DATA MINING 2
Logistic Regression

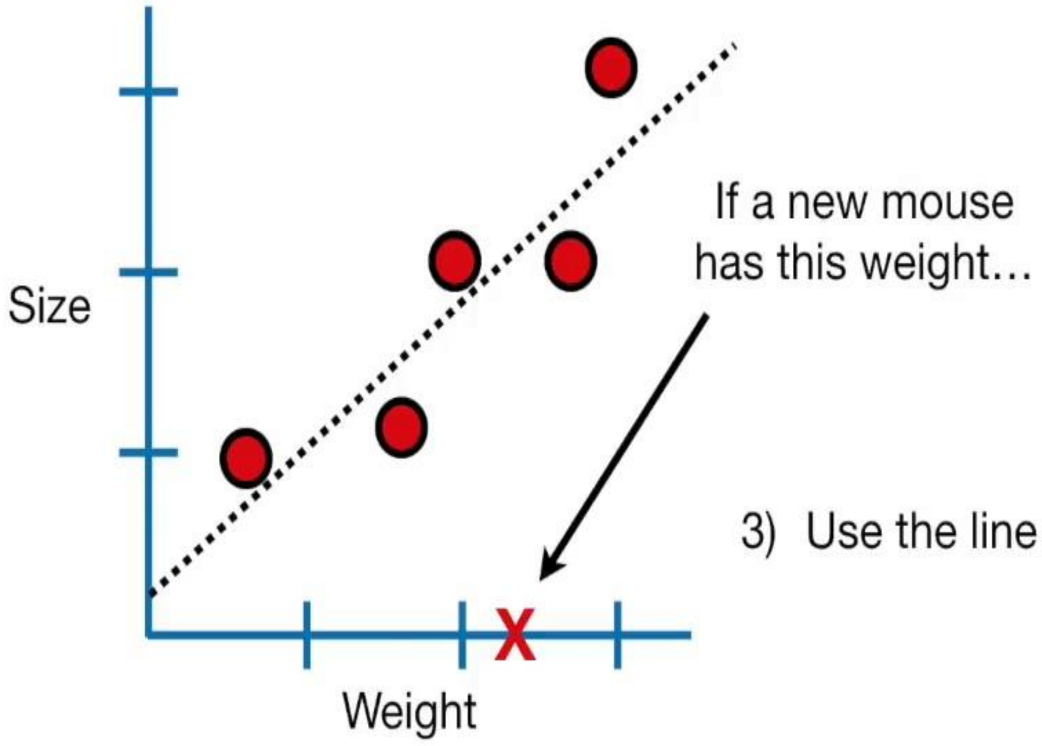
Riccardo Guidotti

a.a. 2023/2024

Contains edited slides from StatQuest

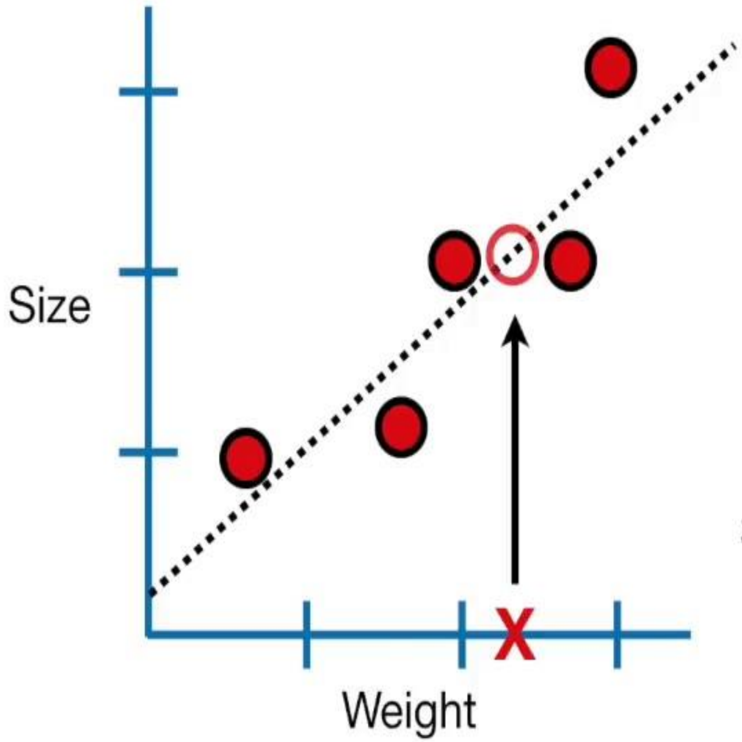
UNIVERSITÀ DI PISA

Recalling Linear Regression



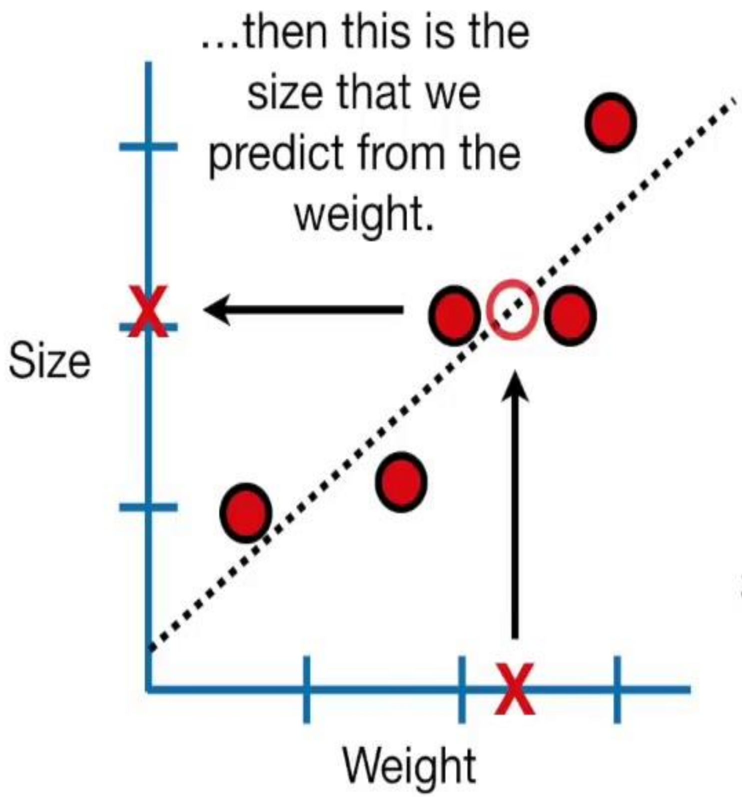
3) Use the line to predict ***size*** given ***weight***.

Recalling Linear Regression



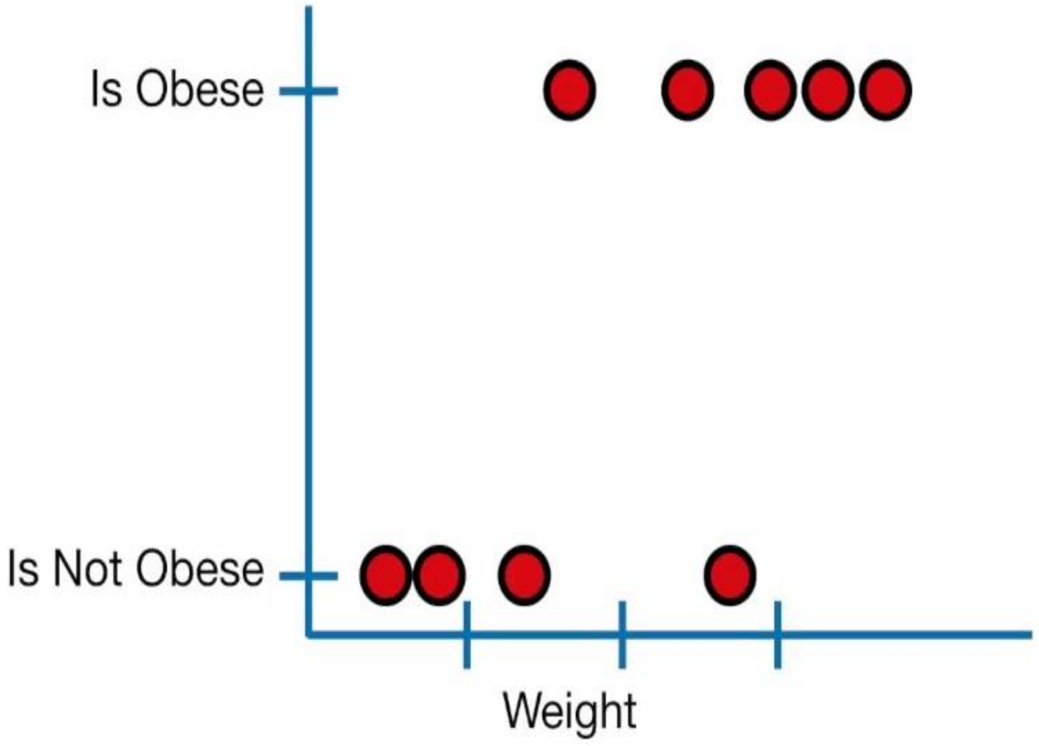
3) Use the line to predict ***size*** given ***weight***.

Recalling Linear Regression



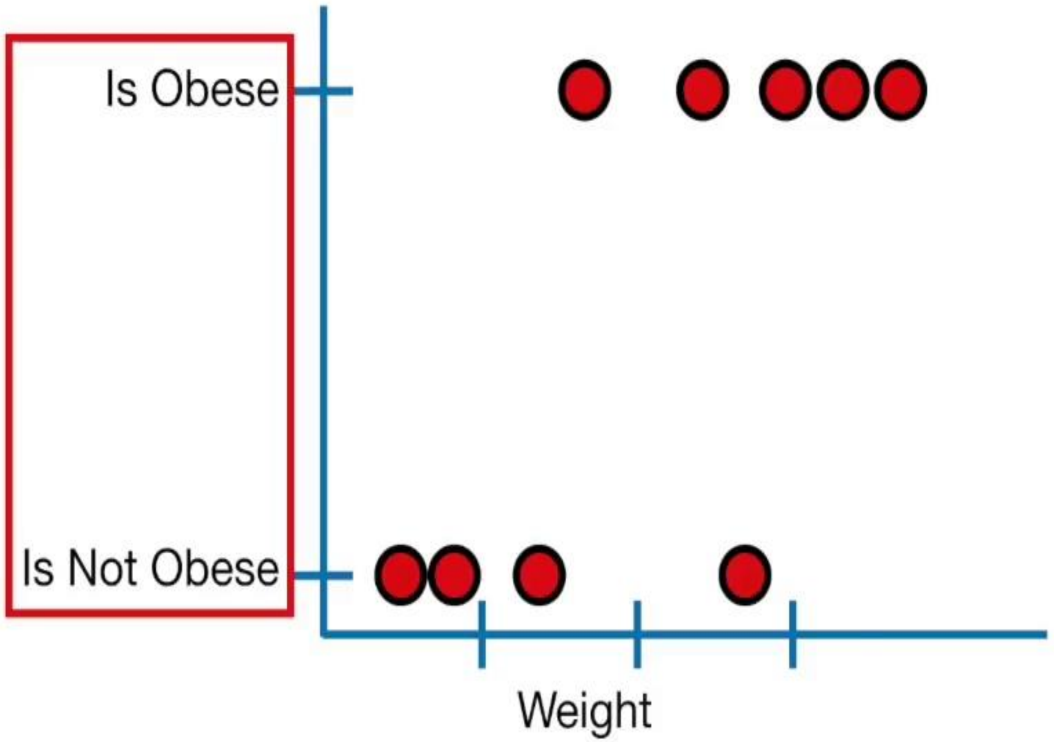
3) Use the line to predict ***size*** given ***weight***.

Logistic Regression

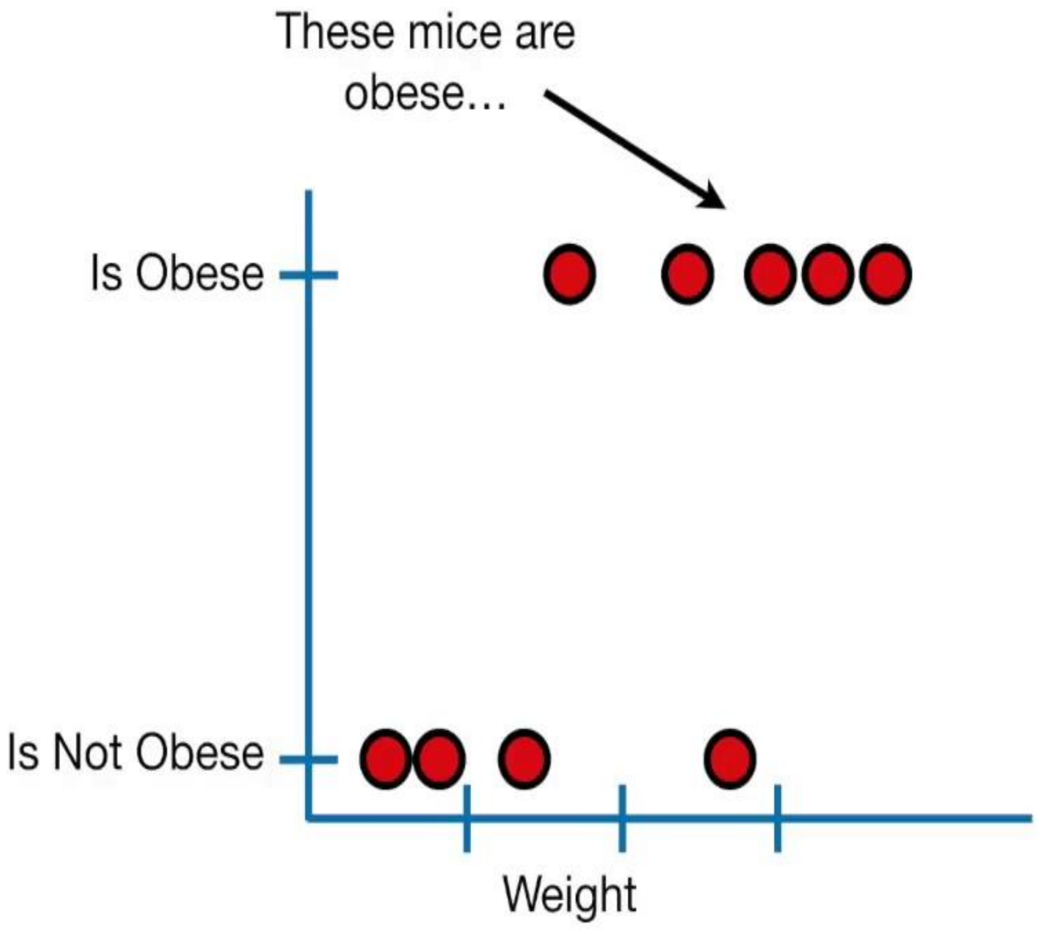


Logistic Regression

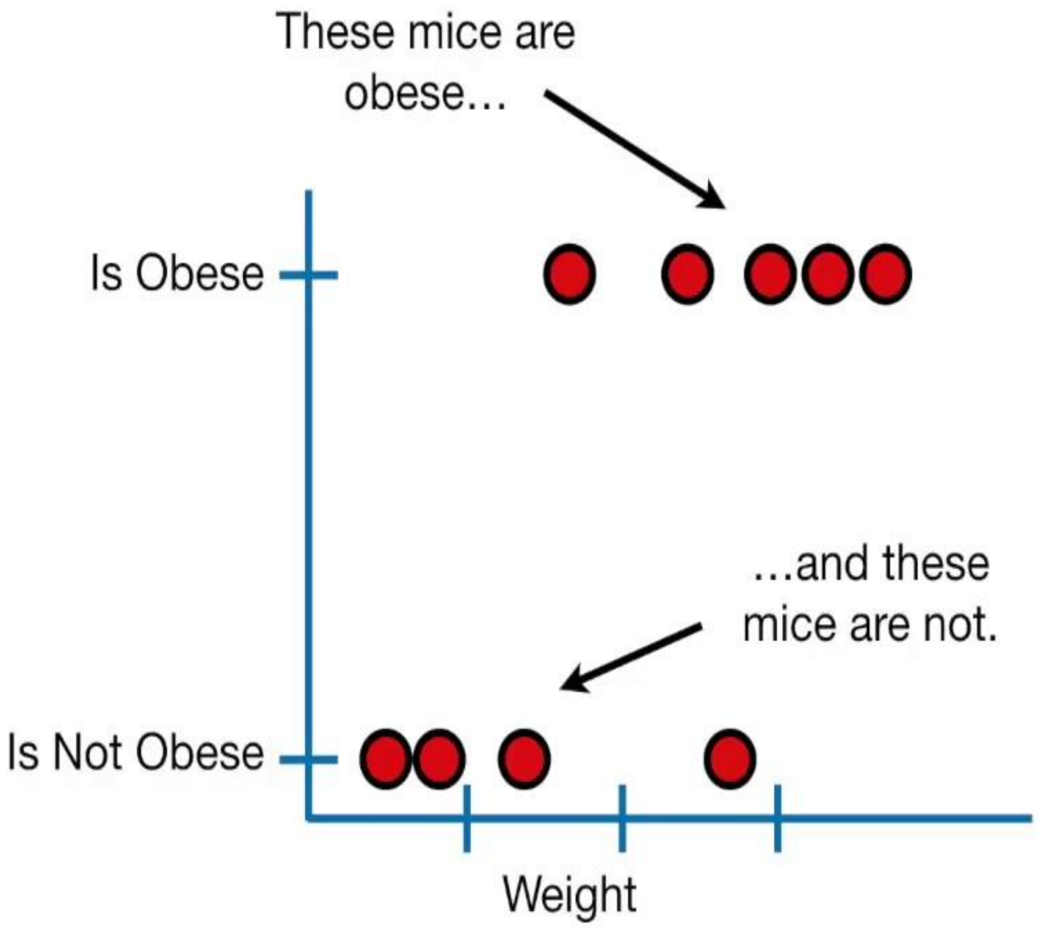
Logistic regression predicts whether something is ***True*** or ***False***, instead of predicting something continuous like ***size***.



