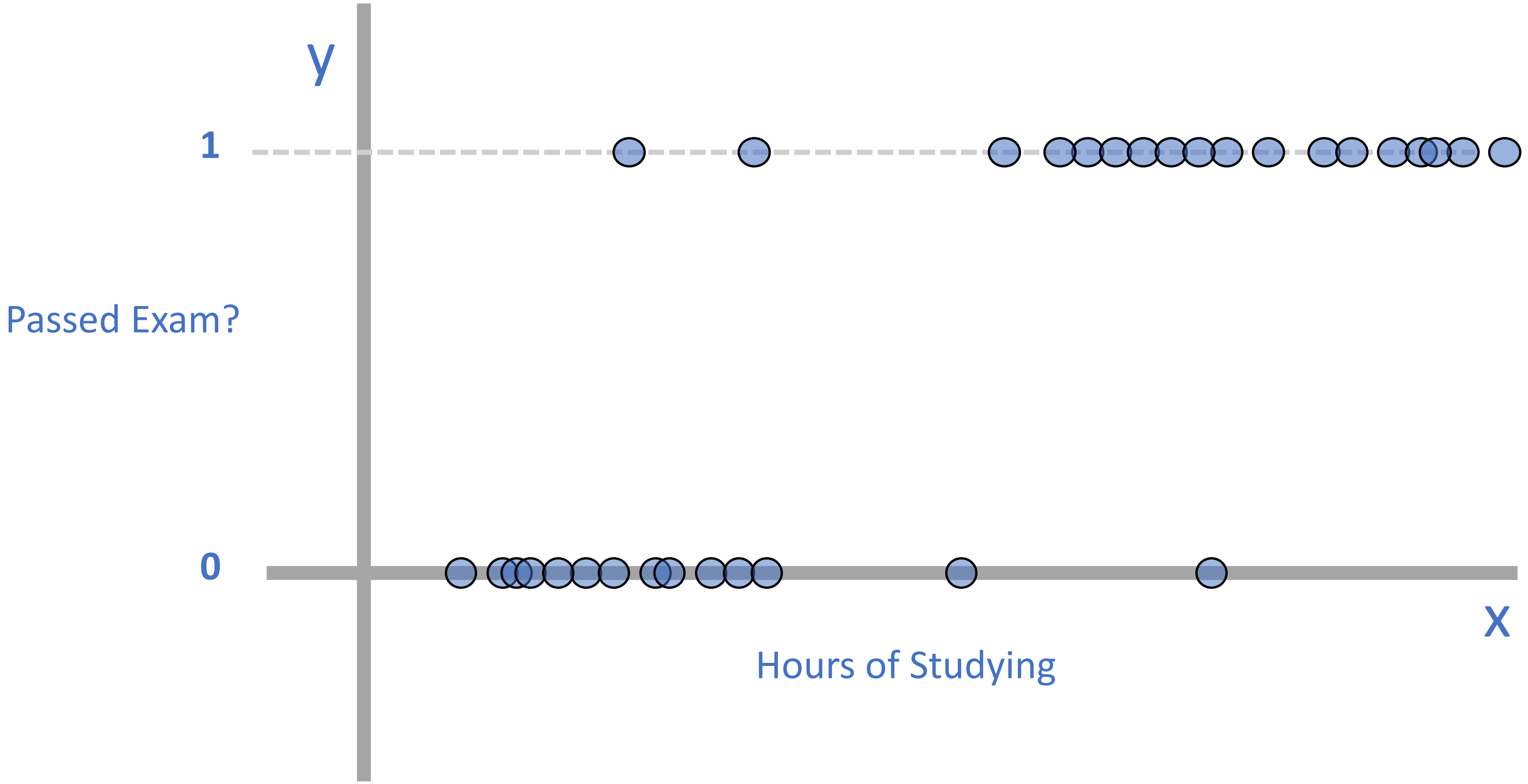
Logistic Regression Classification

Consider this set of binary data



HPSI Human Performance Systems, Inc.

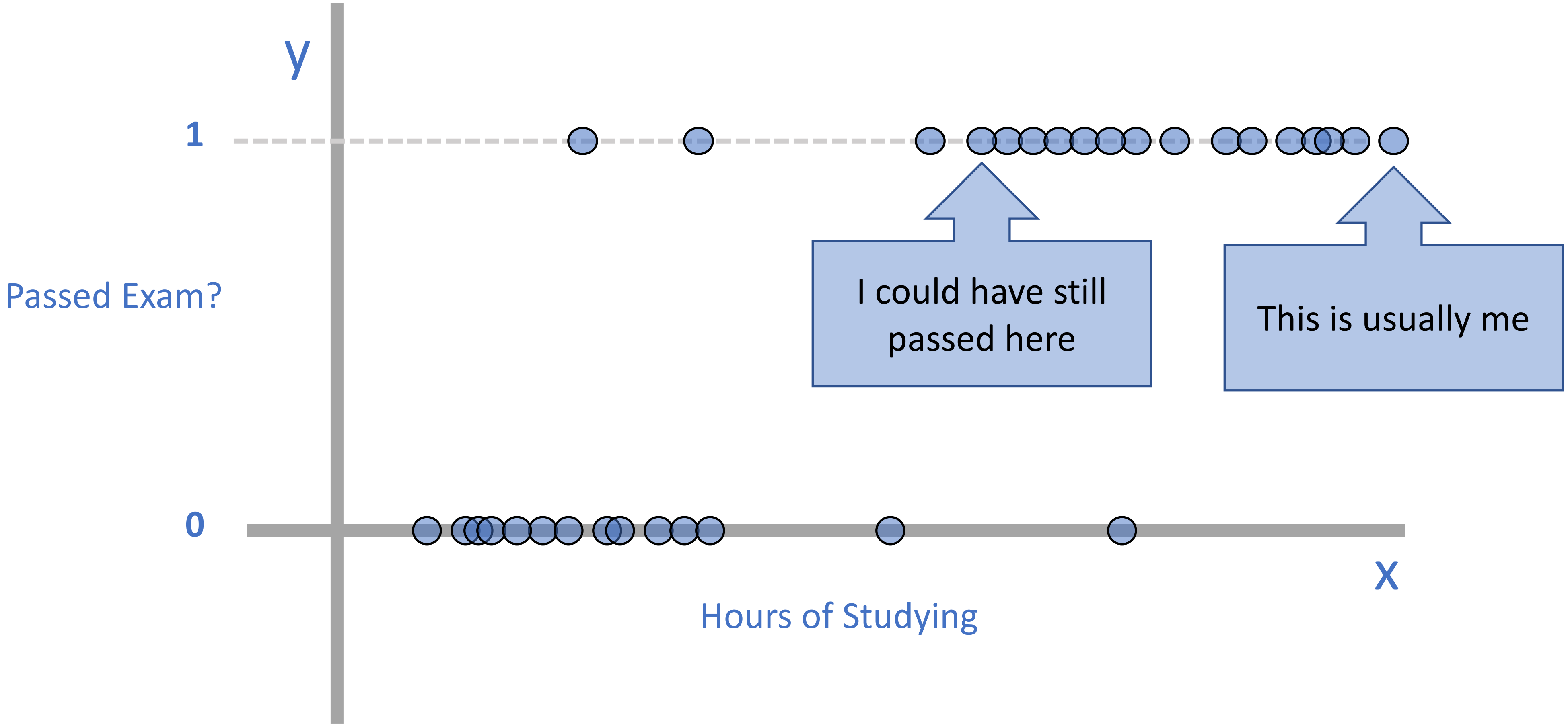
Consider this set of binary data



HPSI
Human Performance Systems, Inc.