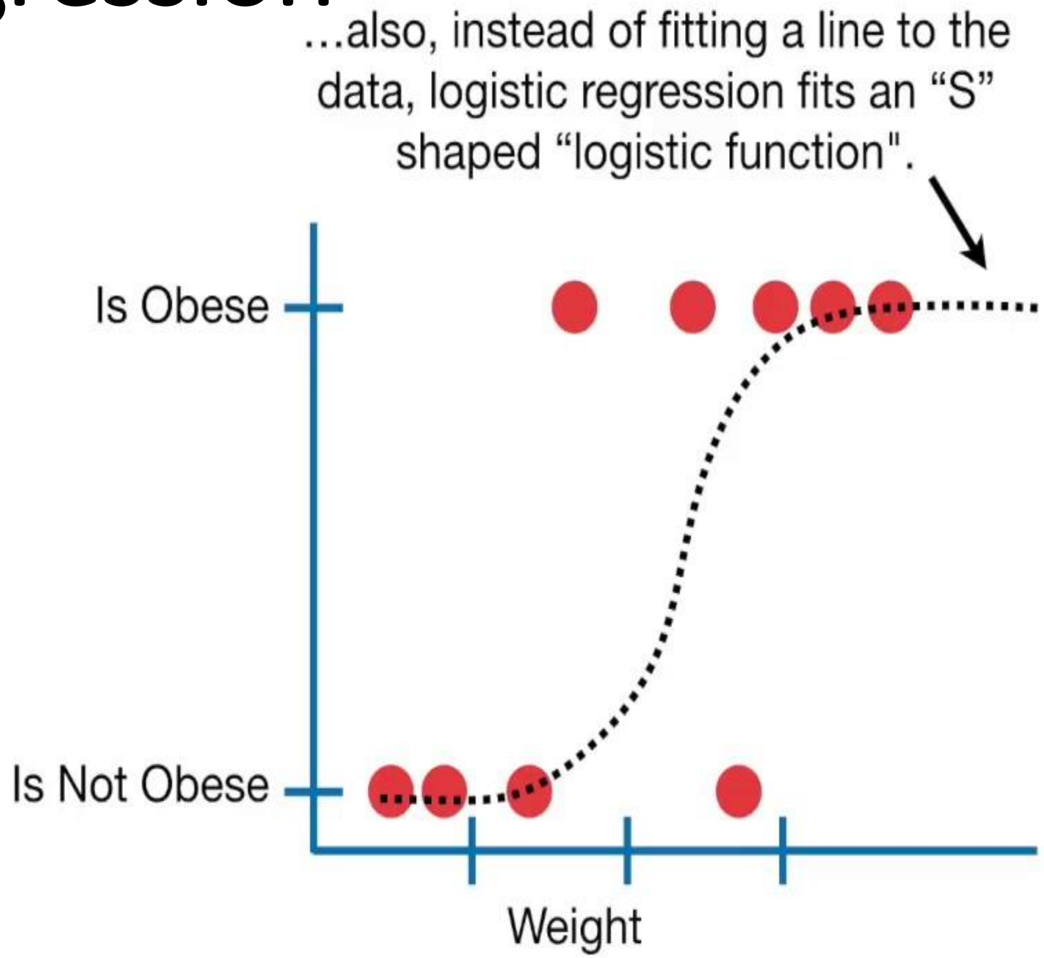
Logistic Regression



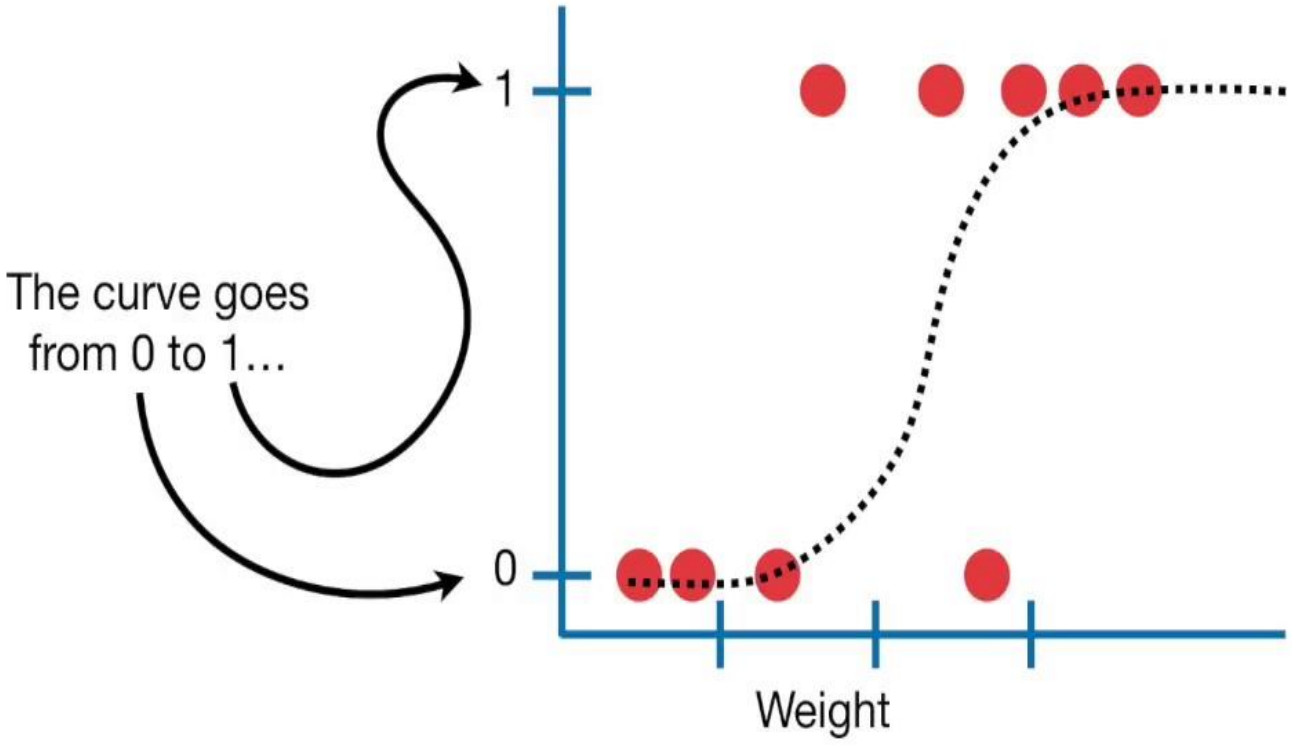
Logistic Regression



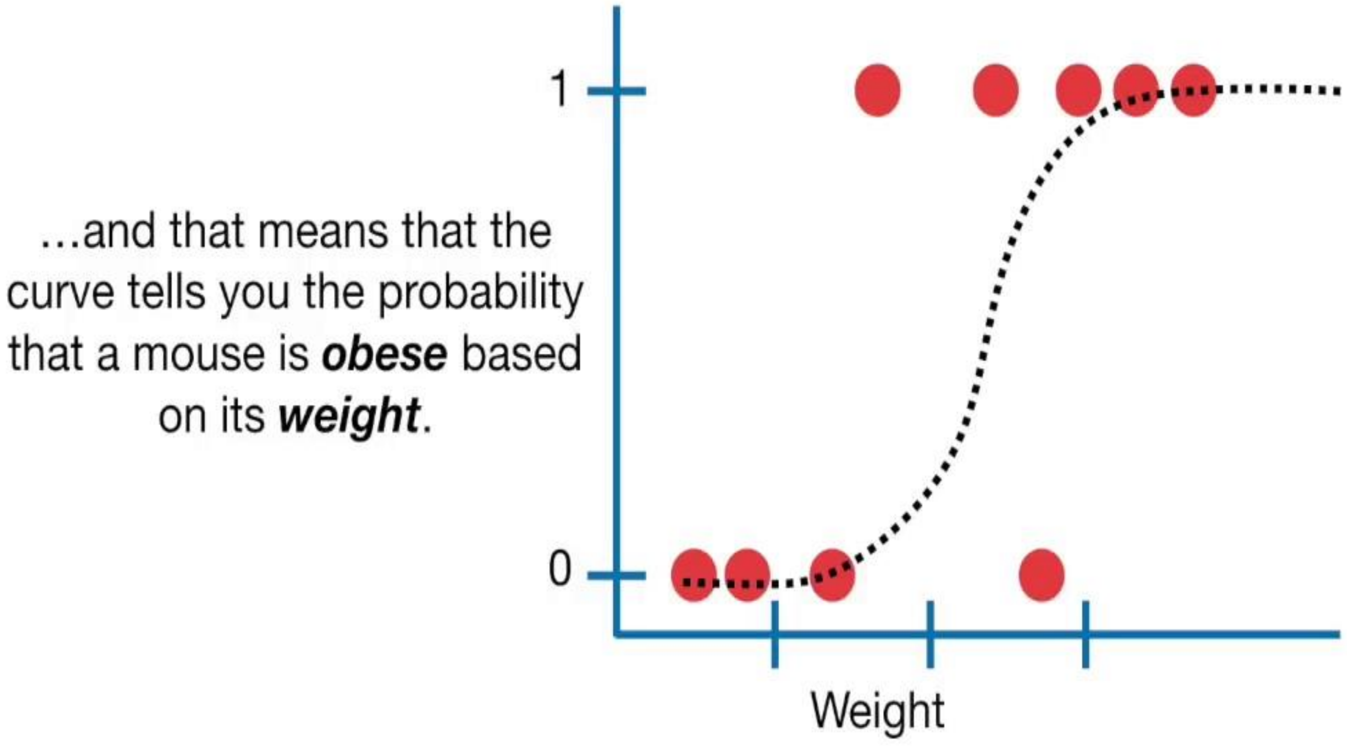
Logistic Regression



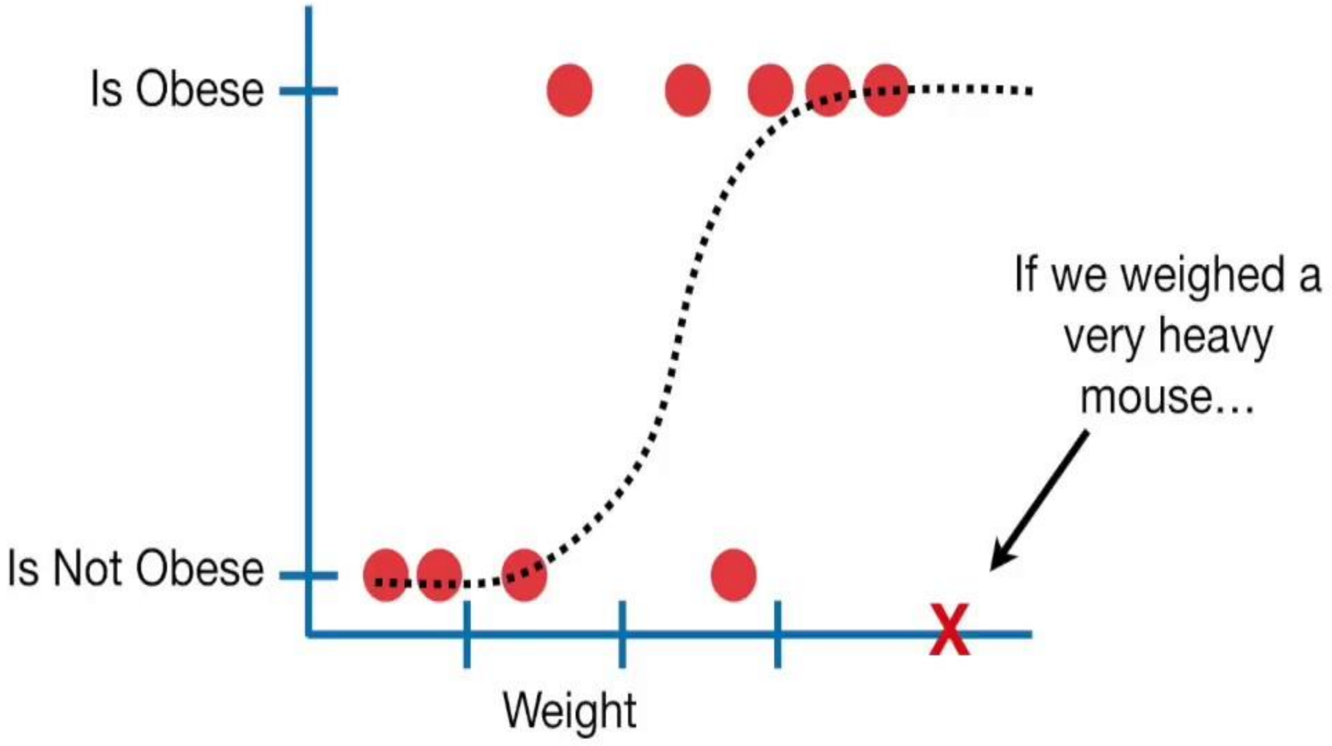
Logistic Regression



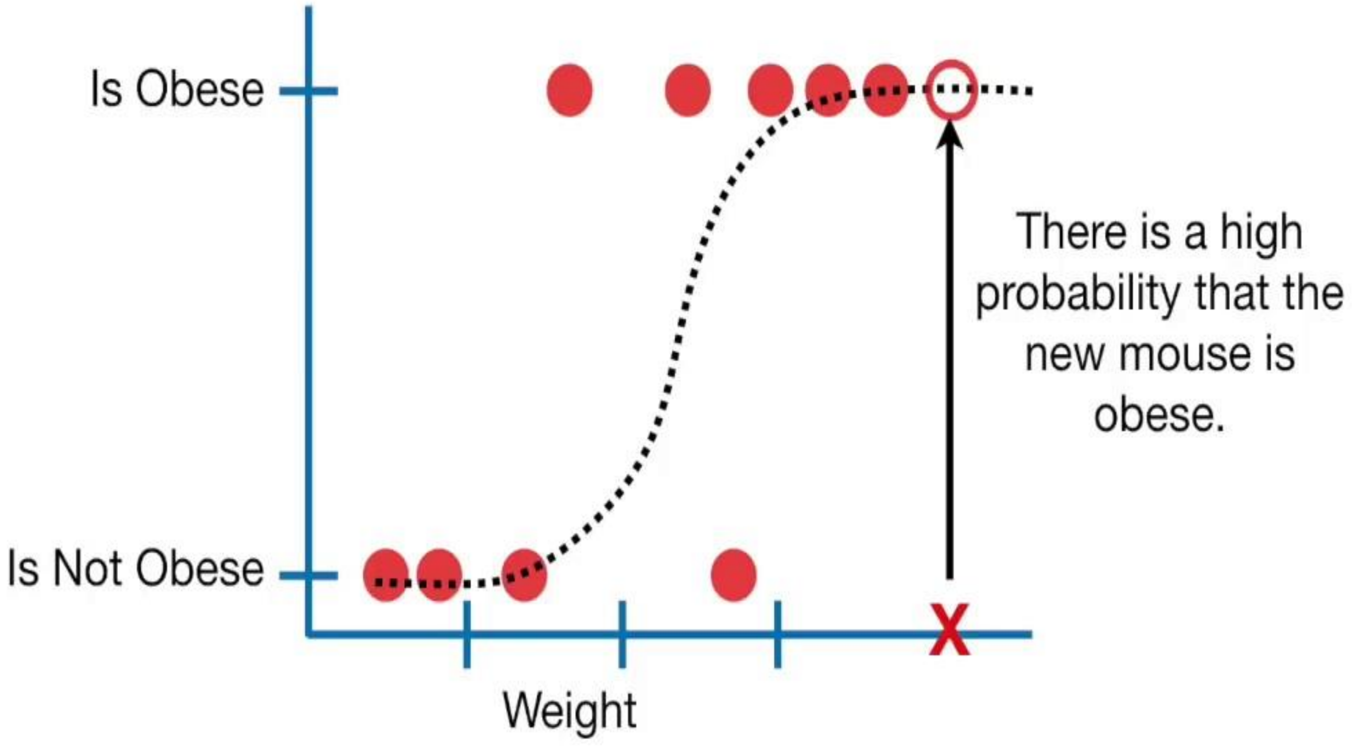
Logistic Regression



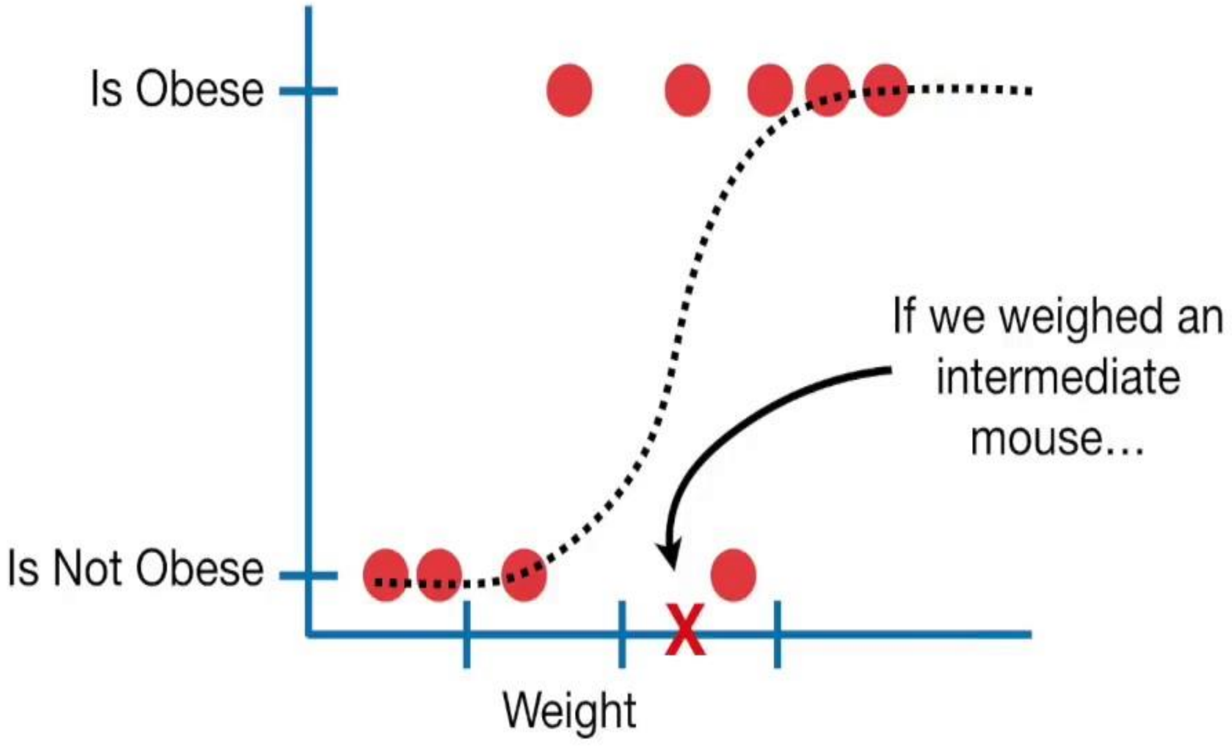
Logistic Regression



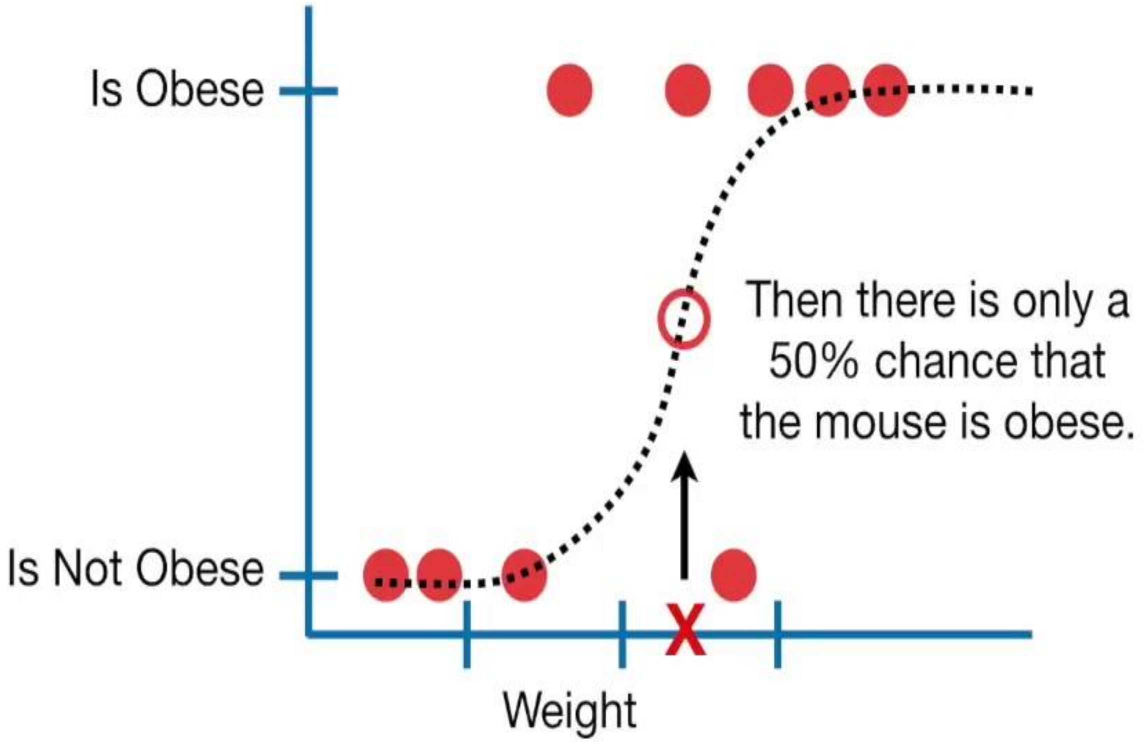
Logistic Regression



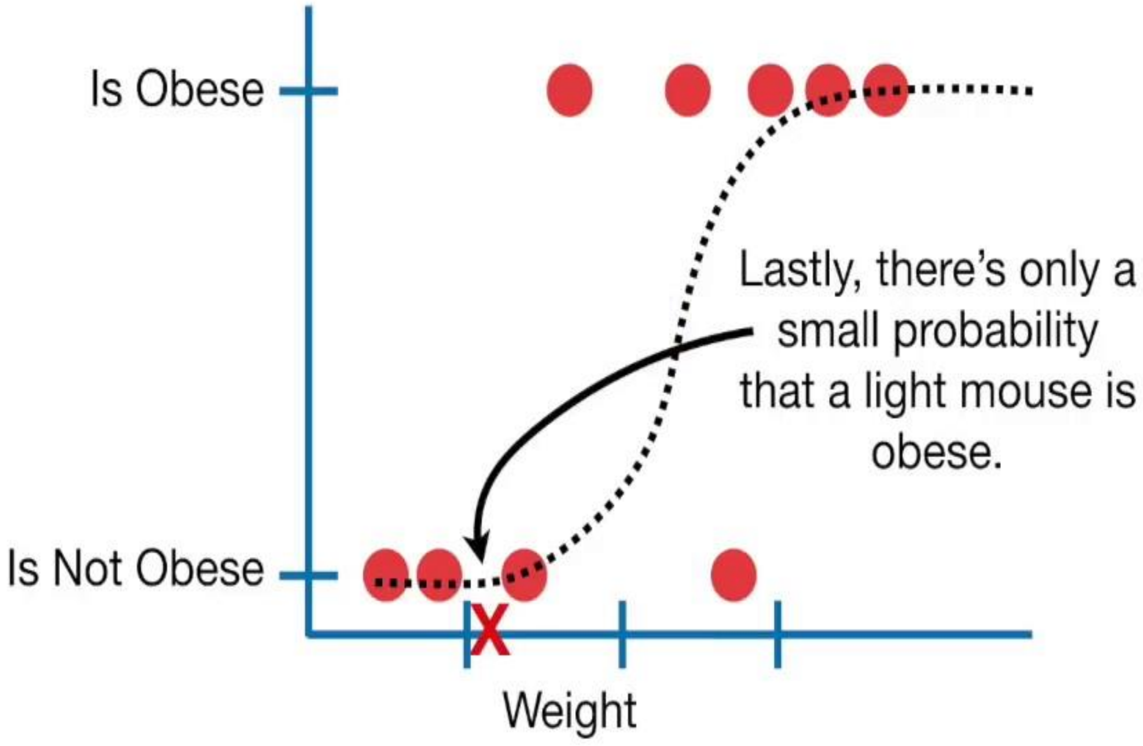
Logistic Regression



Logistic Regression

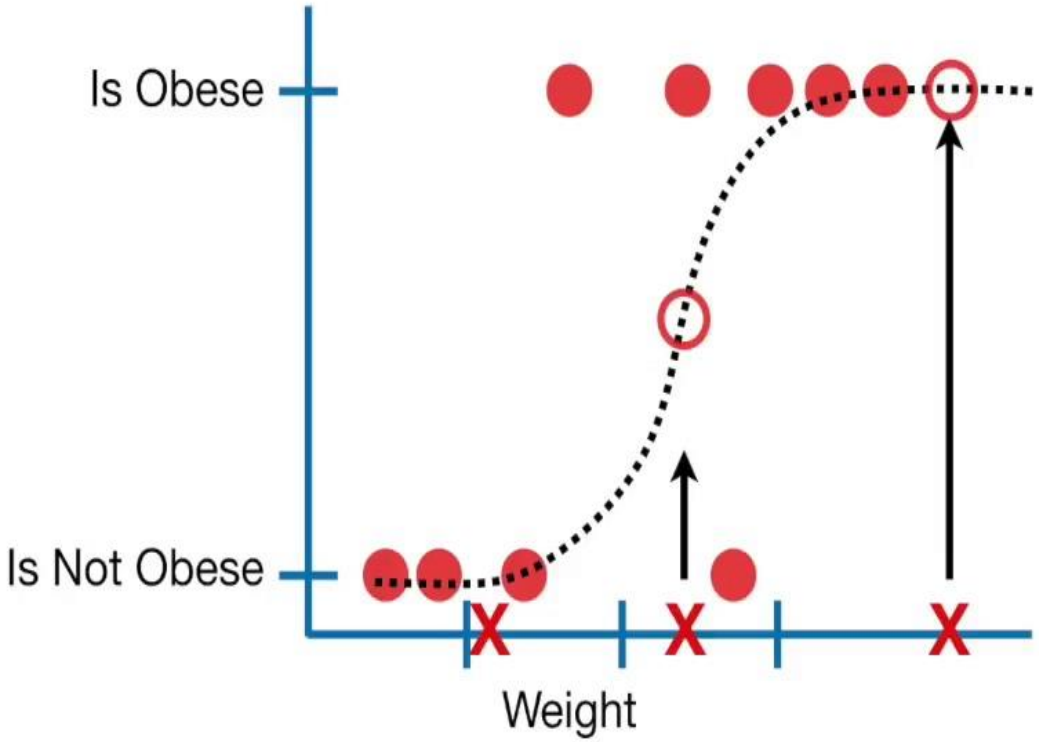