Logistic Regression Classification

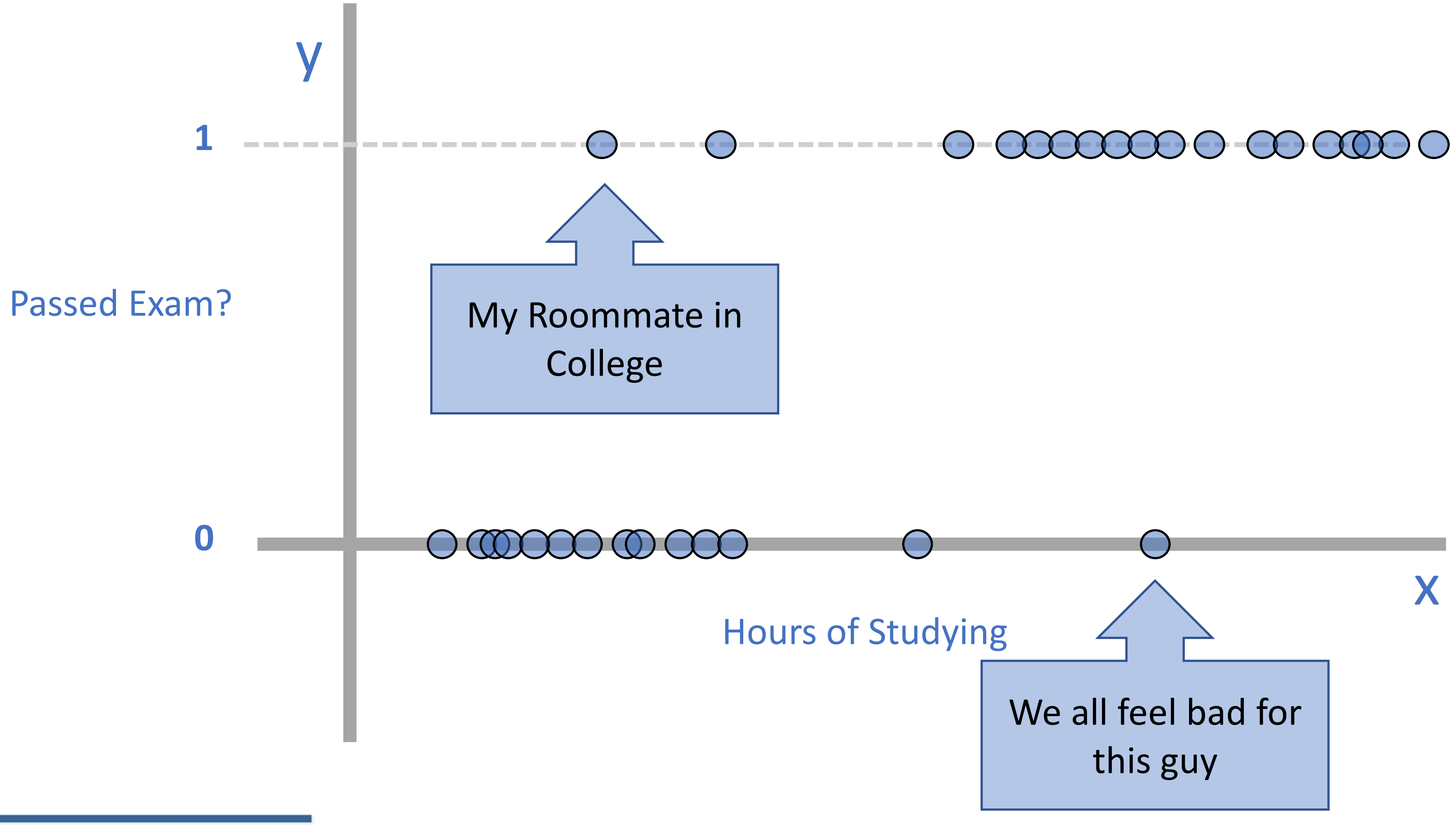
Consider this set of binary data



HPSI
Human Performance Systems, Inc.

Logistic Regression Classification

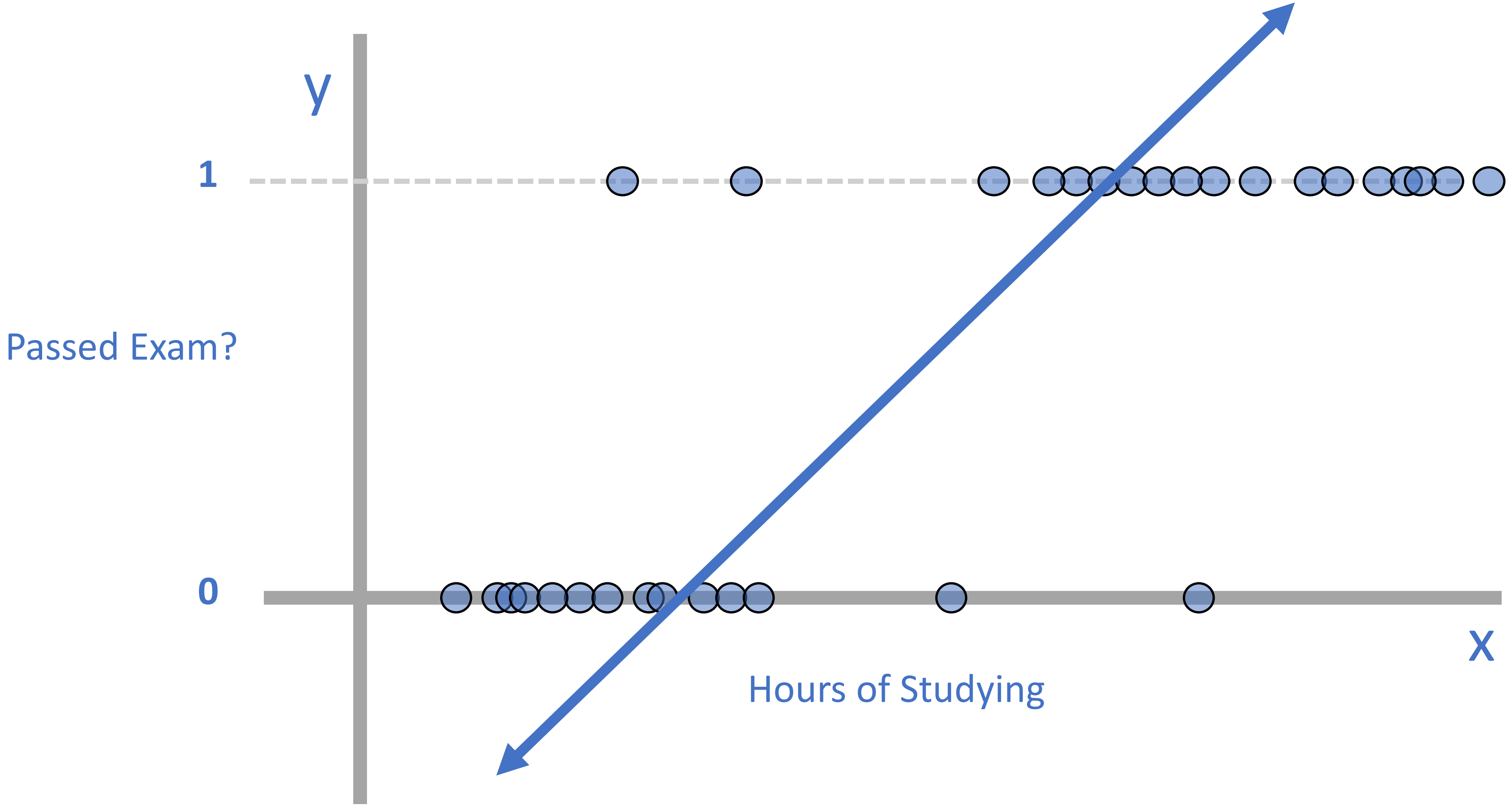
Consider this set of binary data



HPSI Human Performance Systems, Inc.

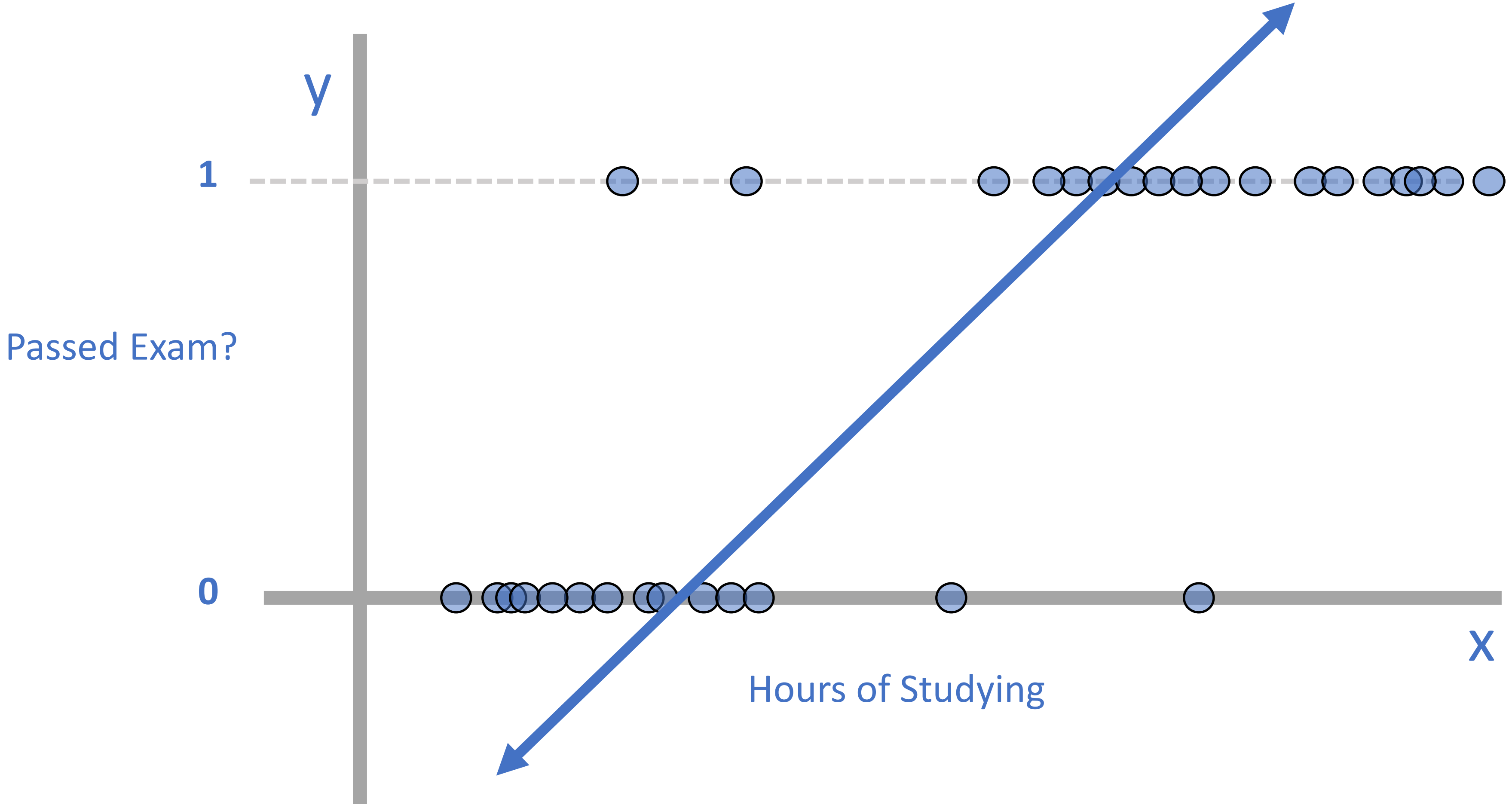
Logistic Regression Classification

- Linear Model? – Aside from being binary, there's nothing special about (y)



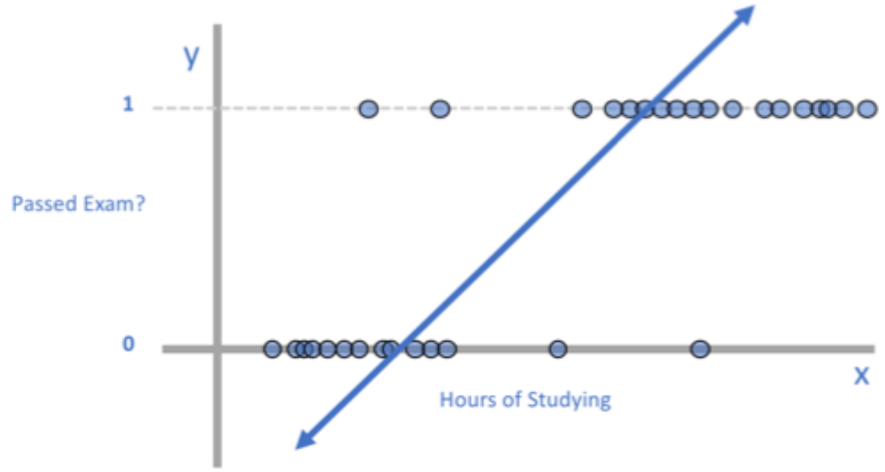
Logistic Regression Classification

- The value of “Passed Exam” is higher if a student studies more



Logistic Regression Classification

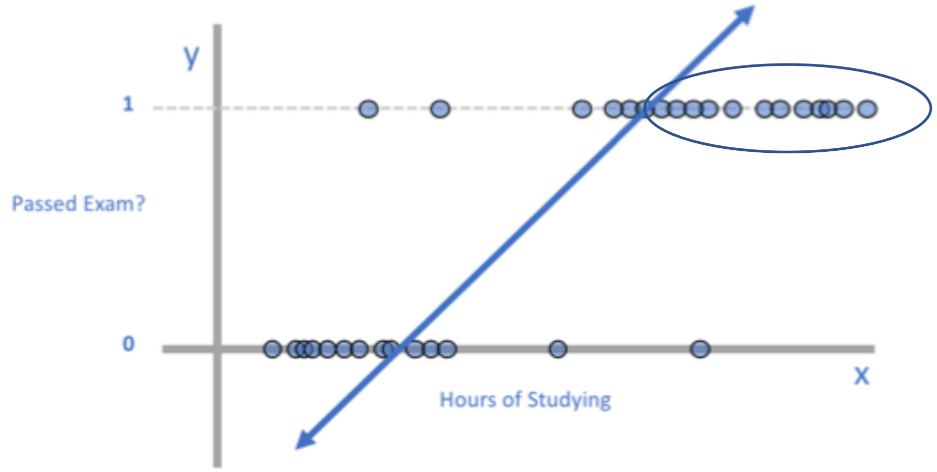
- **Linear Model**
  - $Pass = \beta_0 + \beta_1 hours\ of\ studying + \varepsilon$
- **Problem**
  - We want to see what makes the dependent variable change from a 0 to a 1
  - This can also be interpreted as what increases the likelihood of passing, or $P(pass = 1)$ which we can simply denote as p.
  - We should then be able to re-write the linear model as
  - $P(pass = 1) = p = \beta_0 + \beta_1 hours\ of\ studying + \varepsilon$
  - Every additional hour of studying increases the probability of passing by x%



Logistic Regression Classification

$$P(pass = 1) = \beta_0 + \beta_1 hours\ of\ studying + \ \varepsilon$$

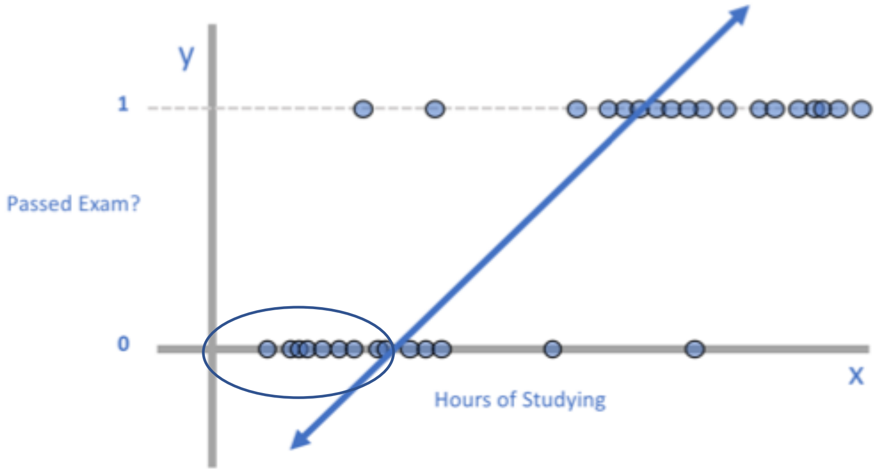
- Recall
  - Probabilities are bounded by $0 \leq p \leq 1$
  - If we were to look at students who studied this many hours, our model would predict a probability greater than 1



Logistic Regression Classification

$$P(pass = 1) = \beta_0 + \beta_1 hours\ of\ studying + \varepsilon$$

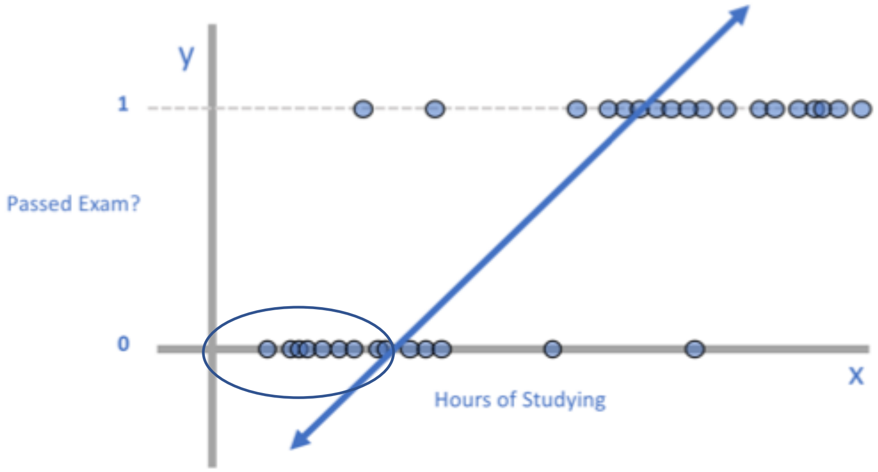
- Recall
    - Probabilities are bounded by $0 \leq p \leq 1$
    - If we were to look at students who studied this many hours, our model would predict a probability greater than 1
    - Likewise, here our model would predict a probability less than 0



Logistic Regression Classification

$$P(pass = 1) = \beta_0 + \beta_1 hours\ of\ studying + \varepsilon$$

- Recall
  - Probabilities are bounded by $0 \leq p \leq 1$
  - If we were to look at students who studied this many hours, our model would predict a probability greater than 1
  - Likewise, here our model would predict a probability less than 0
- In addition to violating the laws of probability, our model would not maintain normally distributed residuals (which is another requirement of a linear regression model)