Logistic Regression

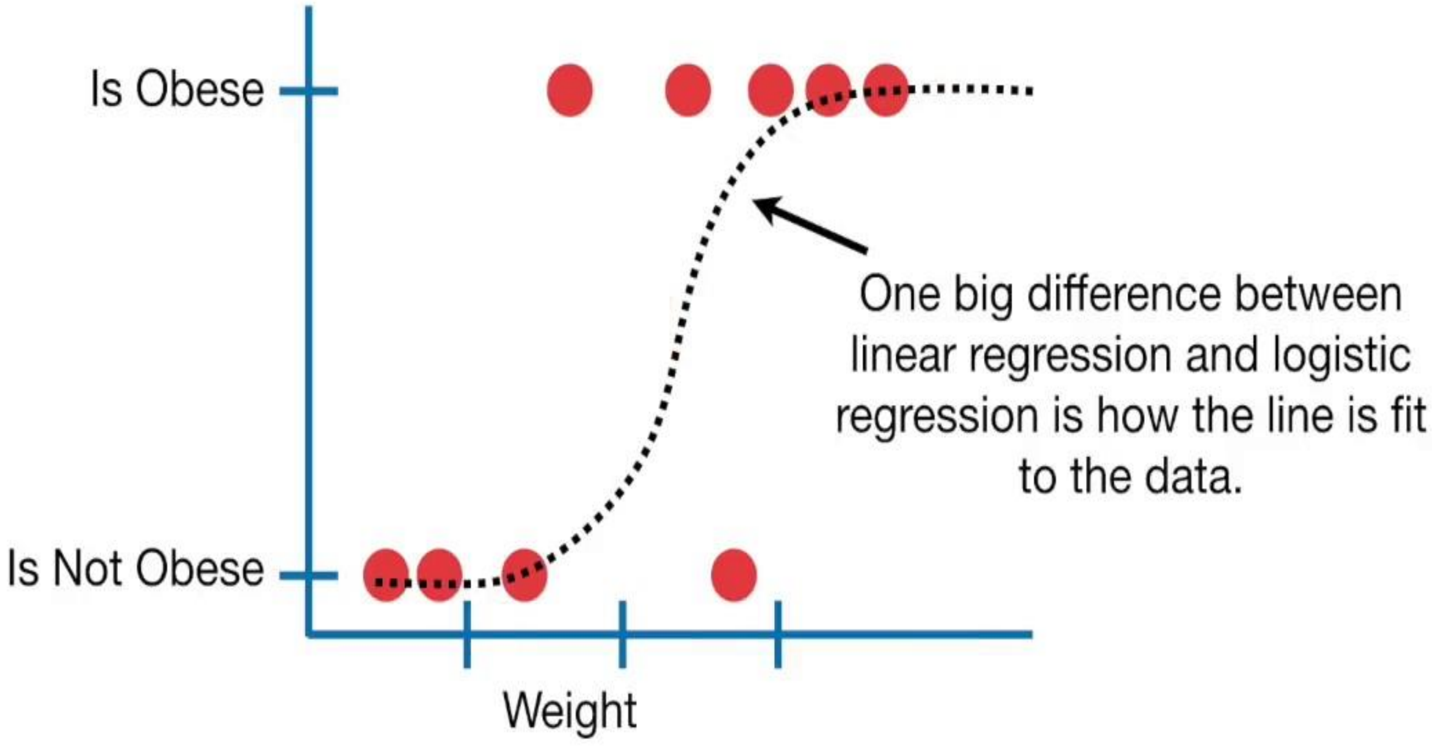


Logistic Regression

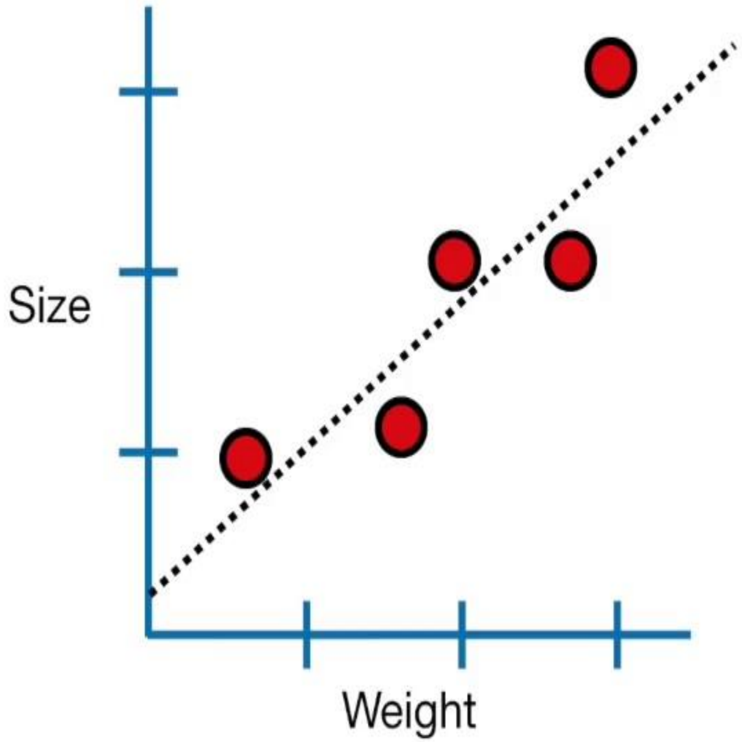
Although logistic regression tells the probability that a mouse is obese or not, it's usually used for classification.



Linear vs Logistic

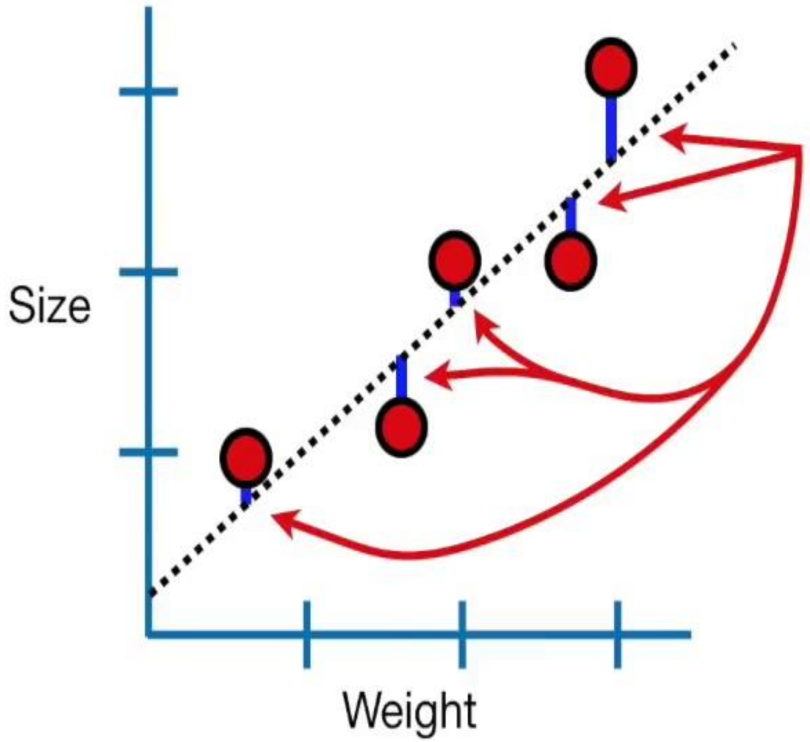


Linear vs Logistic



With linear regression, we fit the line using "least squares".

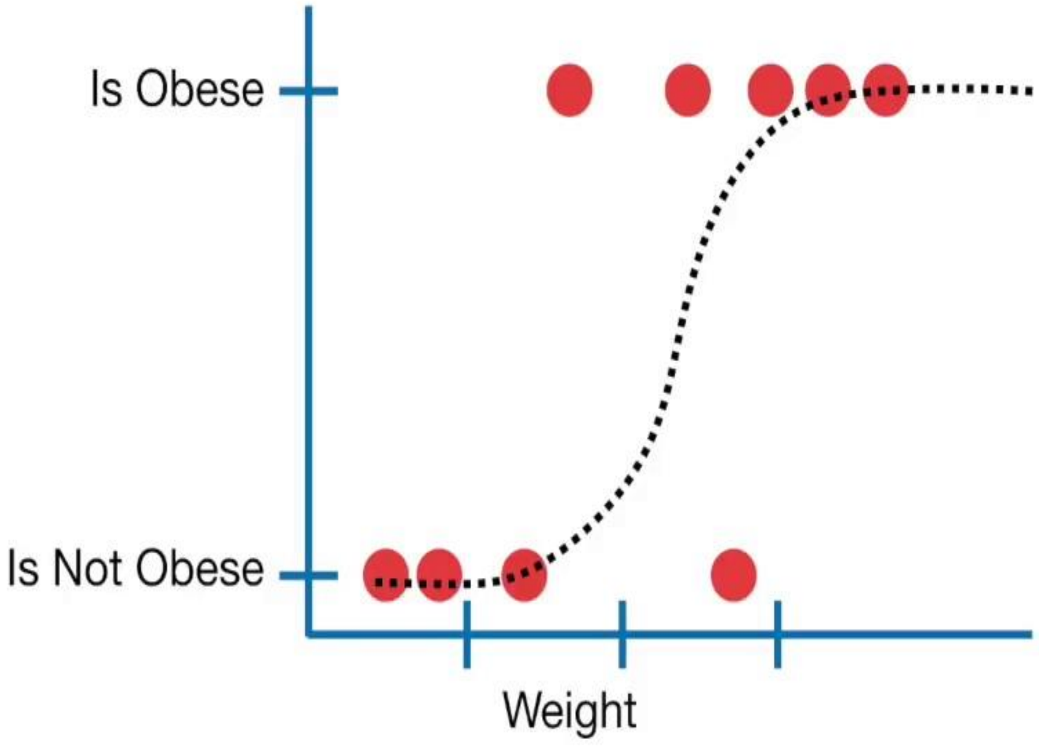
Linear vs Logistic



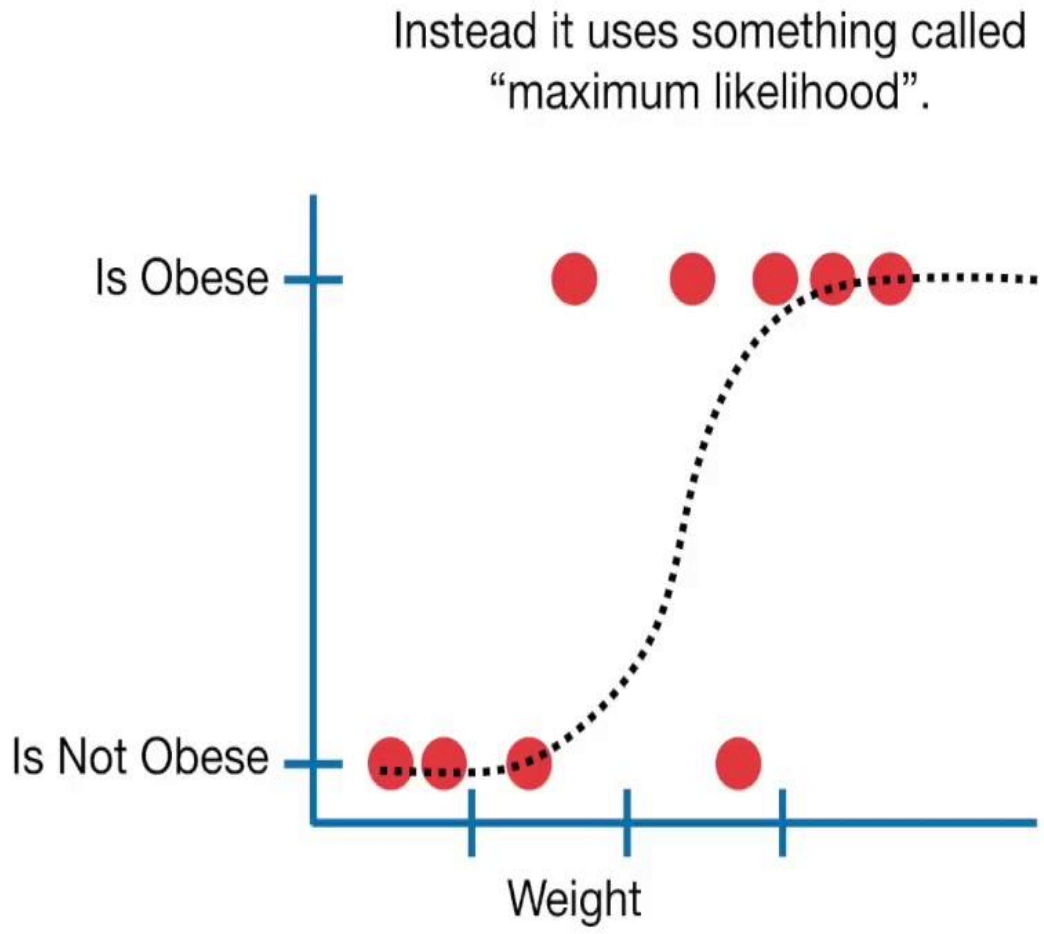
In other words, we find the line that minimizes the sum of the squares of these residuals.