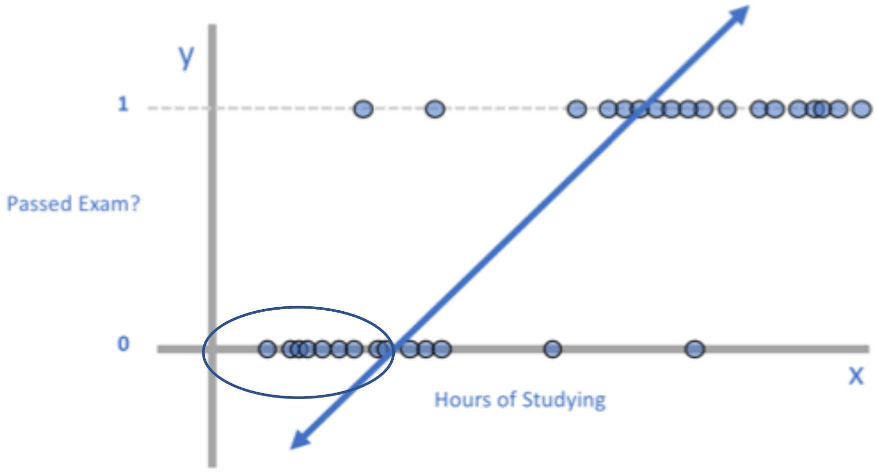


HPSI
Human Performance Systems, Inc.

Logistic Regression Classification

$$P(pass = 1) = \beta_0 + \beta_1 hours\ of\ studying + \ \varepsilon$$

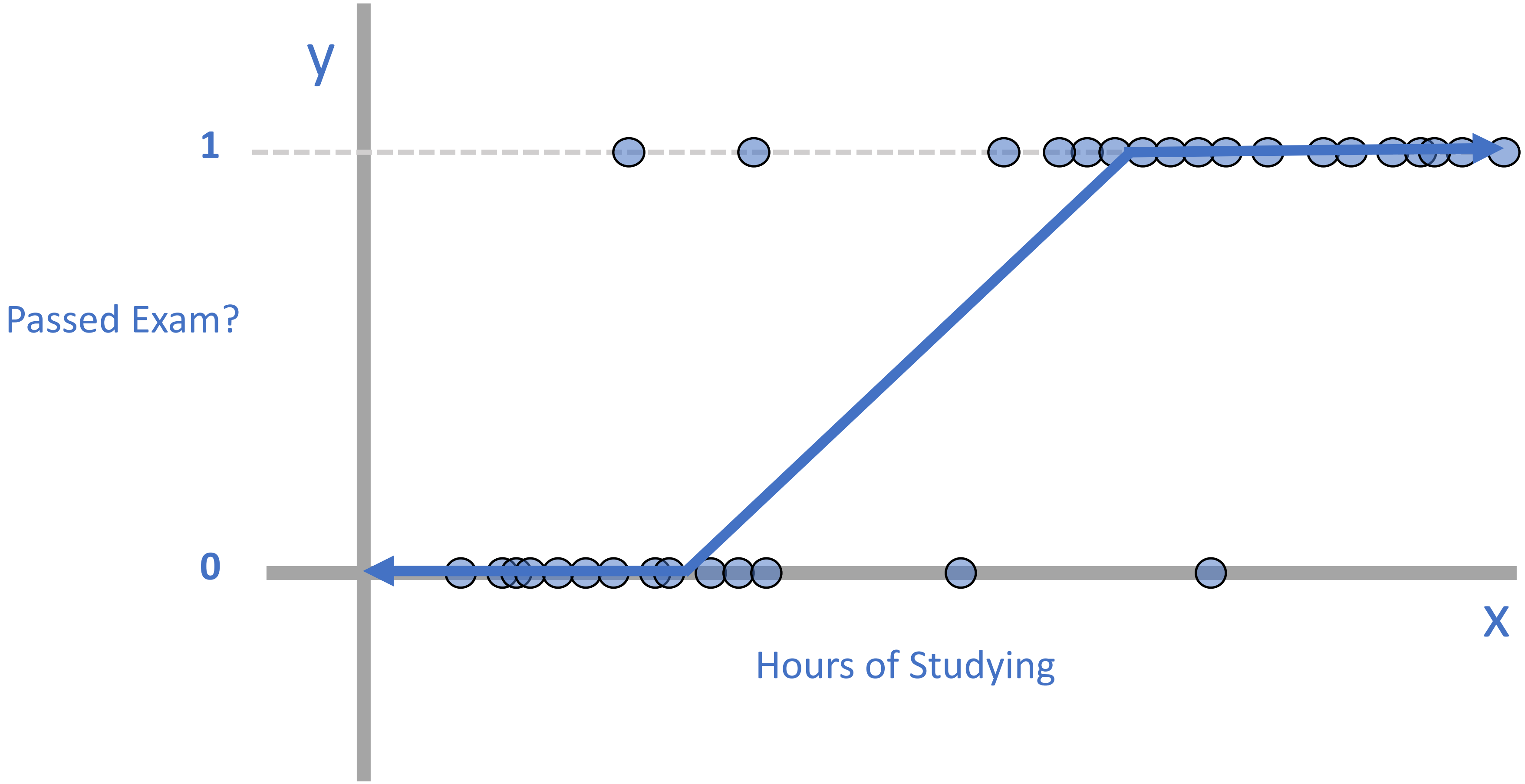
- Recall
  - Probabilities are bounded by $0 \leq p \leq 1$
  - If we were to look at students who studied this many hours, our model would predict a probability greater than 1
  - Likewise, here our model would predict a probability less than 0
- In addition to violating the laws of probability, our model would not maintain normally distributed residuals (which is another requirement of a linear regression model)
- What can we do to fix this?



HPSI
Human Performance Systems, Inc.

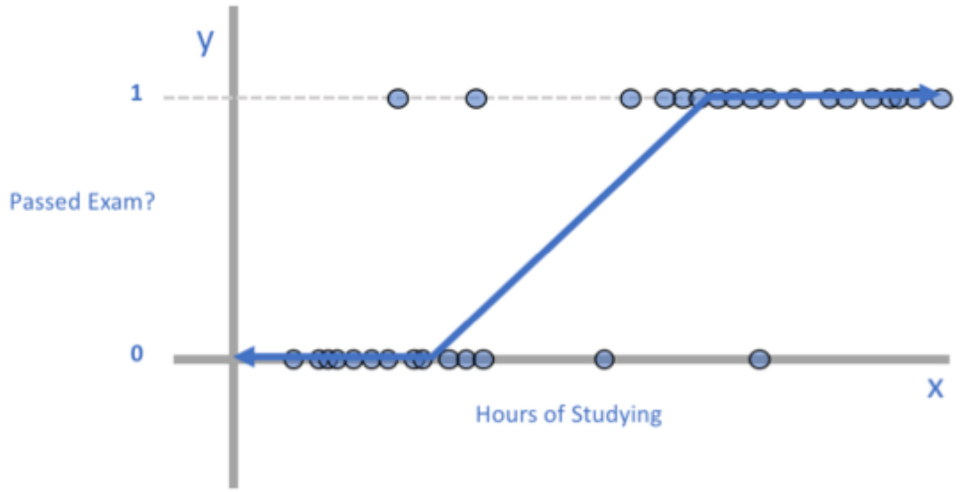
Logistic Regression Classification

- We could cap the probabilities at 0 and 1



Logistic Regression Classification

- Fixing the prior approach
  - We need to somehow constrain p such that $0 \leq p \leq 1$
  - We know $P(pass) = f(hours\ of\ studying)$ but the linear function didn't work
  - Let's try to develop a new function $f(hours\ of\ studying)$ that satisfies this criteria



HPSI
Human Performance Systems, Inc.

Logistic Regression Classification

- Two Steps
  1. It must always be positive (since $0 \leq p(pass)$)
     - $|\ x\ |$?
     - x^2 ?
     - What about $p(pass) = e^{\beta_0+\beta_1 * hours\ of\ studying}$?
     - This works, but there are times when it would be greater than 1
  2. It must always be less than 1 ($p(pass) \leq 1$)
     - If you think about proportions, any number that is divided by a number slightly greater than it will give us a number smaller than 1
     - What if we just add 1 to the denominator?
     - $p(pass) = \frac{(e^{\beta_0+\beta_1 * hours\ of\ studying})}{(e^{\beta_0+\beta_1 * hours\ of\ studying})+1}$
     - Note that we could have added any small number (ε) and this condition would have been met, but we use 1 for reasons that will become clear shortly

HPSI Human Performance Systems, Inc.

Logistic Regression Classification

- The previous expression:

$$p(pass) = p = \frac{(e^{\beta_0 + \beta_1 * hours\ of\ studying})}{(e^{\beta_0 + \beta_1 * hours\ of\ studying}) + 1}$$

- After applying some algebra, can be re-written as:

$$\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 * hours\ of\ studying$$

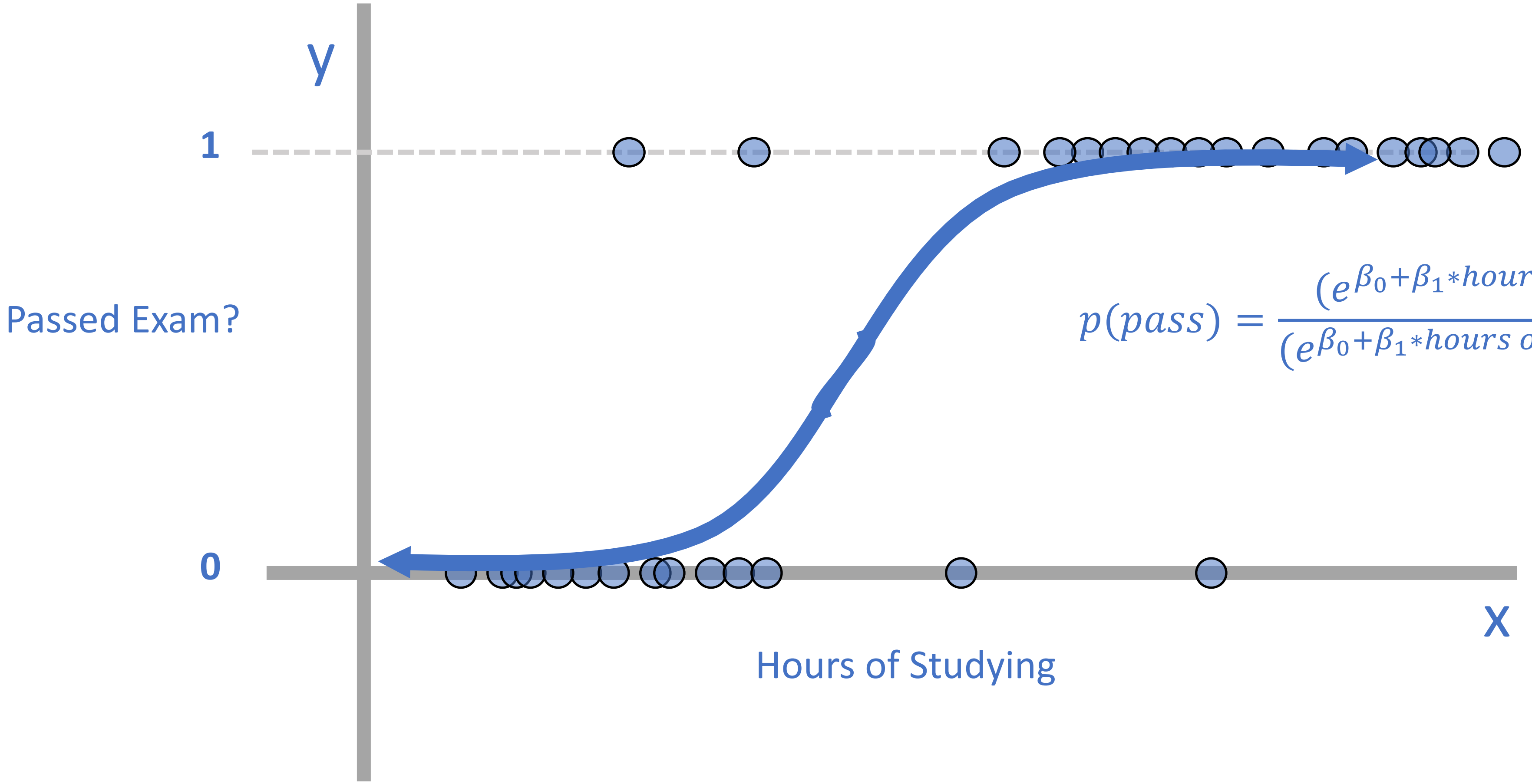
HPSI
Human Performance Systems, Inc.

Logistic Regression Classification

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 * hours\ of\ studying$$

- Does this look familiar?
  - Yes!
  - It's in the form of a standard linear model
- So, even though the probability of a student passing is not a linear function of study-hours, the simple transformation is a linear function of study-hours
- This is the equation used in logistic regression