Linear vs Logistic

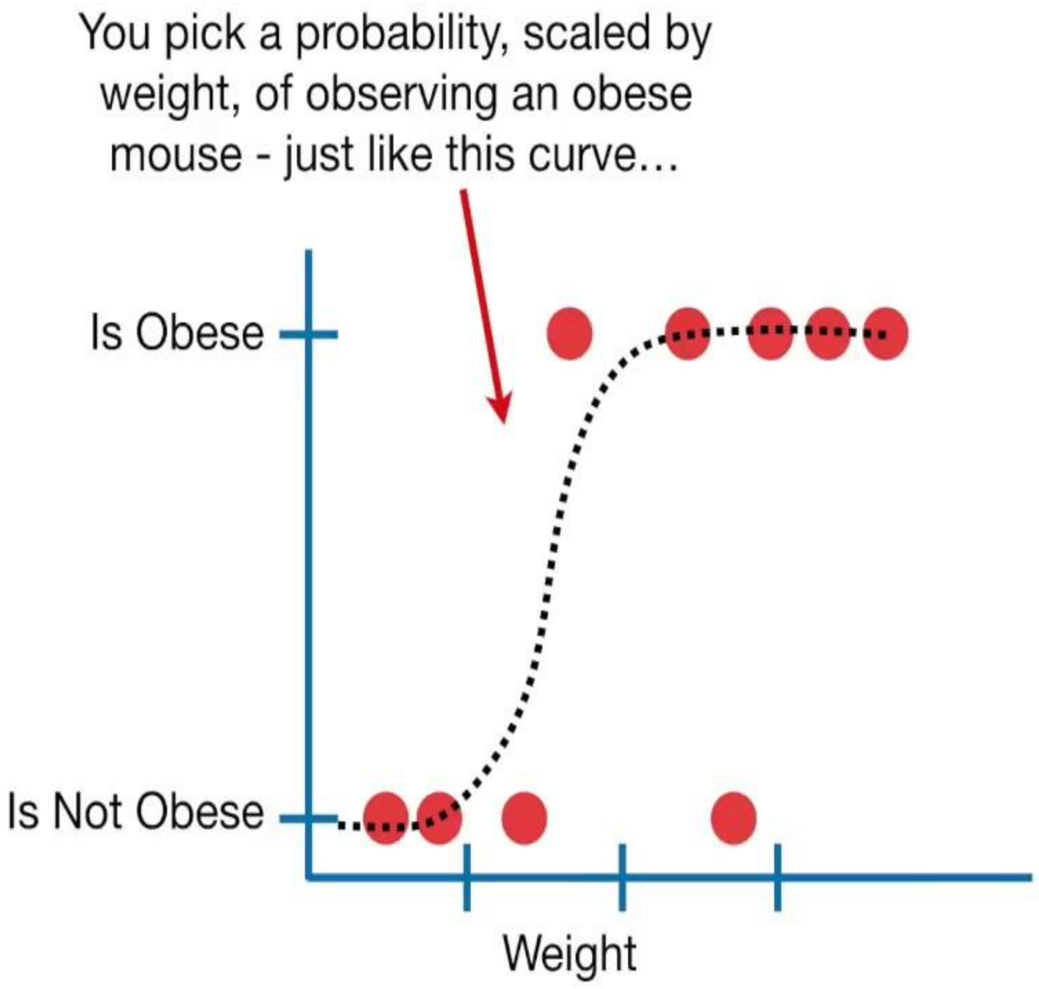
Logistic regression doesn't have the same concept of a "residual", so it can't use least squares



Linear vs Logistic

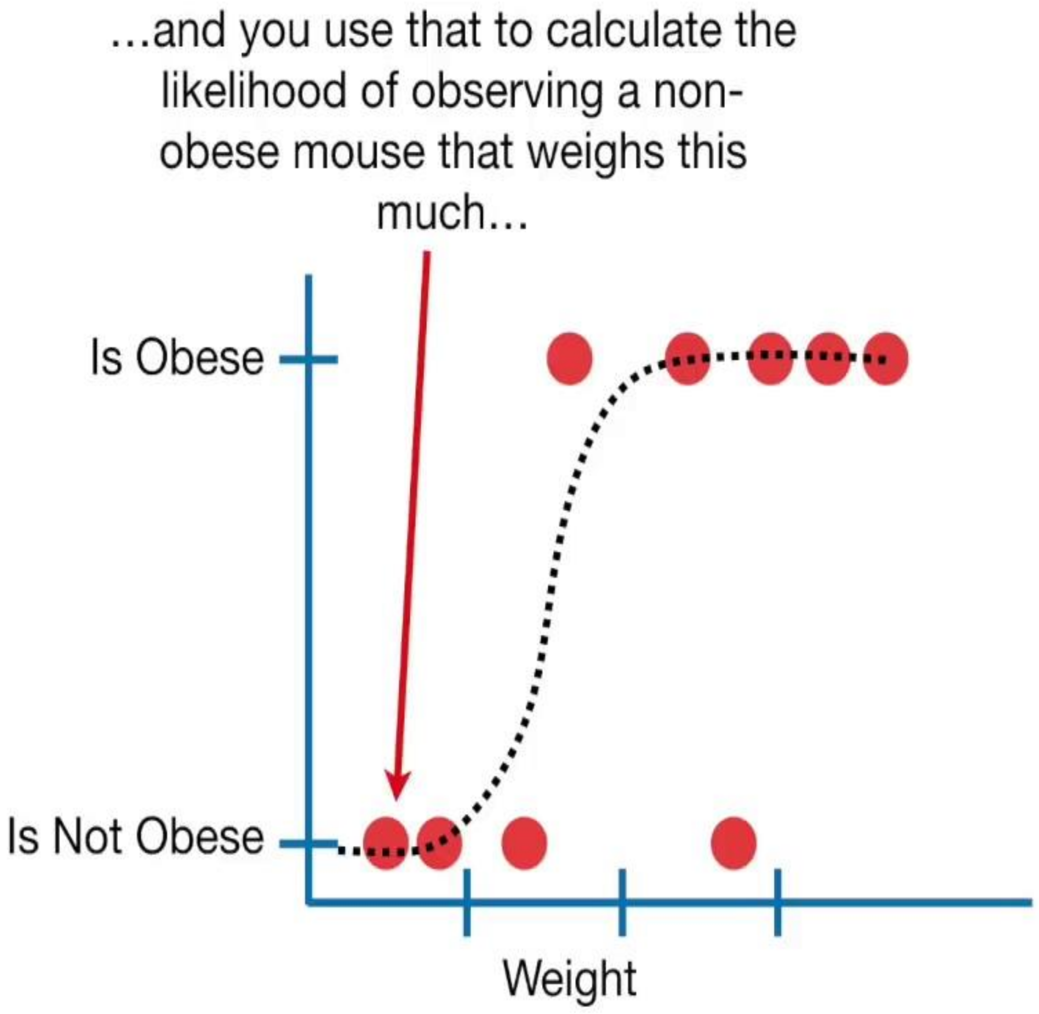


MLE



You pick a probability, scaled by weight, of observing an obese mouse - just like this curve...
Is Obese
Is Not Obese
Weight

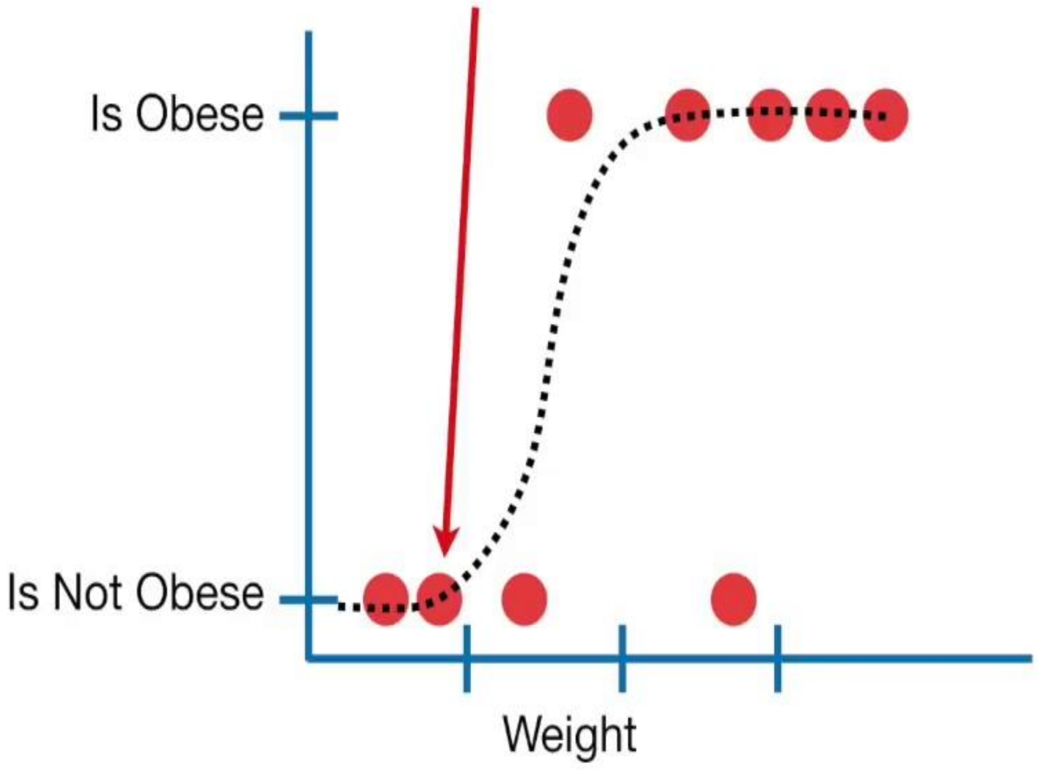
MLE



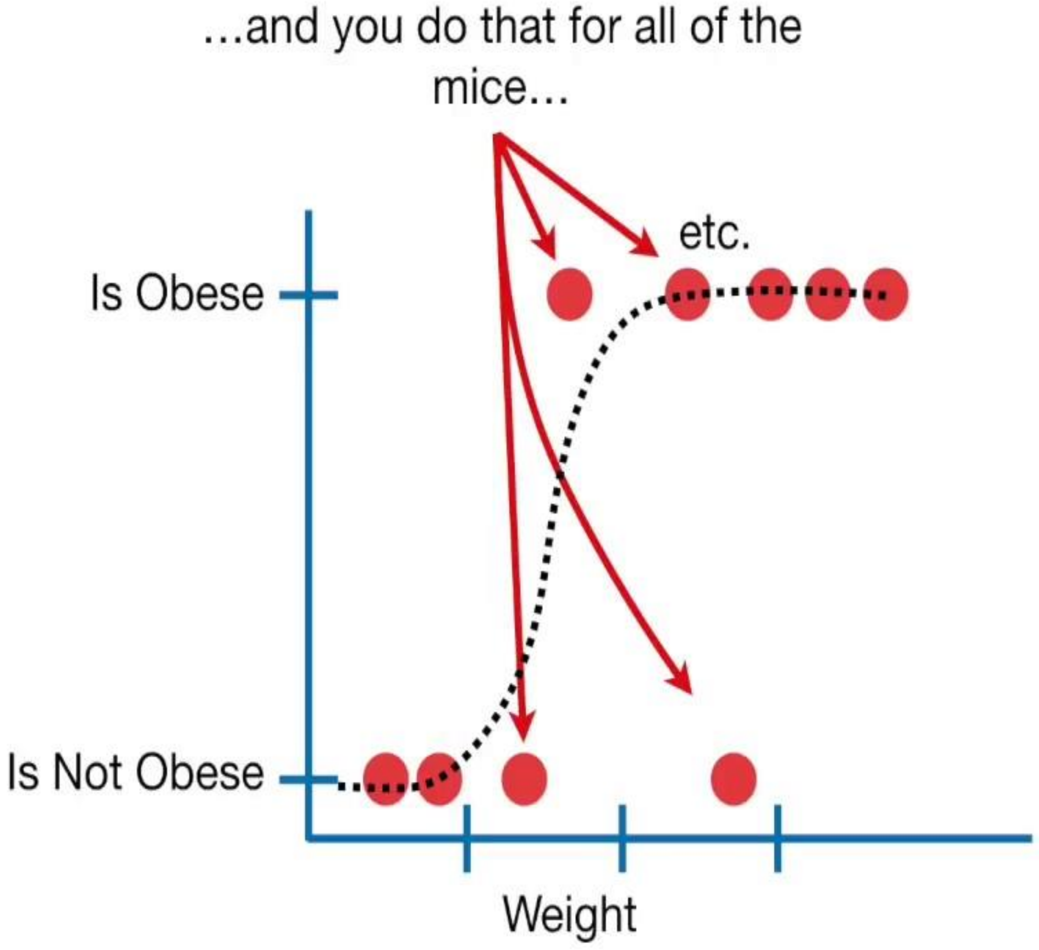
...and you use that to calculate the likelihood of observing a non-obese mouse that weighs this much...
Is Obese
Is Not Obese
Weight

MLE

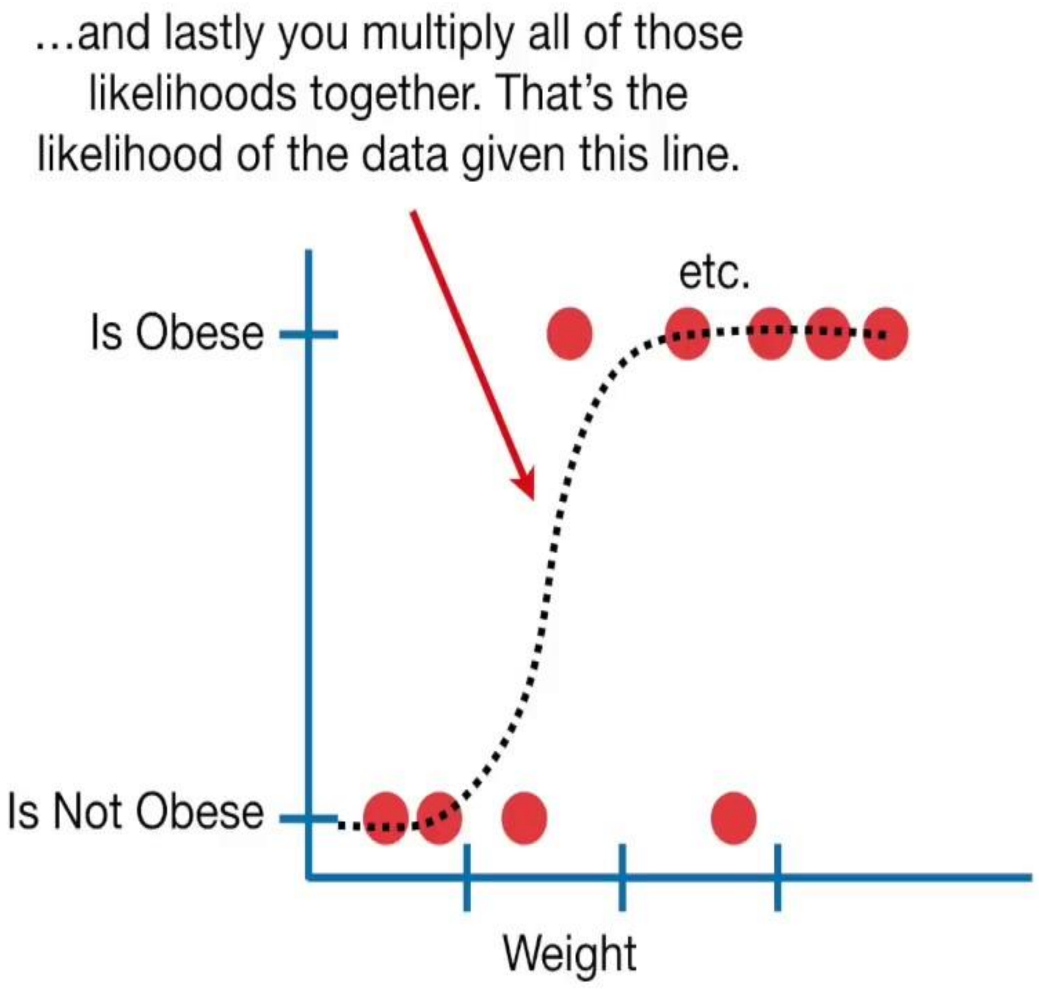
...and then you calculate the likelihood of observing this mouse...



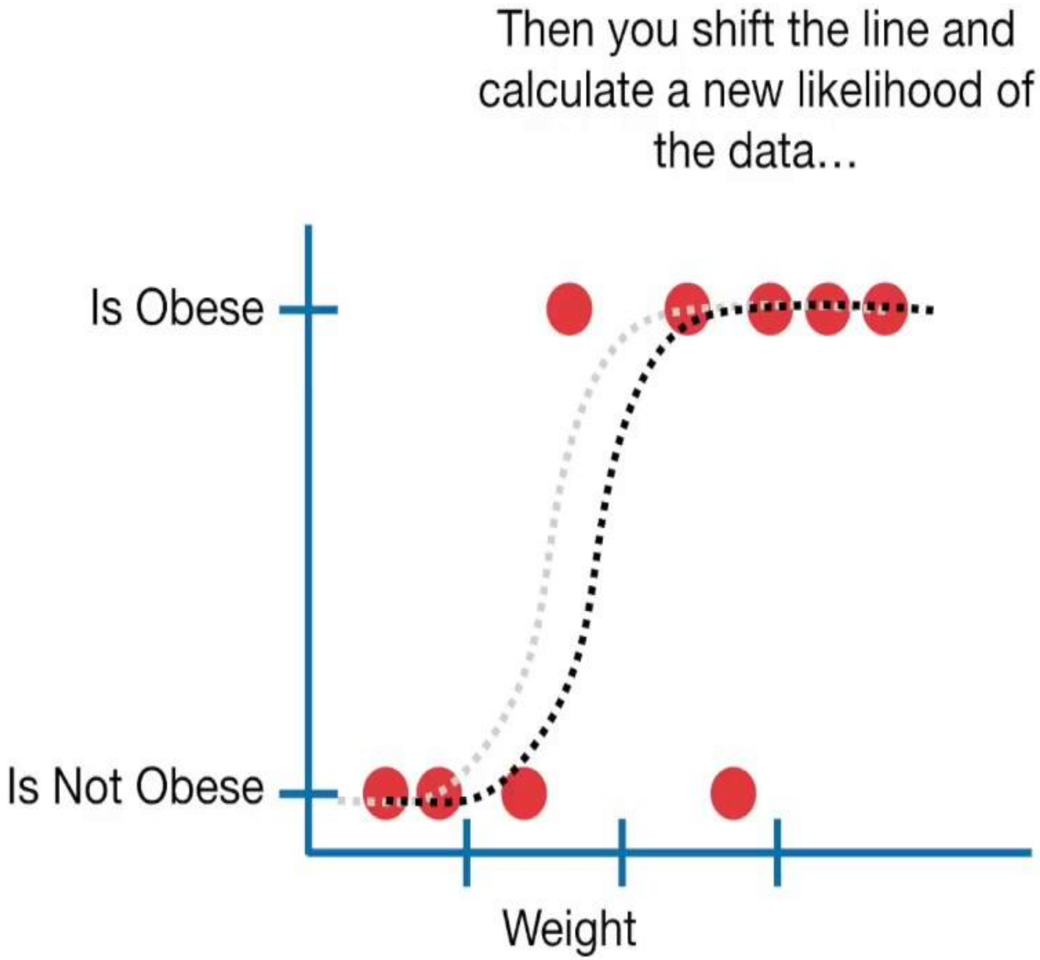
MLE


...and you do that for all of the mice...
etc.
Is Obese
Is Not Obese
Weight

MLE


...and lastly you multiply all of those likelihoods together. That's the likelihood of the data given this line.
etc.
Is Obese
Is Not Obese
Weight