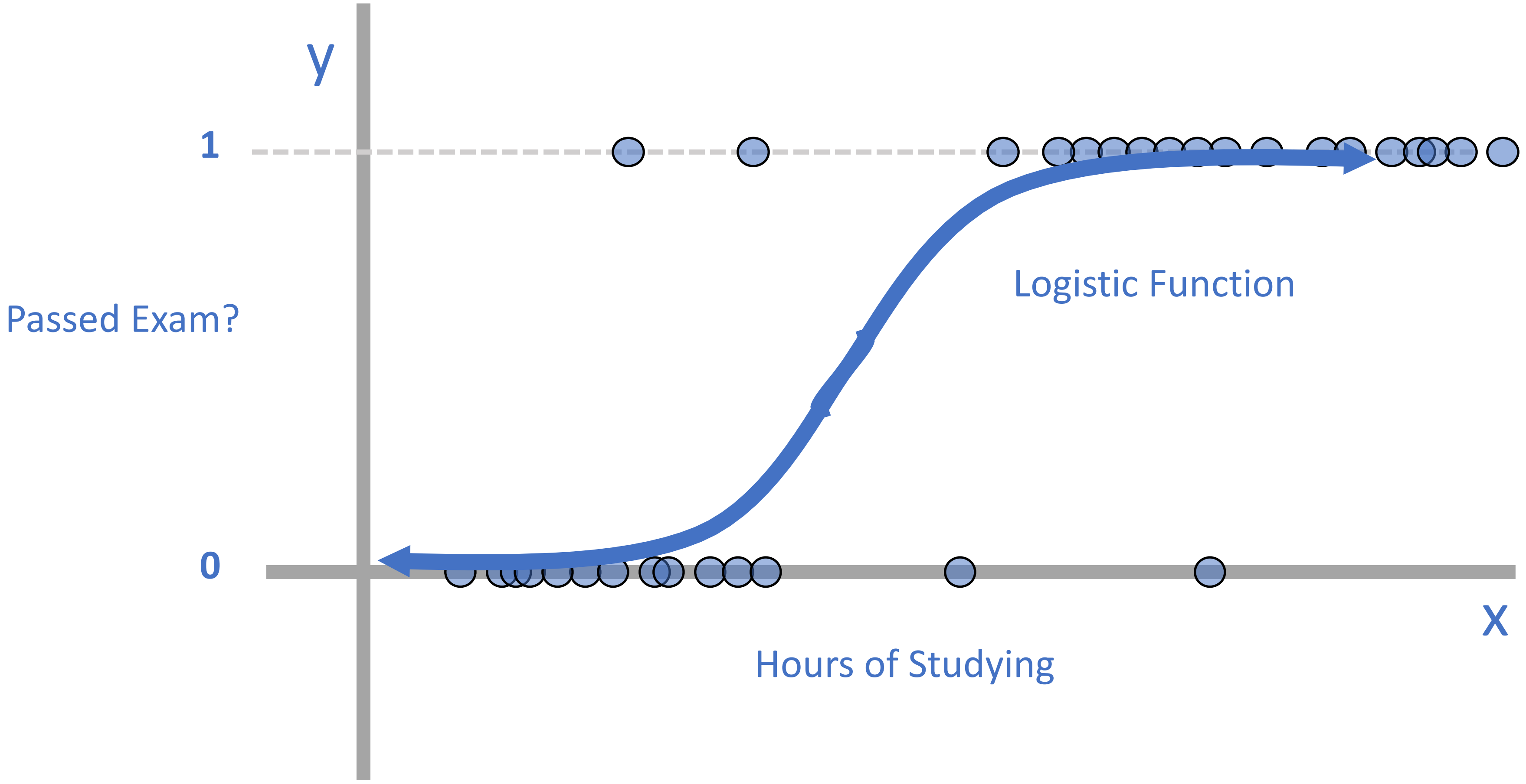
Logistic Regression



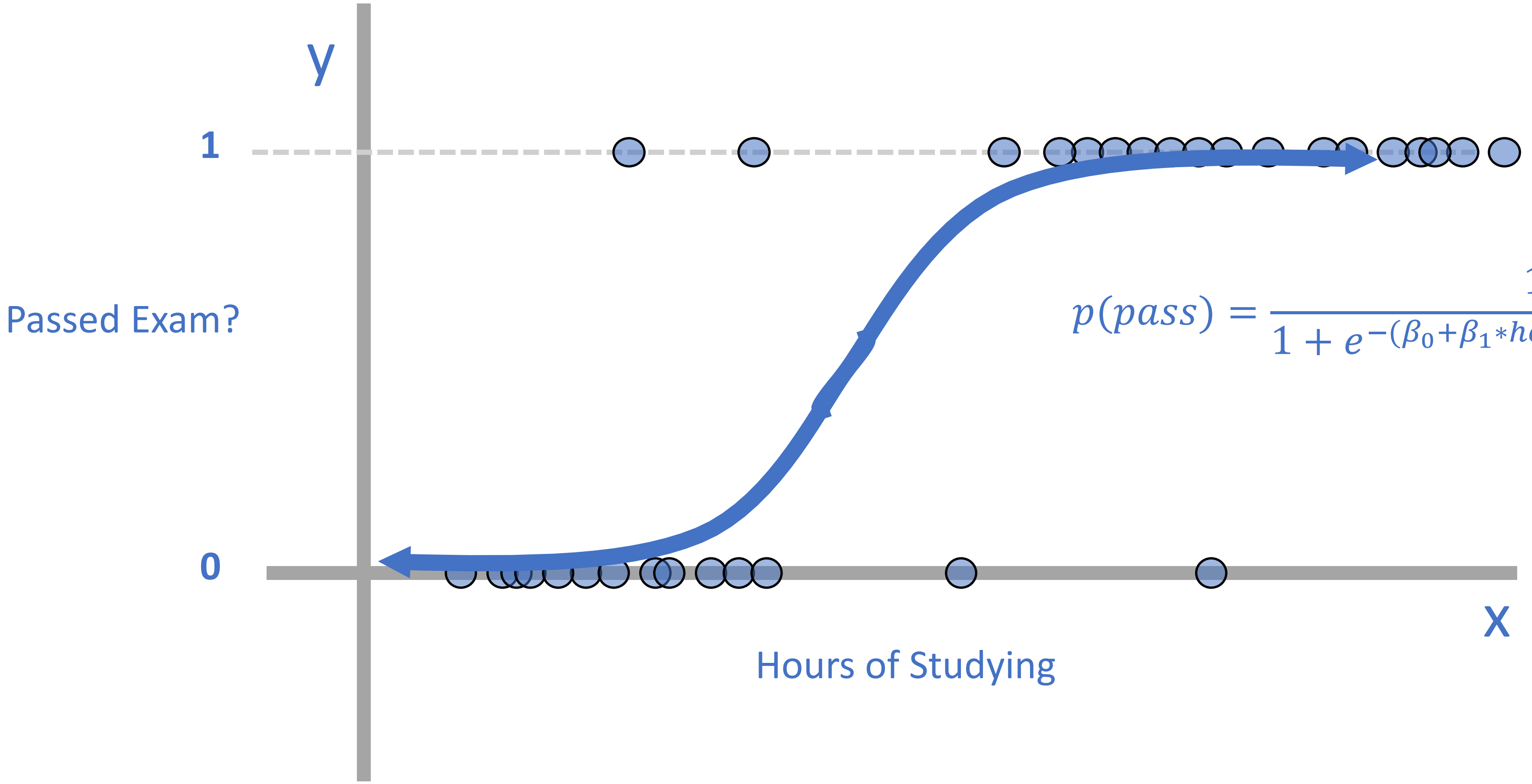
$$p(pass) = \frac{(e^{\beta_0+\beta_1 * hours\ of\ studying})}{(e^{\beta_0+\beta_1 * hours\ of\ studying}) + 1}$$

Logistic Regression Classification

Logistic Regression



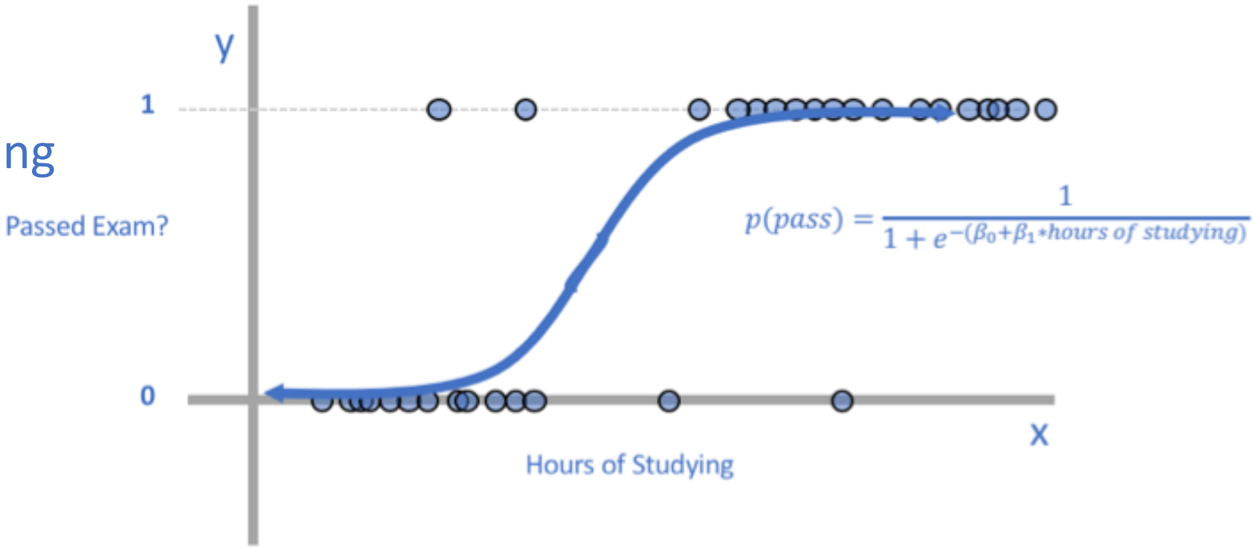
Logistic Regression



$$p(pass) = \frac{1}{1 + e^{-(\beta_0+\beta_1 * hours\ of\ studying)}}$$

Logistic Regression Classification

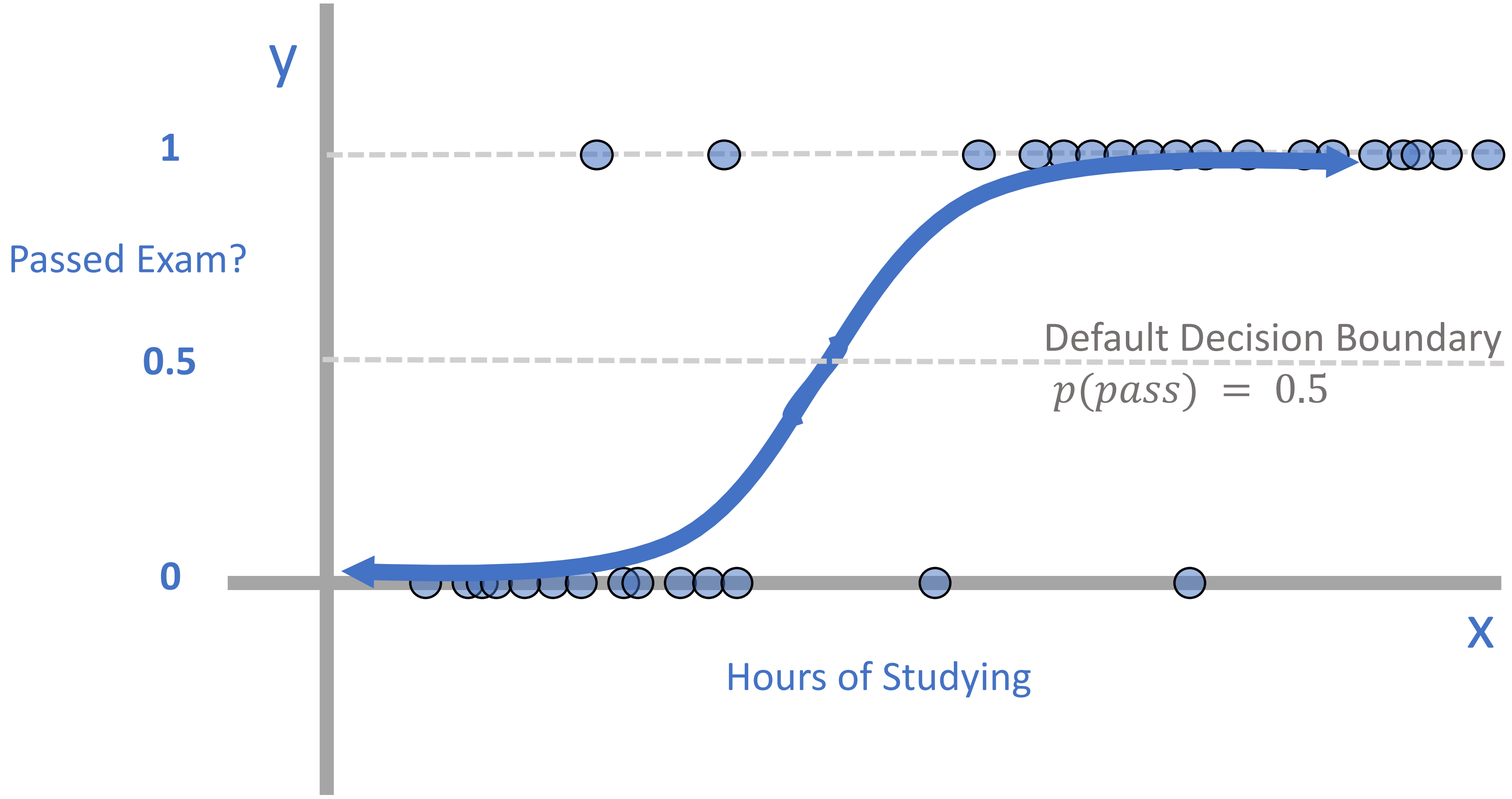
- Note that the probability of passing is now between 0 and 1
  - $0 \leq p \leq 1$
- We now have a continuous function
- As study-hours approach 0, the probability of passing goes (asymptotically) to zero
- As study-hours approach infinity, probability of passing goes (asymptotically) to 1



HPSI
Human Performance Systems, Inc.

Logistic Regression Classification

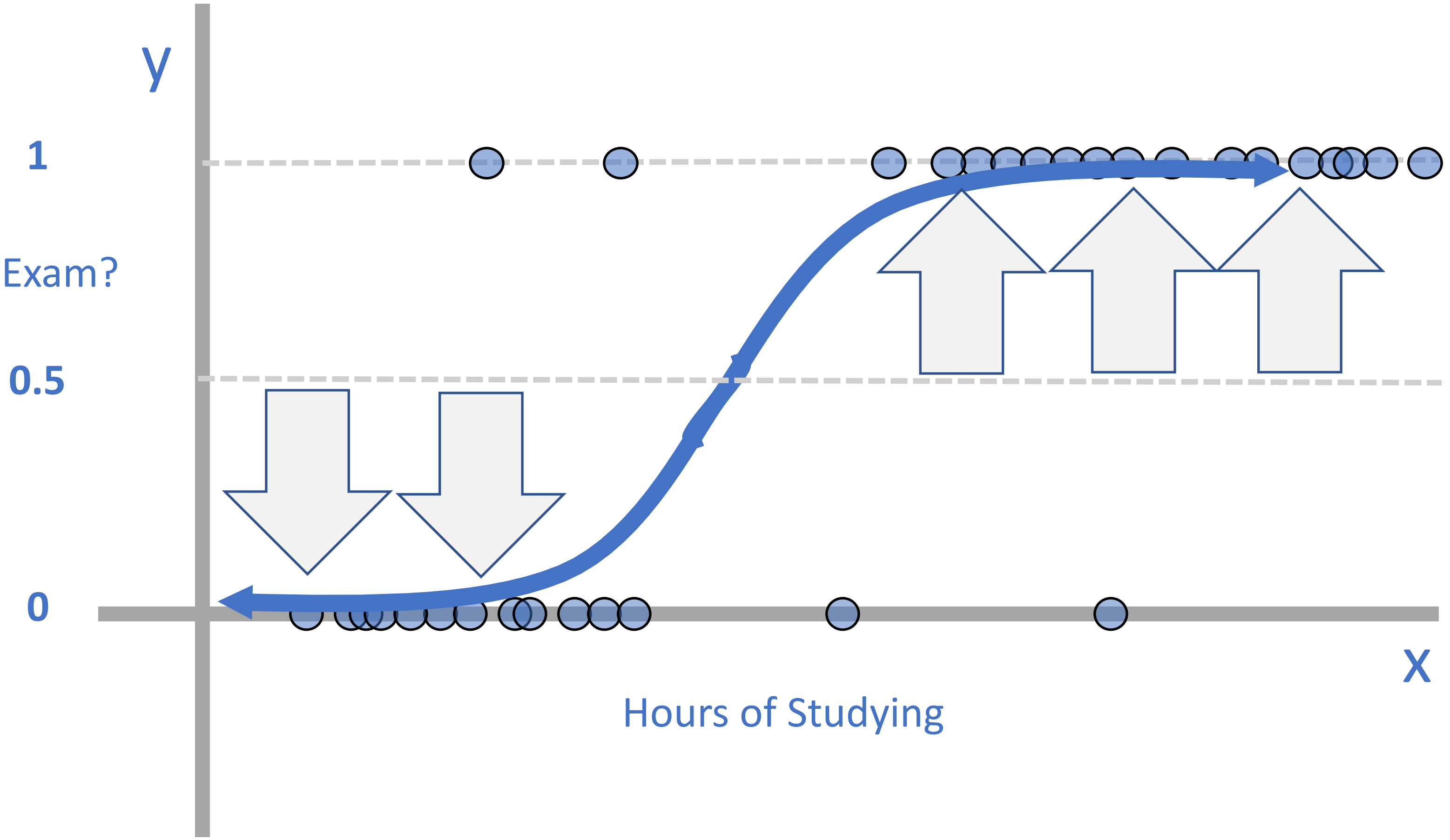
Decision Boundary



HPSI
Human Performance Systems, Inc.

Logistic Regression Classification

Decision Boundary



HPSI Human Performance Systems, Inc.

Interpreting Binary Logistic Regression Model Output

- Interpreting the coefficients from a Logistic Regression model is different from a Linear Regression model
- Consider these results from a fitted logistic regression model
  - $\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 * study\_hrs$
  - $\beta_0 = -4.077$
  - $\beta_1 = 1.5046$

HPSI
Human Performance Systems, Inc.

Interpreting Binary Logistic Regression Model Output

- Interpreting the coefficients from a Logistic Regression model is different from a Linear Regression model
- Consider these results from a fitted logistic regression model
  - $\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 * study\_hrs$
  - $\beta_0 = -4.077$
  - $\beta_1 = 1.5046$
  - $\ln\left(\frac{p}{1-p}\right) = -4.077 + 1.5046 * study\_hrs$

HPSI Human Performance Systems, Inc.

Interpreting Binary Logistic Regression Model Output

- Interpreting the coefficients from a Logistic Regression model is different from a Linear Regression model
- Consider these results from a fitted logistic regression model
  - $\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 * study\_hrs$
  - $\beta_0 = -4.077$
  - $\beta_1 = 1.5046$
  - $\ln\left(\frac{p}{1-p}\right) = -4.077 + 1.5046 * study\_hrs$
  - For every additional unit increase in $study\_hrs$, $\ln\left(\frac{p}{1-p}\right)$ increases by 1.5046 units?

Interpreting Binary Logistic Regression Model Output

- Interpreting the coefficients from a Logistic Regression model is different from a Linear Regression model
- Consider these results from a fitted logistic regression model
  - $\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 * study\_hrs$
  - $\beta_0 = -4.077$
  - $\beta_1 = 1.5046$
  - $\ln\left(\frac{p}{1-p}\right) = -4.077 + 1.5046 * study\_hrs$
  - For every additional unit increase in $study\_hrs$, $\ln\left(\frac{p}{1-p}\right)$ increases by 1.5046 units?
  - But what does that mean?

Interpreting Binary Logistic Regression Model Output

- If we have

$$y^* = logit = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 * study\_hrs$$

- y^* is called the "logit" function
  - A logit is defined as the natural log of the odds
- Exponentiating both sides of equation results in:

$$\frac{p}{1-p} = e^{(\beta_0 + \beta_1 * study\_hrs\ )}$$

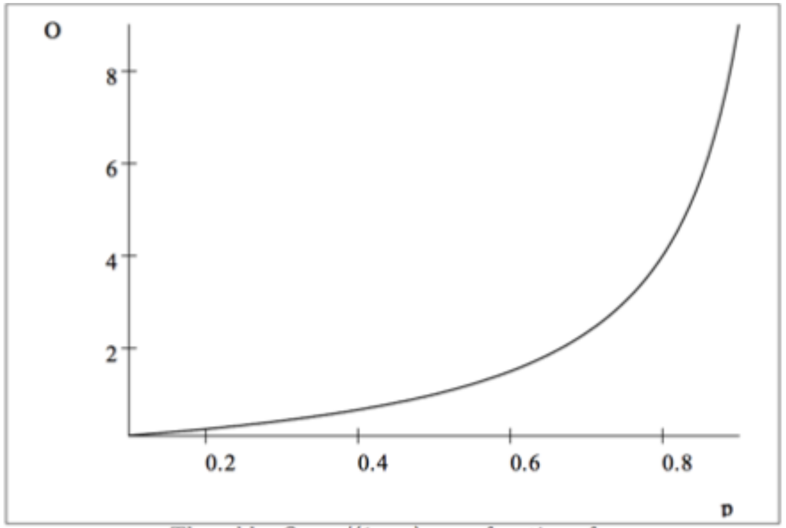
- We can exponentiate each of the coefficients to generate their corresponding odds ratios