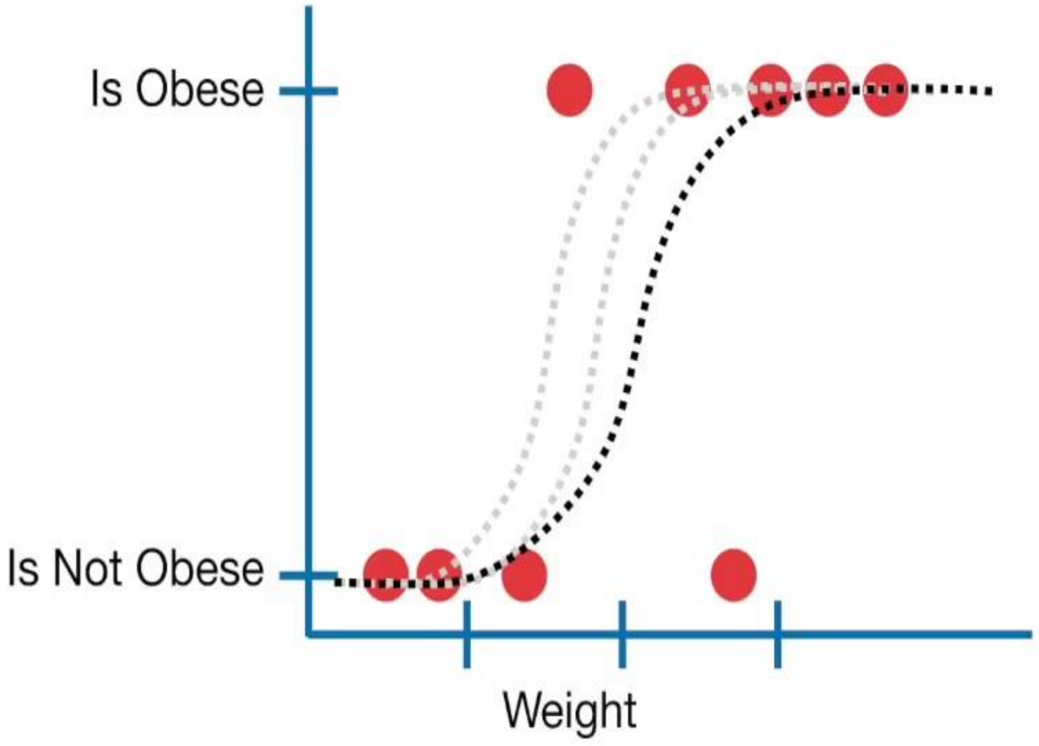
MLE



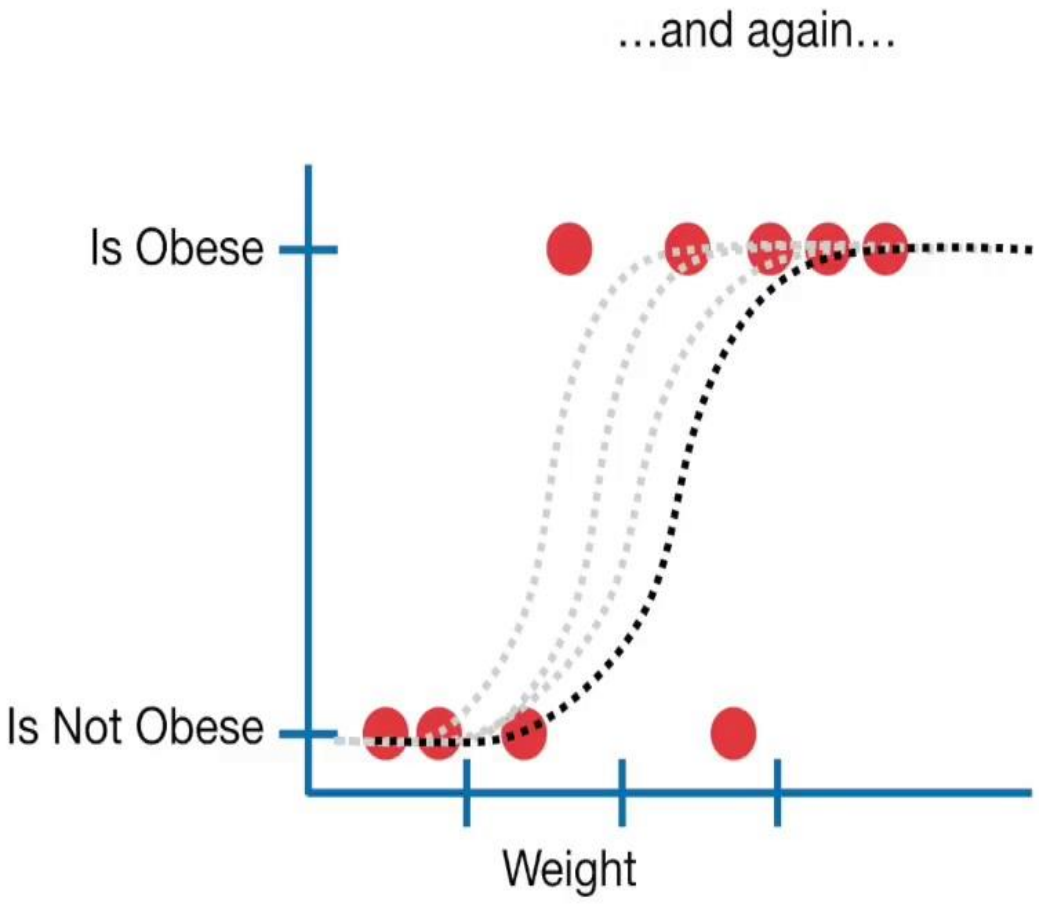
Then you shift the line and calculate a new likelihood of the data...
Is Obese
Is Not Obese
Weight

MLE

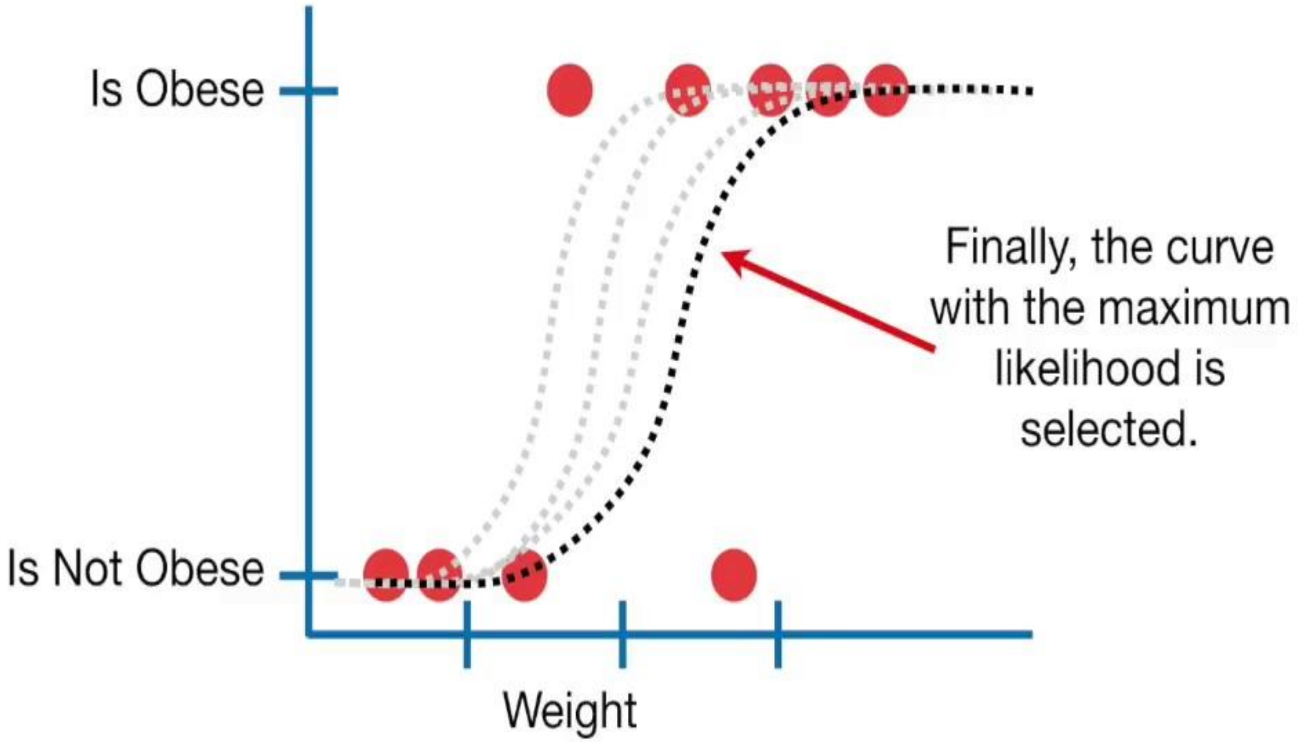
…then shift the line and
calculate the likelihood again…



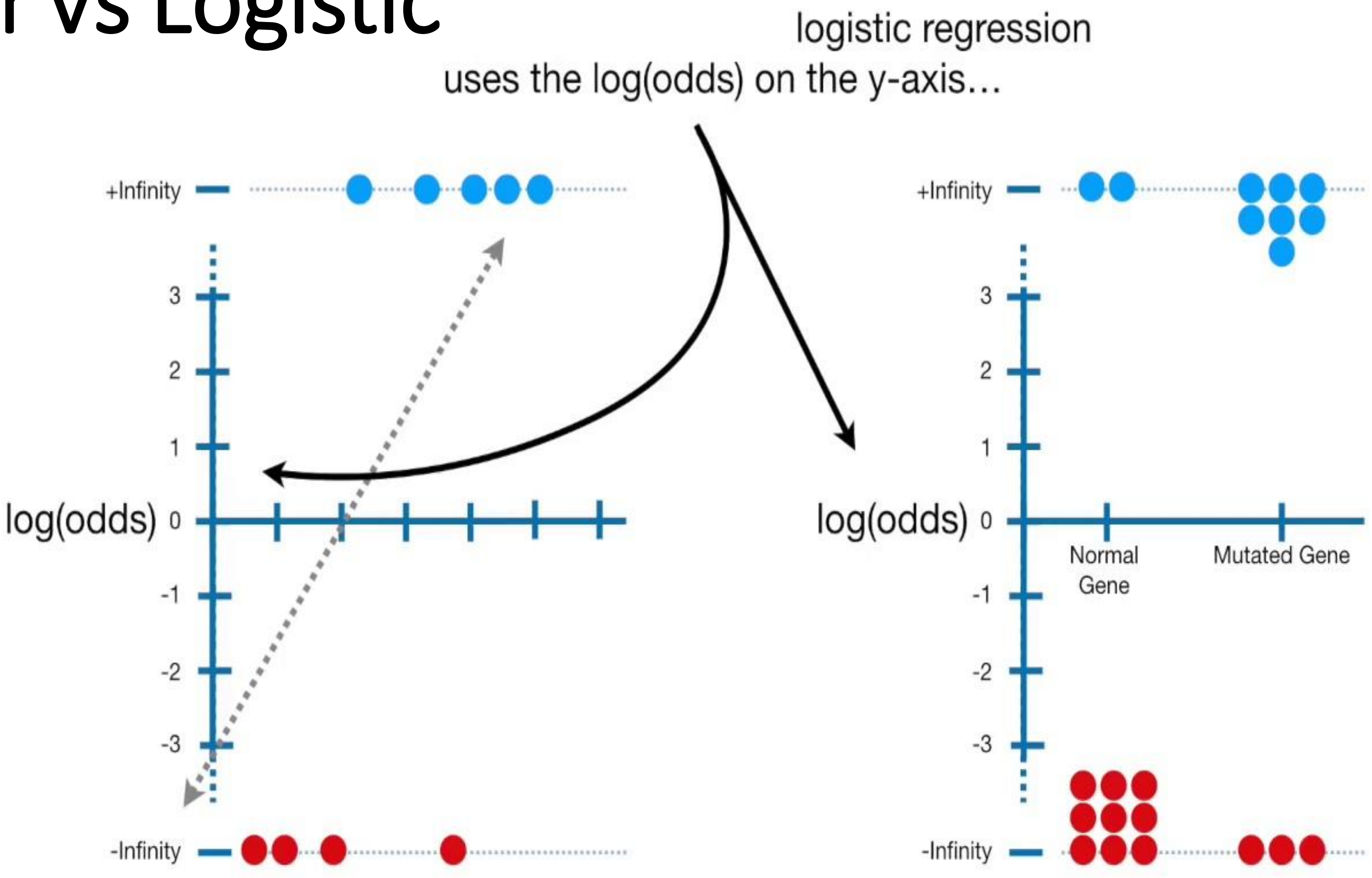
MLE



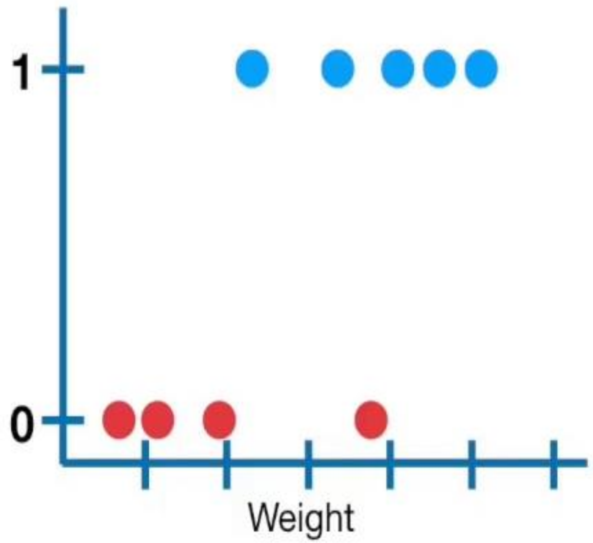
MLE


Is Obese
Is Not Obese
Weight
Finally, the curve with the maximum likelihood is selected.