Interpreting Binary Logistic Regression Model Output

- Odds

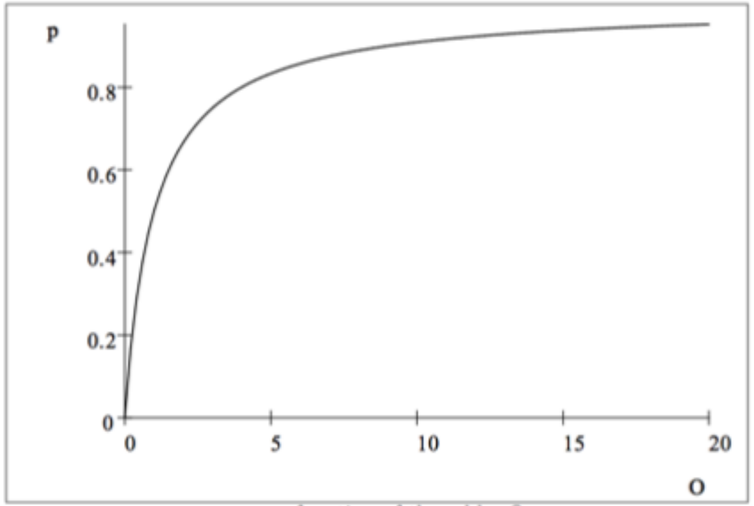
$$Odds(O) = \frac{P}{1-P}$$

  - The odds ratio tells you how a unit increase or decrease in the corresponding input variable affects the odds of passing the test
    - When the odds ratio is greater than 1, it describes a positive relationship
    - An odds ratio less than 1 implies a negative relationship



The odds, $O = p/(1-p)$, as a function of p

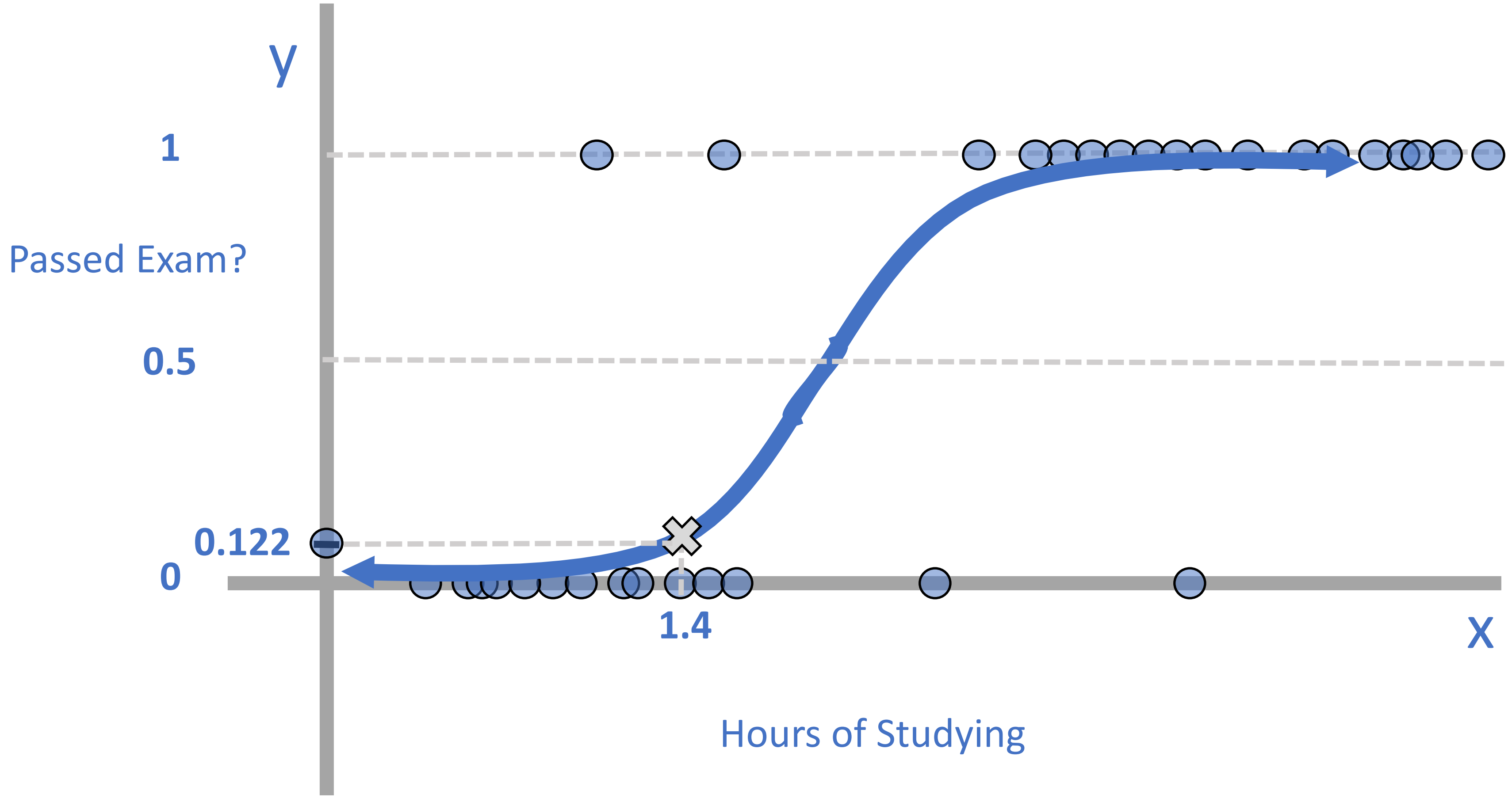
Note that $0 \leq O$, and O is undefined for $p = 1$. Solving $O = \frac{p}{1-p}$ for p, $p = \frac{O}{(O+1)}$



p as a function of the odds, O

Logistic Regression Classification

Visualizing the Logistic Regression Model



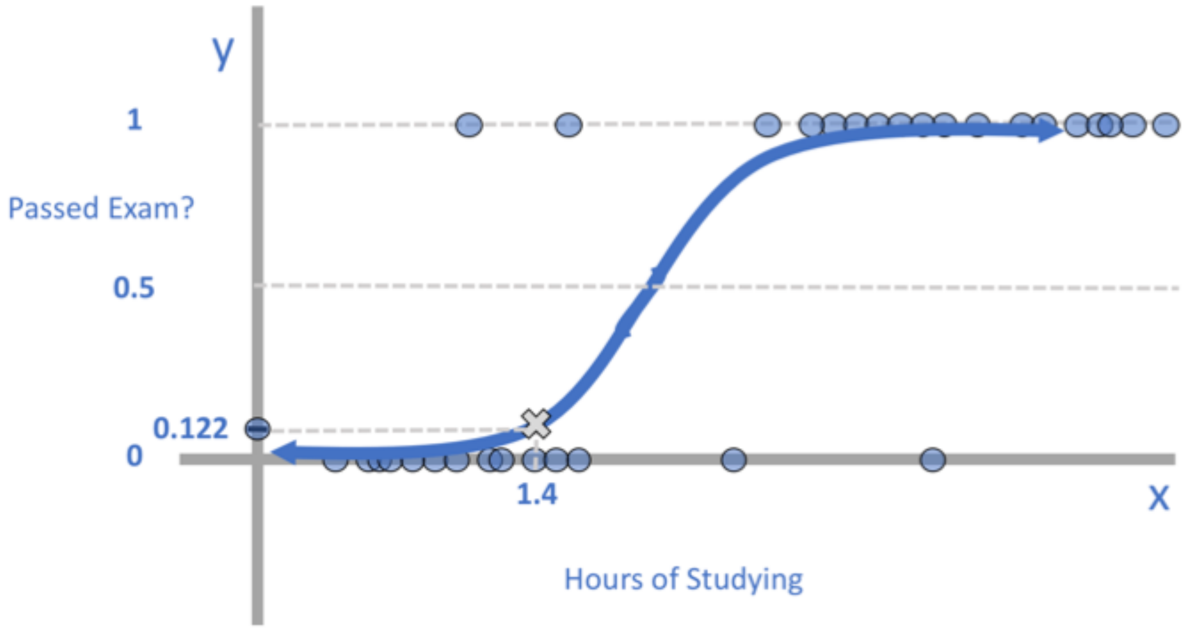
Logistic Regression Classification

- If a student only prepares 1.4 hours for the test
  - $study\_hours$=1.4
  - $\beta_0 = -4.077$
  - $\beta_1 = 1.5046$

$$p(pass) = \frac{(e^{\beta_0+\beta_1 * studyhrs})}{(e^{\beta_0+\beta_1 * studyhrs}) + 1}$$

$$p(pass) = \frac{(e^{-4.077+1.5046*1.4})}{(e^{-4.077+1.5046*1.4}) + 1} = 0.122$$

- Our model predicts a probability of passing of 12.2%
- This falls below the 0.5 threshold and results in a prediction of failing the exam



HPSI
Human Performance Systems, Inc.

Logistic Regression Example

HPSI
Human Performance Systems, Inc.

References

1. *Locally Weighted Learning* by Atkeson, Moore, Schaal
2. *Tuning Locally Weighted Learning* by Schaal, Atkeson, Moore
3. Kruschke, J. K. (2014). Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.
4. Machine Learning, Neural and Statistical Classification, Editors: D. Michie, D.J. Spiegelhalter, C.C. Taylor
5. http://www.dataschool.io/guide-to-logistic-regression/