Linear vs Logistic

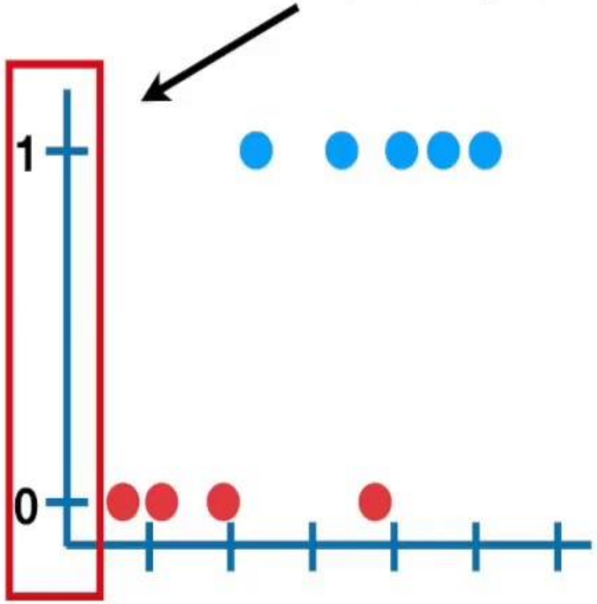


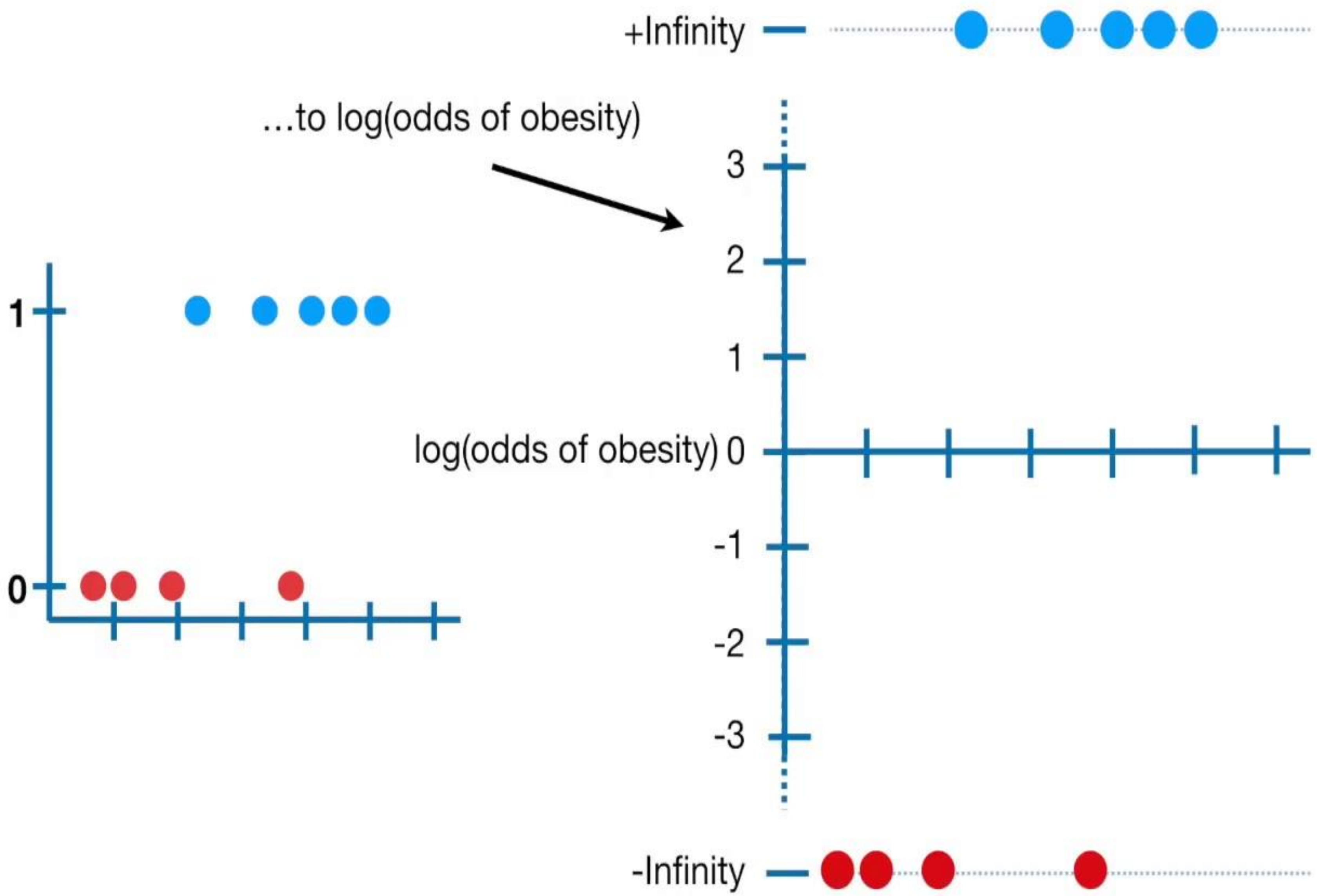
Fit a Line with Logistic Regression

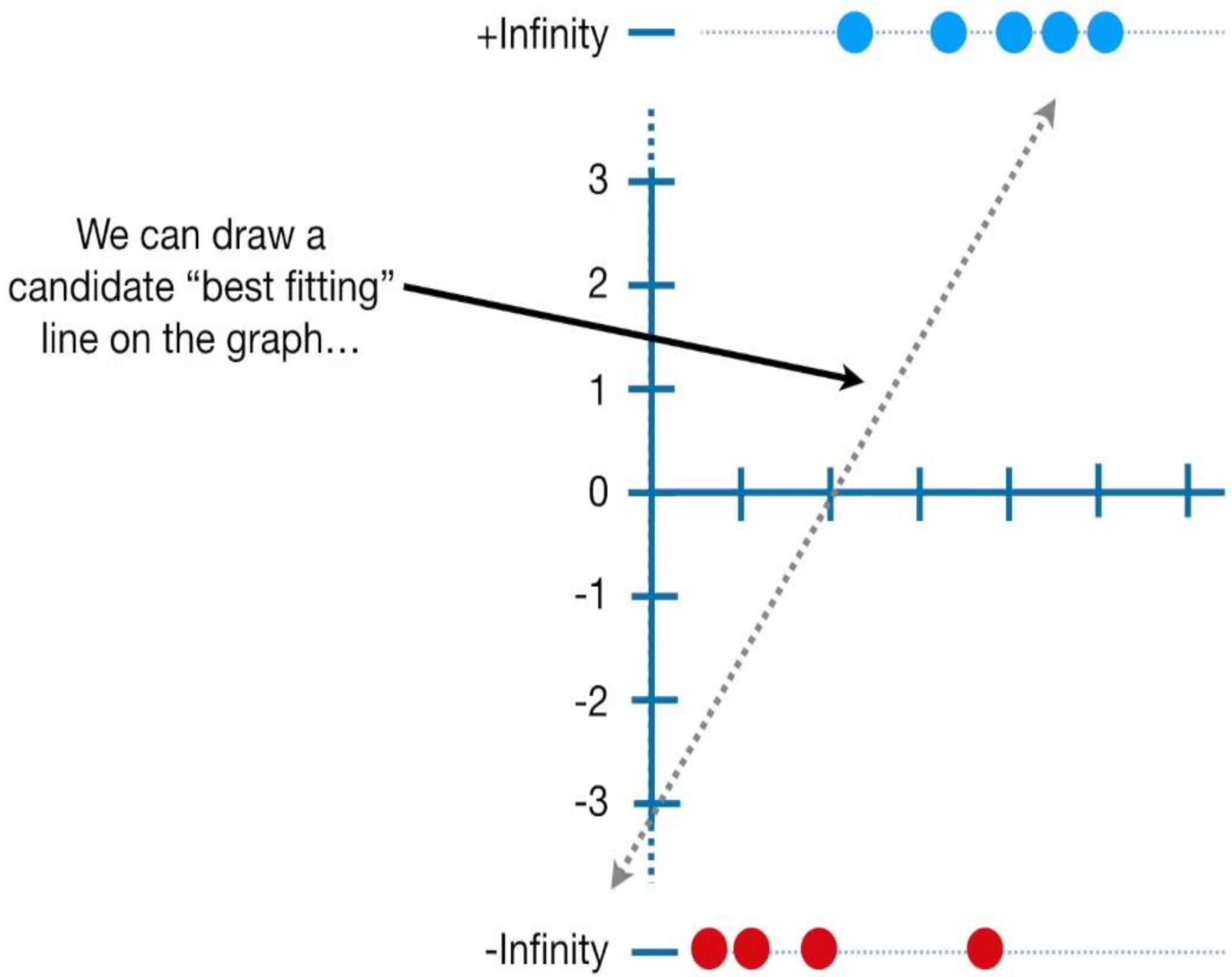


Fit a Line with Logistic Regression

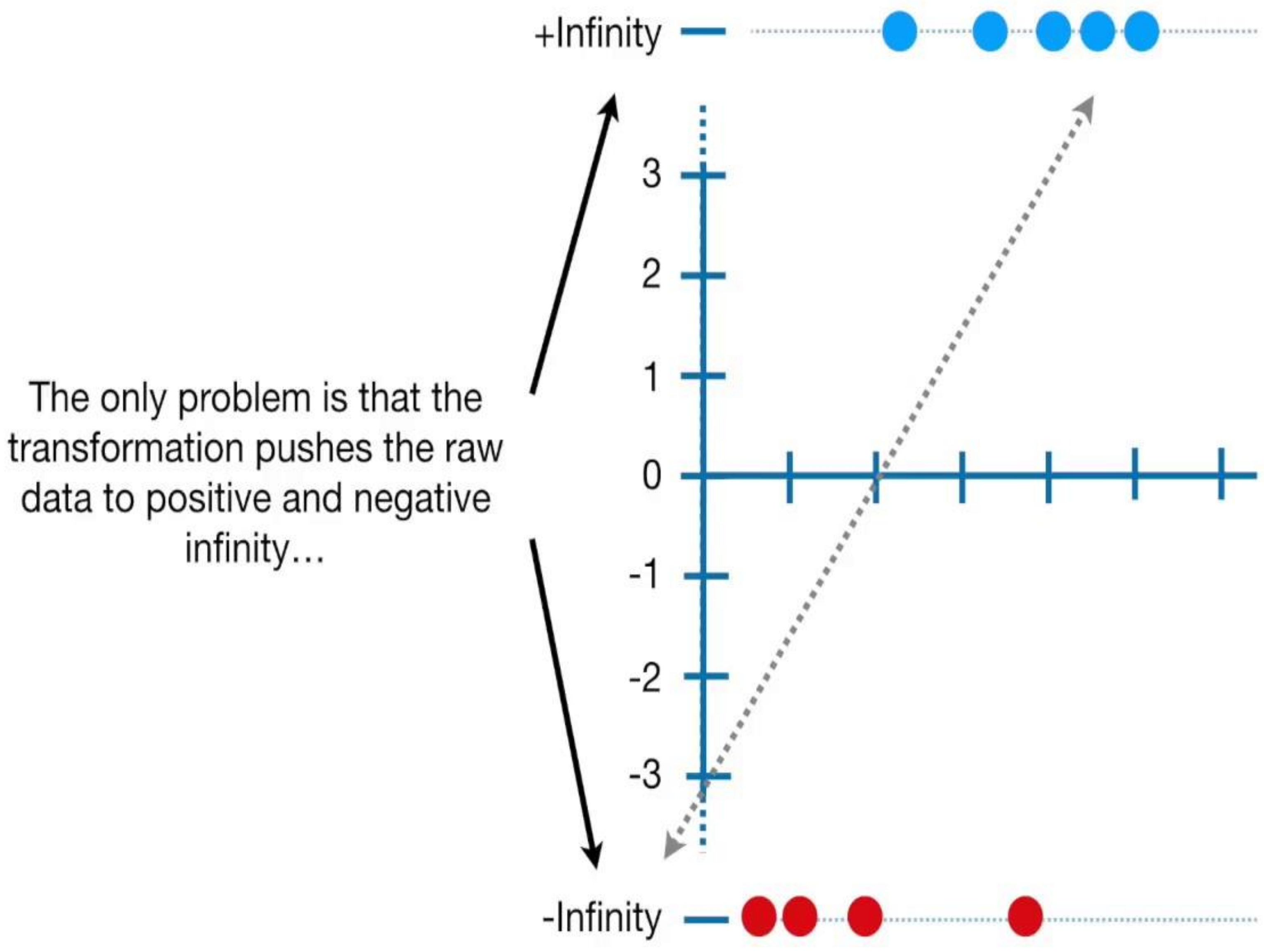
As we know, in logistic regression, we transform the y-axis from the probability of obesity…



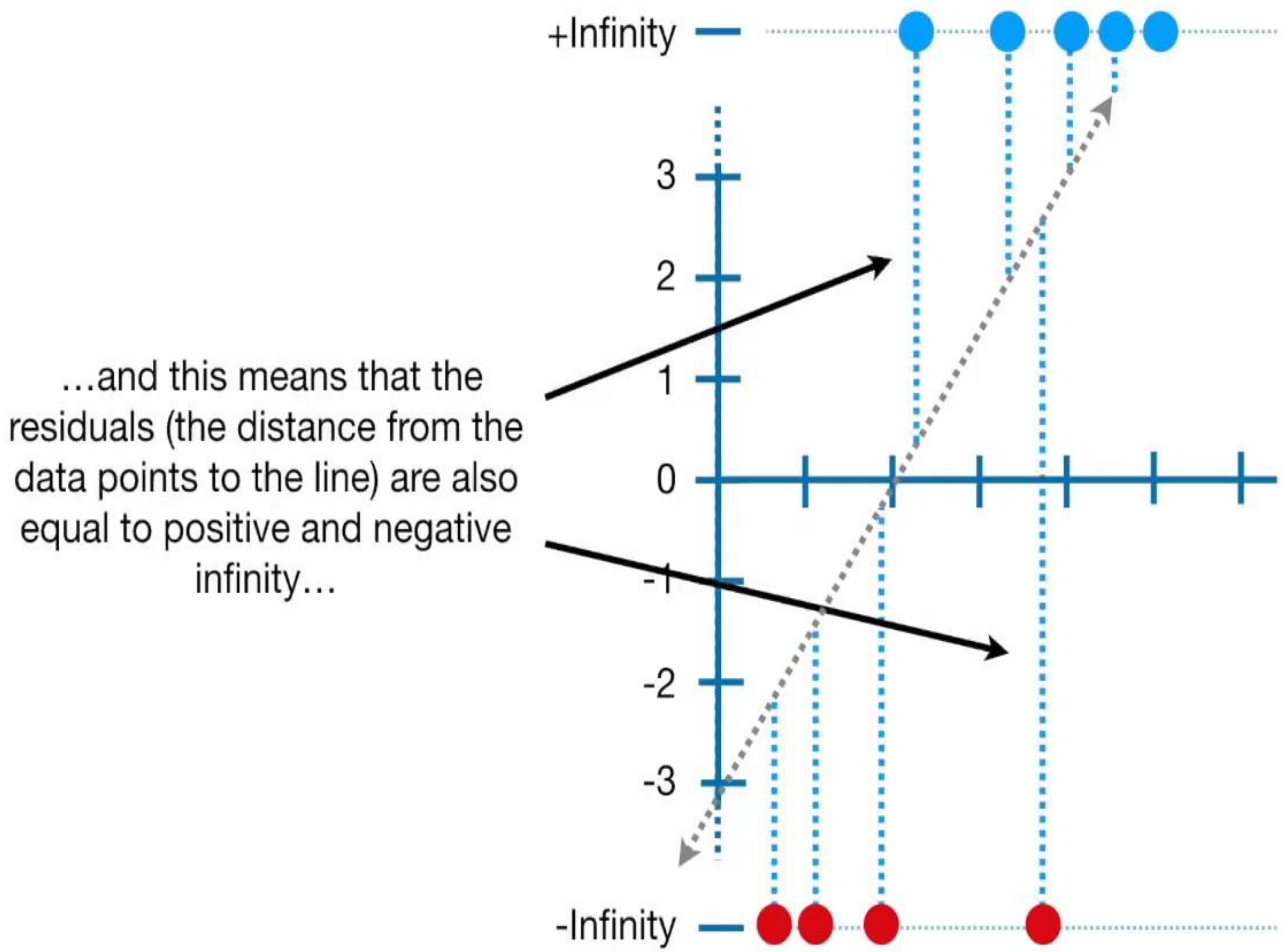

+Infinity
...to log(odds of obesity)
3
2
1
log(odds of obesity) 0
-1
-2
-3
-Infinity
1
0



$$\beta_0 + \beta_1 X$$

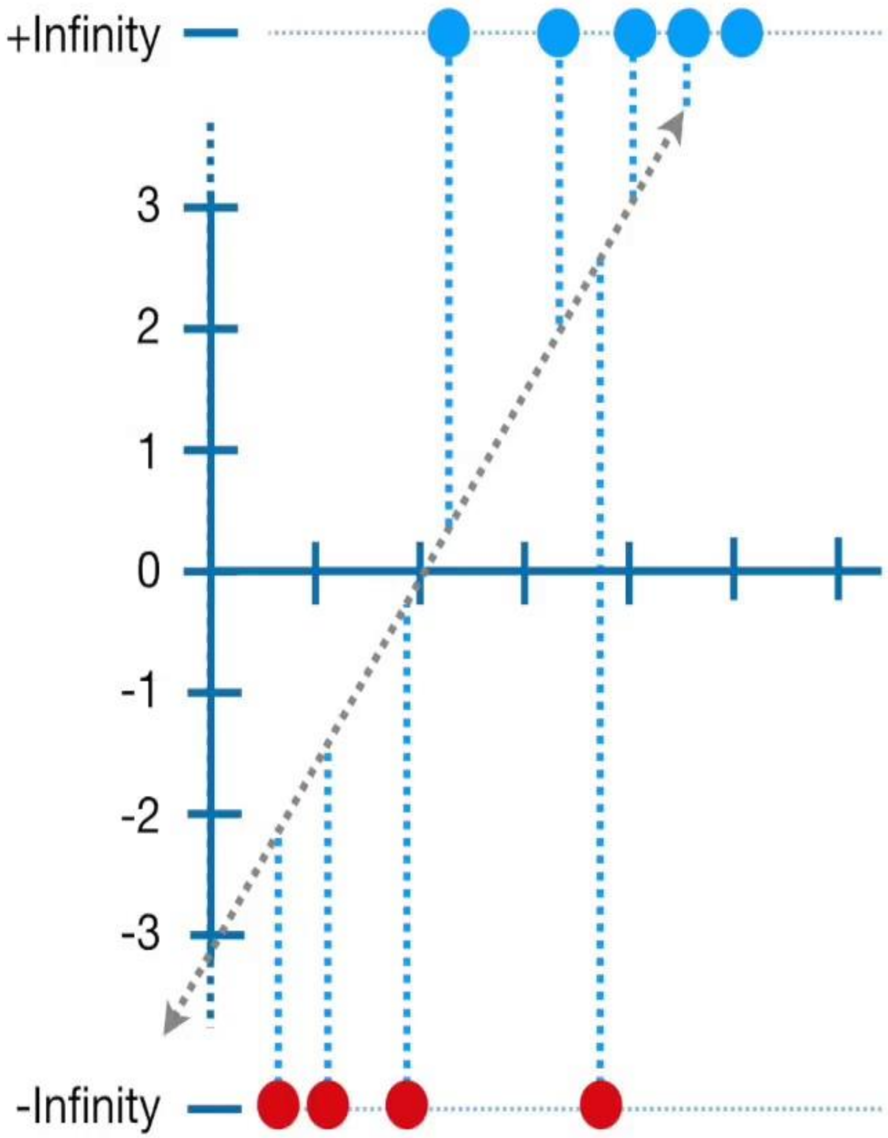


$$\beta_0 + \beta_1 X$$



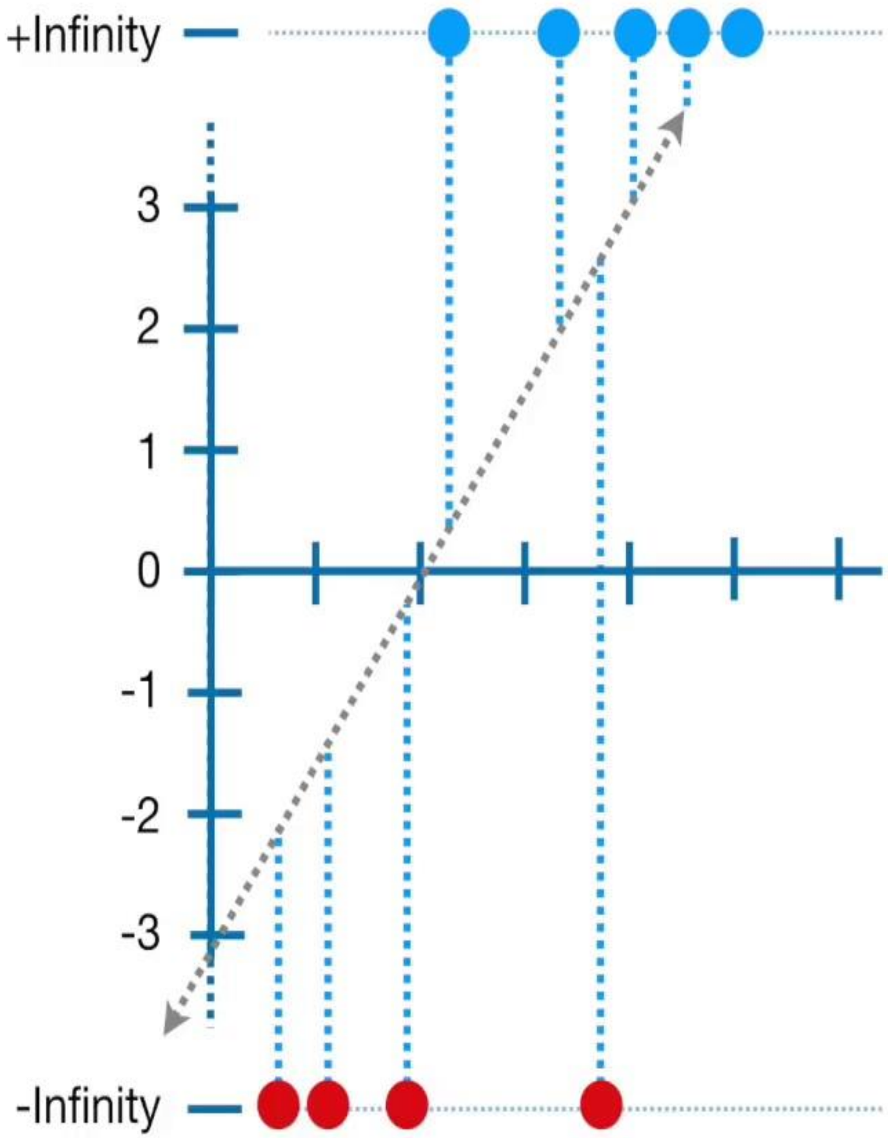
$$\beta_0 + \beta_1 X$$

...and this means we can't use least-squares to find the best fitting line.



$$\beta_0 + \beta_1 X$$

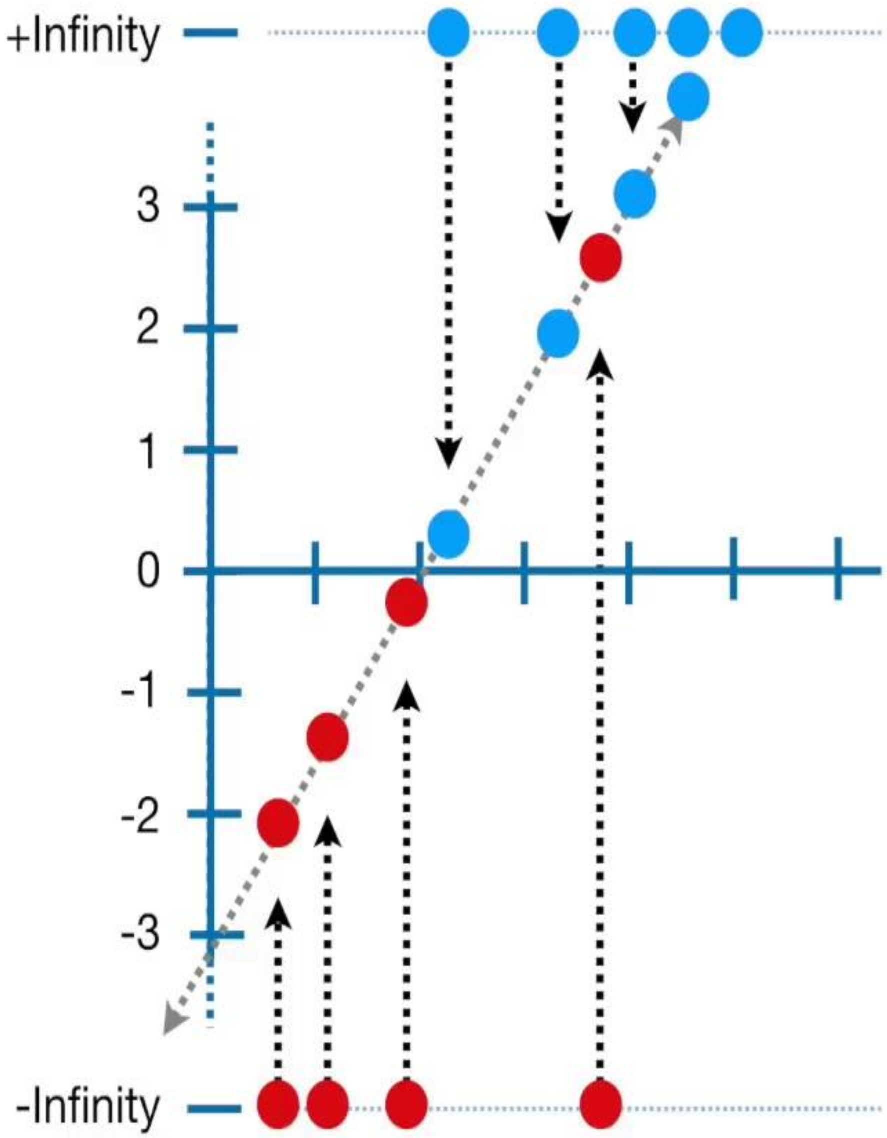
Instead, we use maximum likelihood...



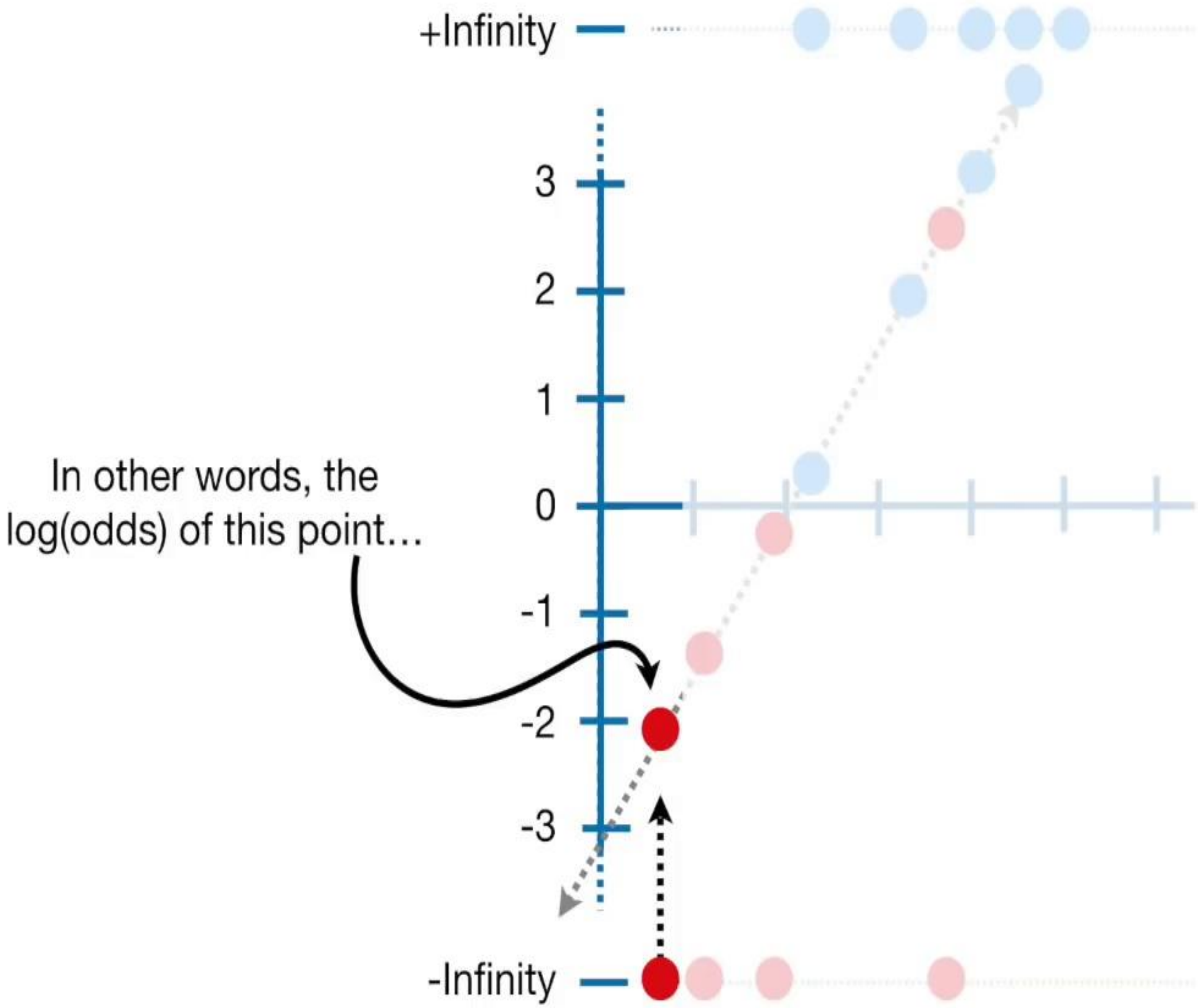
$$\beta_0 + \beta_1 X$$

The first thing we do is project the original data points onto the candidate line.

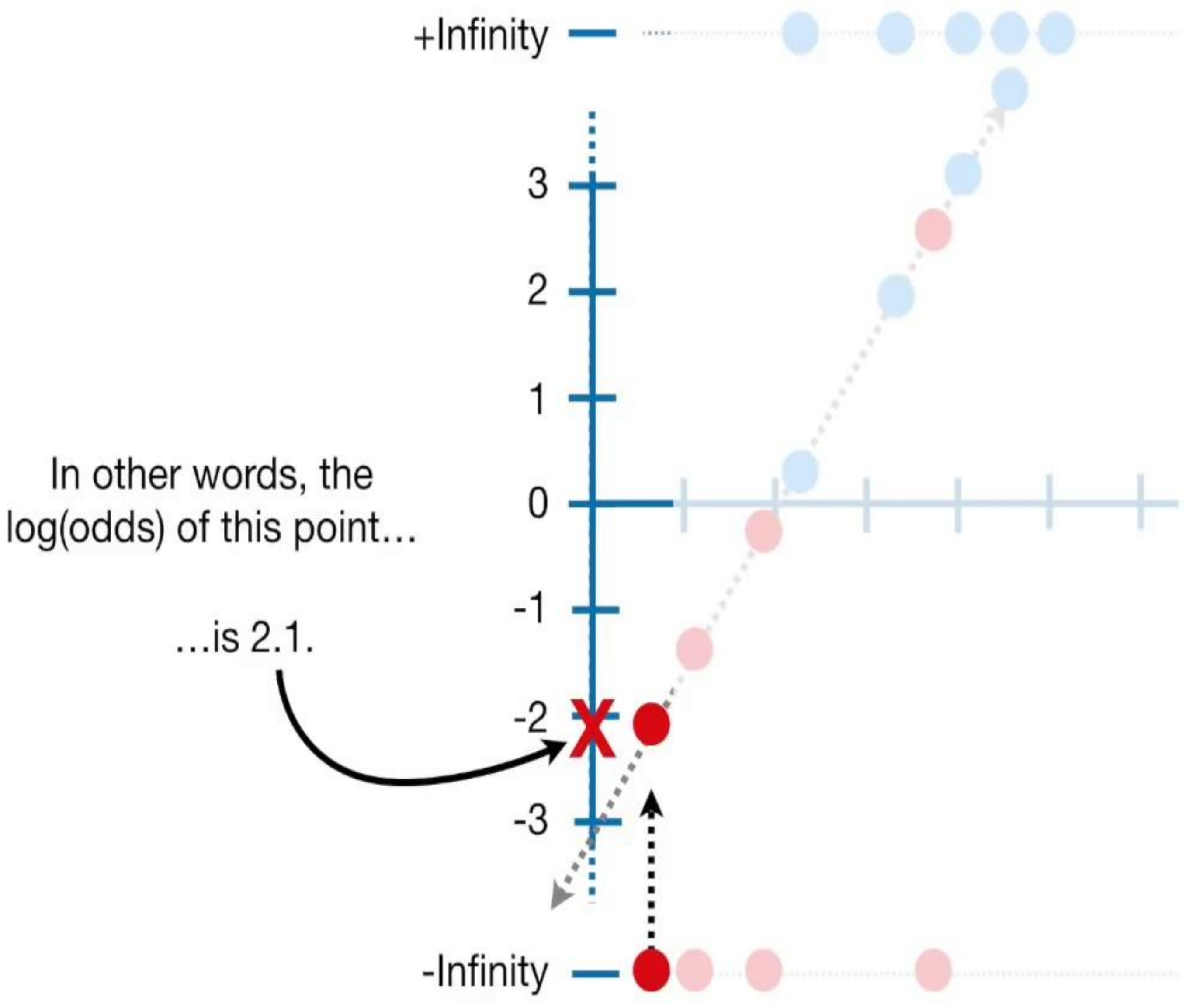
This gives each sample a candidate log(odds) value.



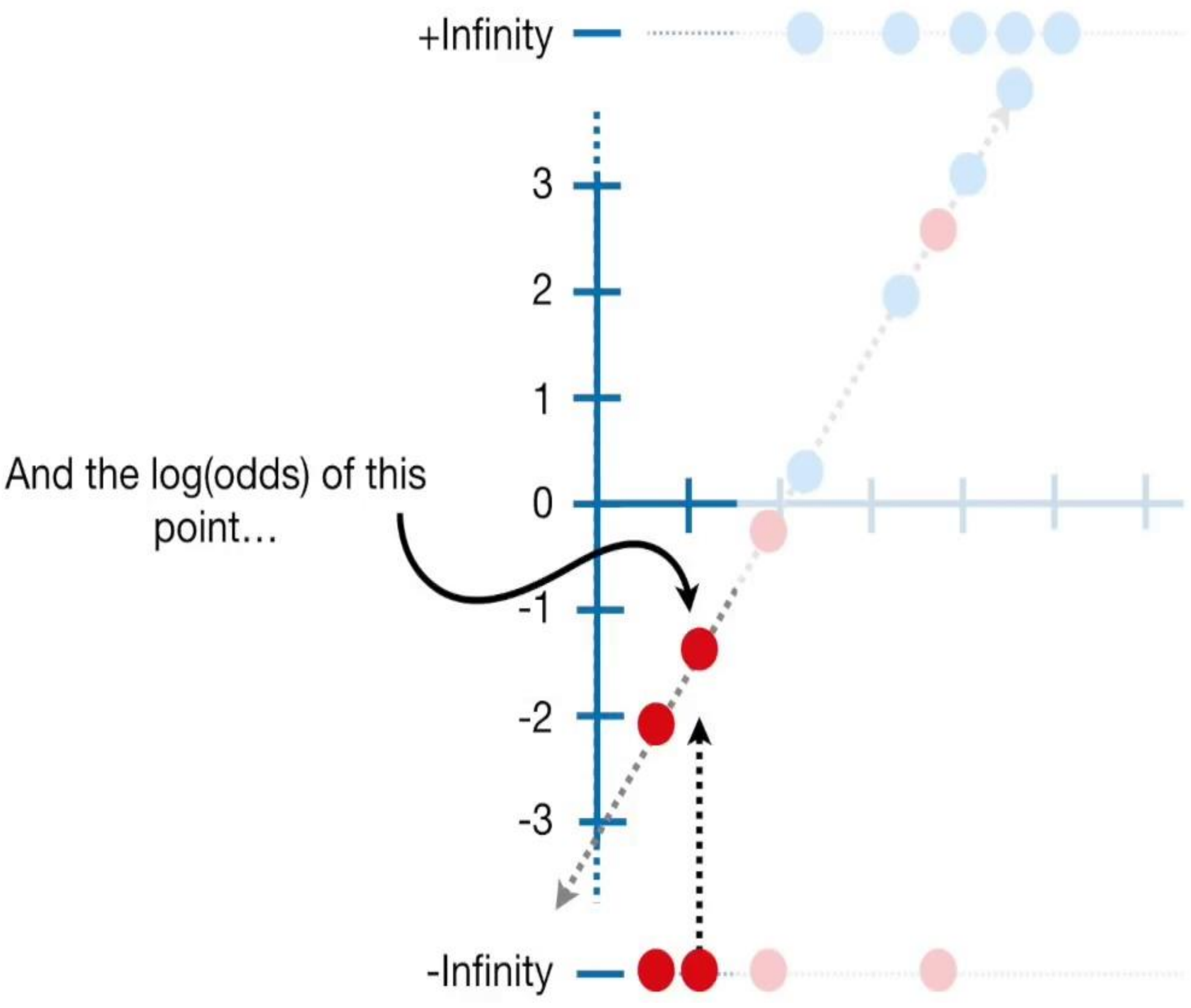
$$\beta_0 + \beta_1 X$$



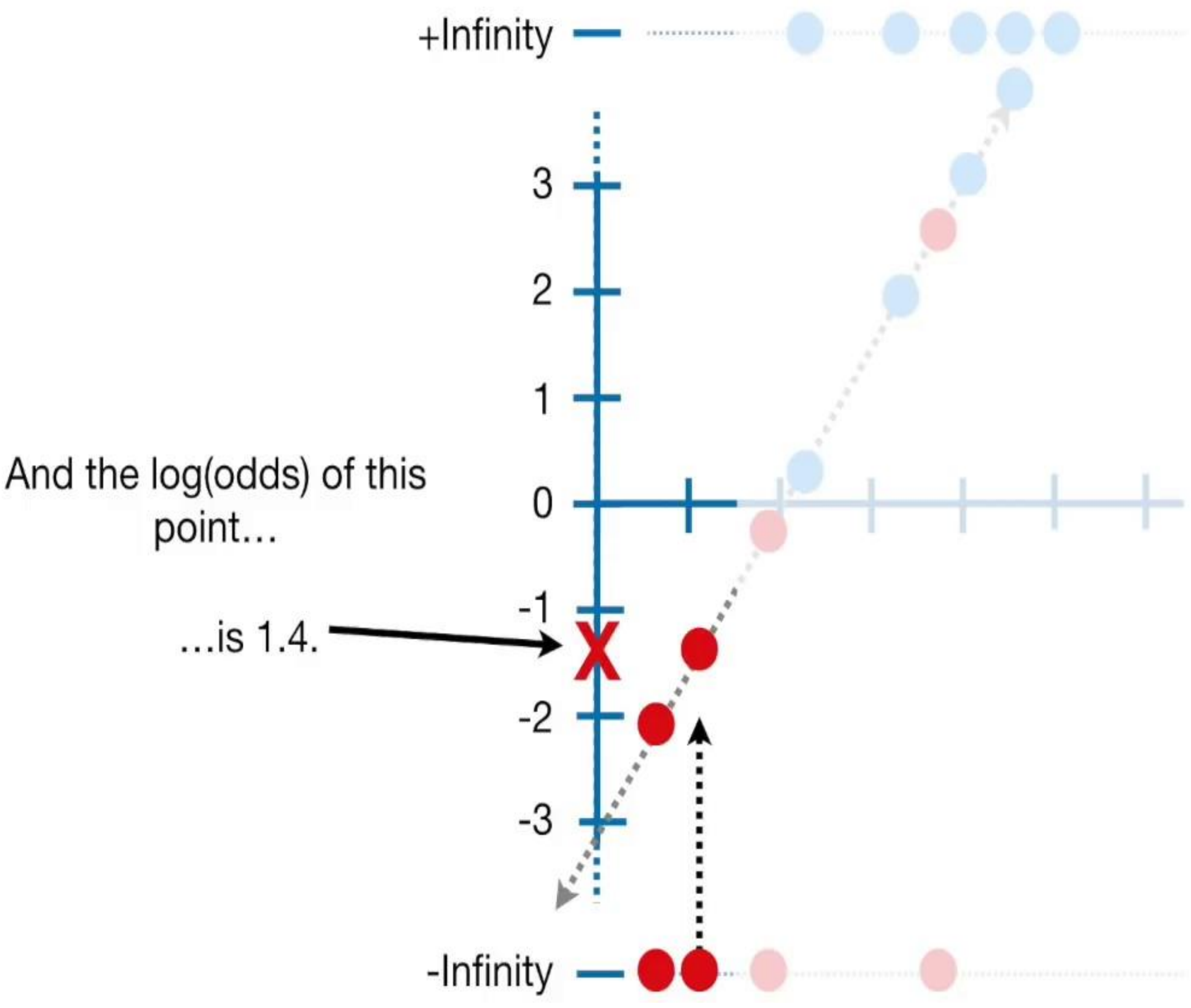
$$\beta_0 + \beta_1 X$$



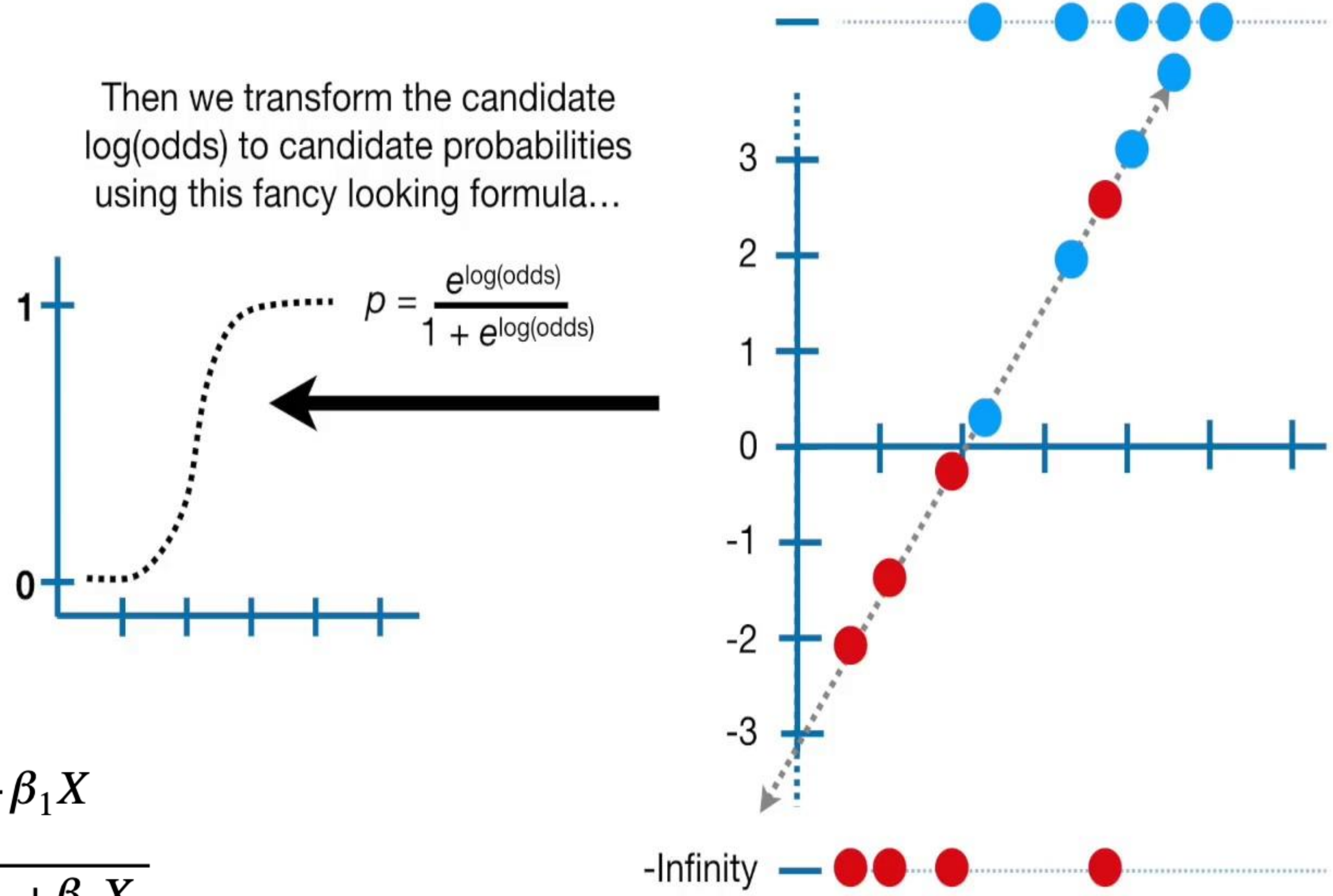
$$\beta_0 + \beta_1 X$$



$$\beta_0 + \beta_1 X$$

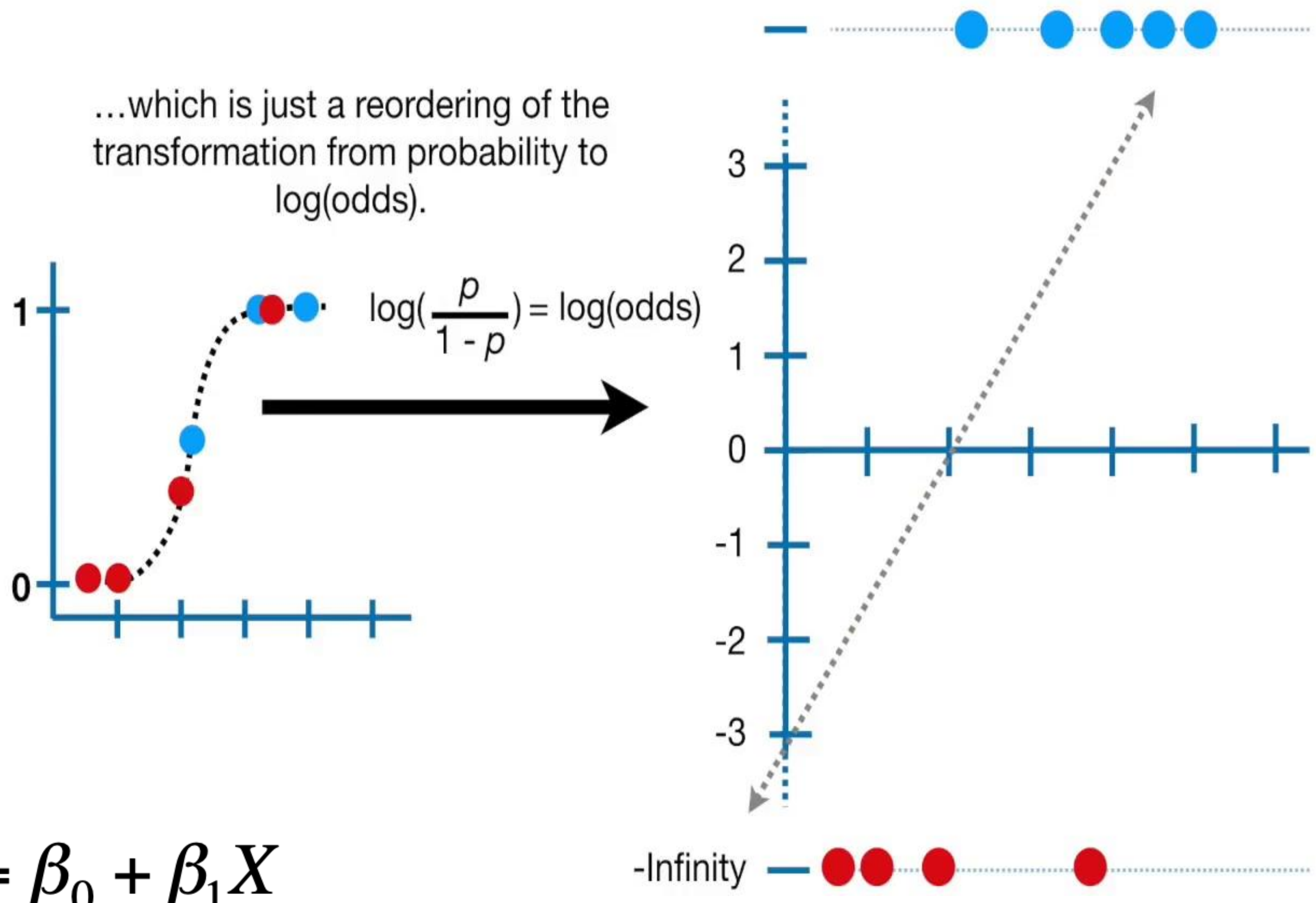


$$\beta_0 + \beta_1 X$$



$$\beta_0 + \beta_1 X$$

$$p = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$



$$\beta_0 + \beta_1 X$$

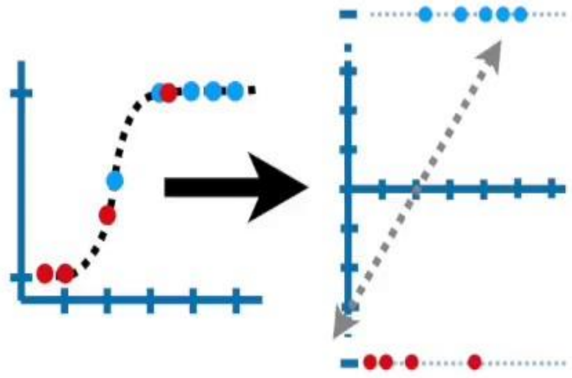
$$\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 X$$

Logistic Regression Equation

$$\log\left(\frac{p}{1-p}\right) = \log(\text{odds})$$

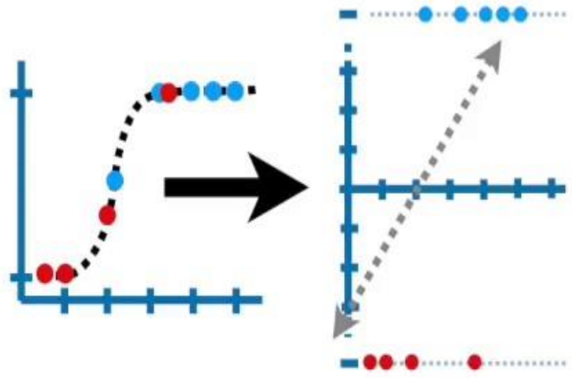
Exponentiate both sides...

$$\frac{p}{1-p} = e^{\log(\text{odds})}$$



Logistic Regression Equation

$$\log\left(\frac{p}{1-p}\right) = \log(\text{odds})$$

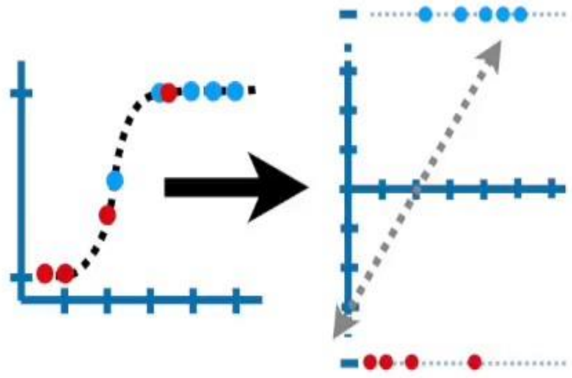


Exponentiate both sides... $$\frac{p}{1-p} = e^{\log(\text{odds})}$$

Multiply both sides by $(1 - p)$... $$p = (1-p)e^{\log(\text{odds})}$$

Logistic Regression Equation

$$\log\left(\frac{p}{1-p}\right) = \log(\text{odds})$$



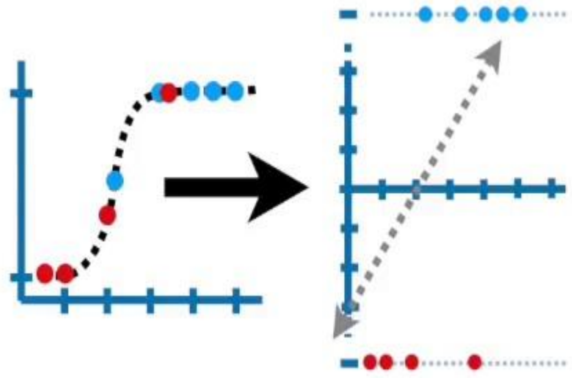
Exponentiate both sides... $\frac{p}{1-p} = e^{\log(\text{odds})}$

Multiply both sides by $(1 - p)$... $p = (1 - p)e^{\log(\text{odds})}$

Multiply $(1 - p)$ and $e^{\log(\text{odds})}$... $p = e^{\log(\text{odds})} - pe^{\log(\text{odds})}$

Logistic Regression Equation

$$\log\left(\frac{p}{1-p}\right) = \log(\text{odds})$$



Exponentiate both sides... $\frac{p}{1-p} = e^{\log(\text{odds})}$

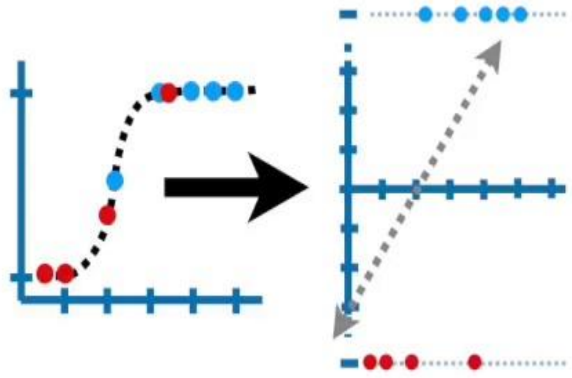
Multiply both sides by $(1 - p)$... $p = (1 - p)e^{\log(\text{odds})}$

Multiply $(1 - p)$ and $e^{\log(\text{odds})}$... $p = e^{\log(\text{odds})} - pe^{\log(\text{odds})}$

Add $pe^{\log(\text{odds})}$ to both sides... $p + pe^{\log(\text{odds})} = e^{\log(\text{odds})}$

Logistic Regression Equation

$$\log\left(\frac{p}{1-p}\right) = \log(\text{odds})$$



Exponentiate both sides... $\frac{p}{1-p} = e^{\log(\text{odds})}$

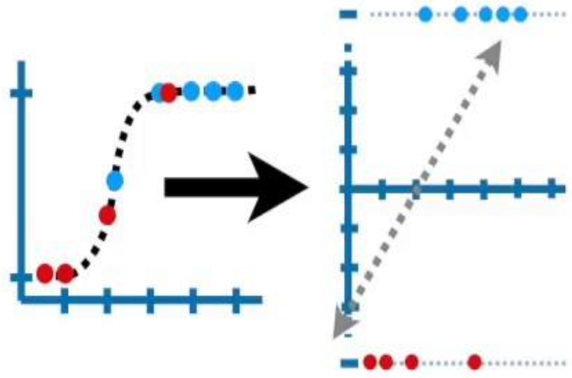
Multiply both sides by (1 - *p*)... $p = (1-p)e^{\log(\text{odds})}$

Multiply (1 - *p*) and $e^{\log(\text{odds})}$... $p = e^{\log(\text{odds})} - pe^{\log(\text{odds})}$

Add $pe^{\log(\text{odds})}$ to both sides... $p + pe^{\log(\text{odds})} = e^{\log(\text{odds})}$

Pull *p* out... $p(1 + e^{\log(\text{odds})}) = e^{\log(\text{odds})}$

Logistic Regression Equation



$$\log\left(\frac{p}{1-p}\right) = \log(\text{odds})$$

Exponentiate both sides... $$\frac{p}{1-p} = e^{\log(\text{odds})}$$

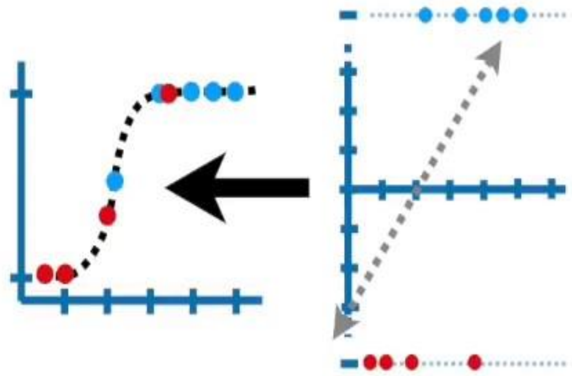
Multiply both sides by $(1 - p)$... $$p = (1 - p)e^{\log(\text{odds})}$$

Multiply $(1 - p)$ and $e^{\log(\text{odds})}$... $$p = e^{\log(\text{odds})} - pe^{\log(\text{odds})}$$

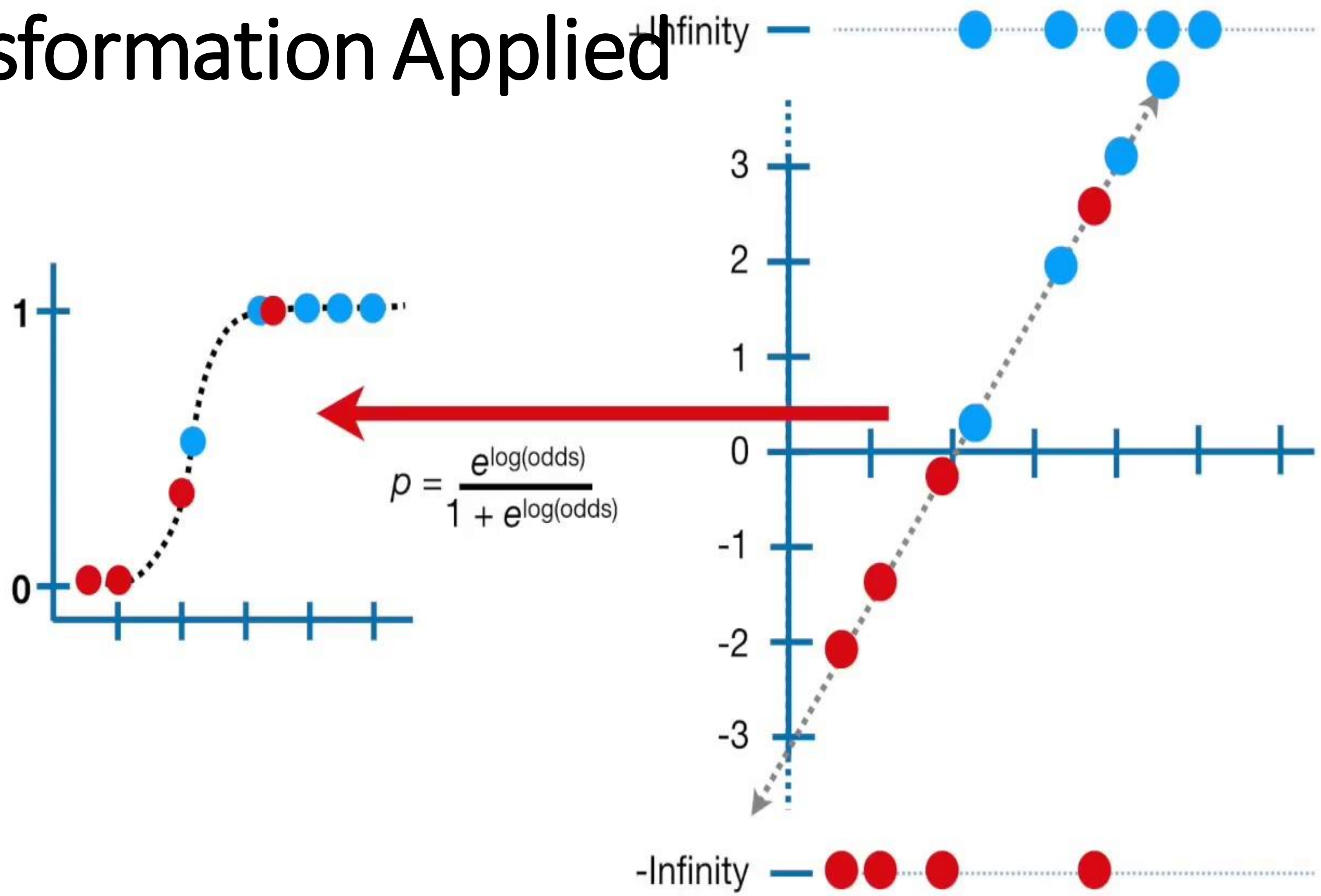
Add $pe^{\log(\text{odds})}$ to both sides... $$p + pe^{\log(\text{odds})} = e^{\log(\text{odds})}$$

Pull p out... $$p(1 + e^{\log(\text{odds})}) = e^{\log(\text{odds})}$$

Divide both sides by $(1 + e^{\log(\text{odds})})$... $$p = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$$

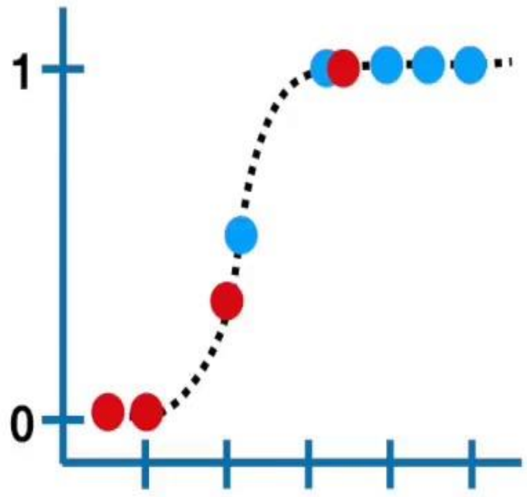


Transformation Applied

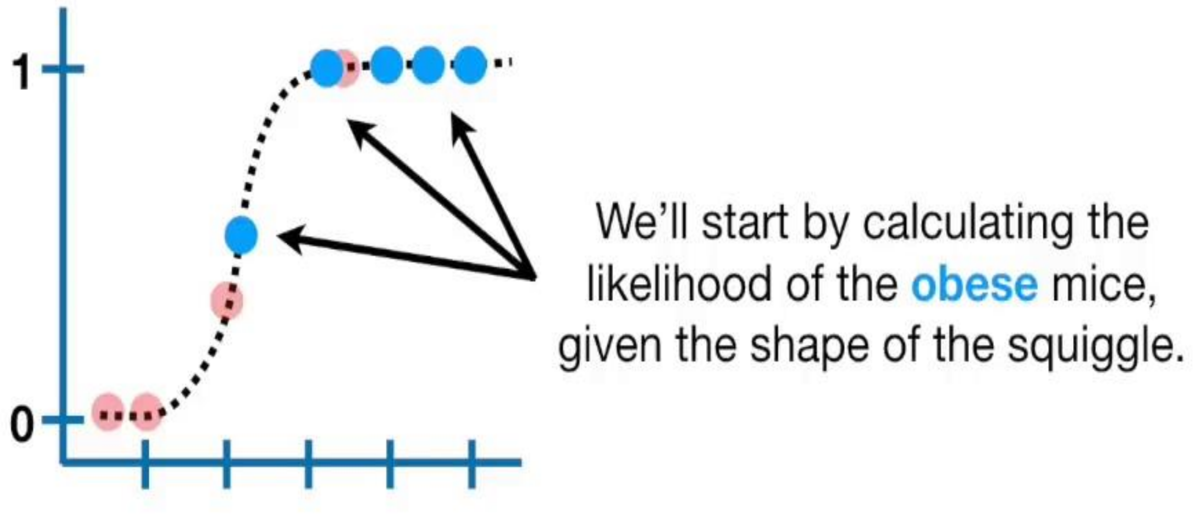


Find Best Line

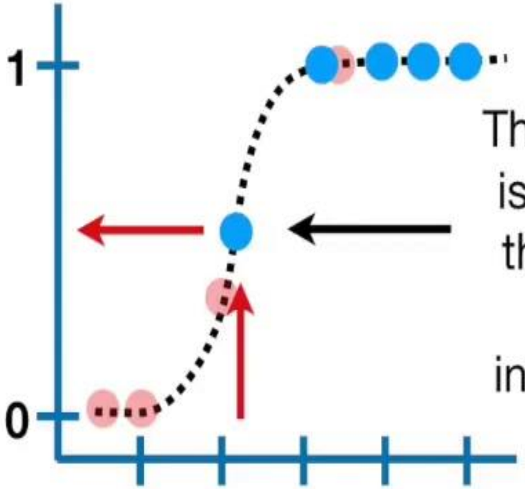
Now we use the observed status (obese or **not obese**) to calculate their likelihood given the shape of the squiggly line.



Find Best Line

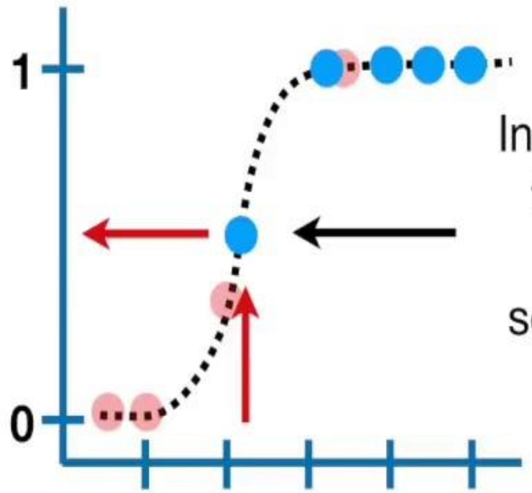


Find Best Line



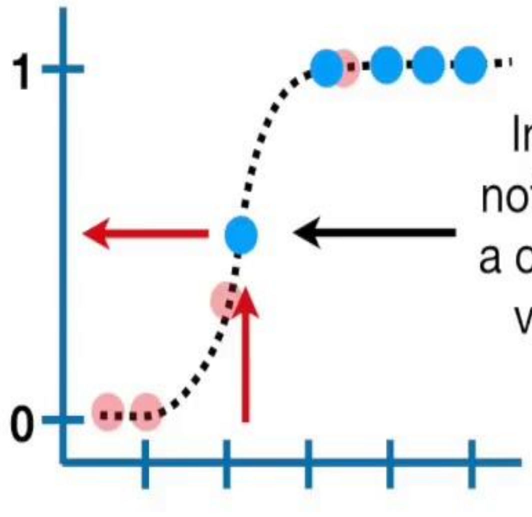
The likelihood that this mouse is obese, given the shape of the squiggle, is the value on the y-axis where point intersects the squiggle, 0.49.

Find Best Line



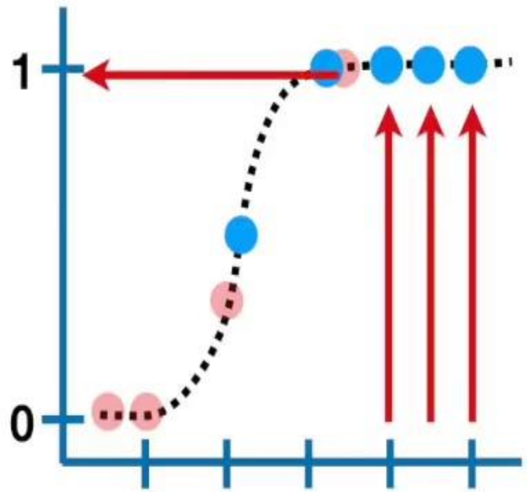
In other words, the likelihood that this mouse is **obese**, given the shape of the squiggle, is the same as the predicted probability.

Find Best Line



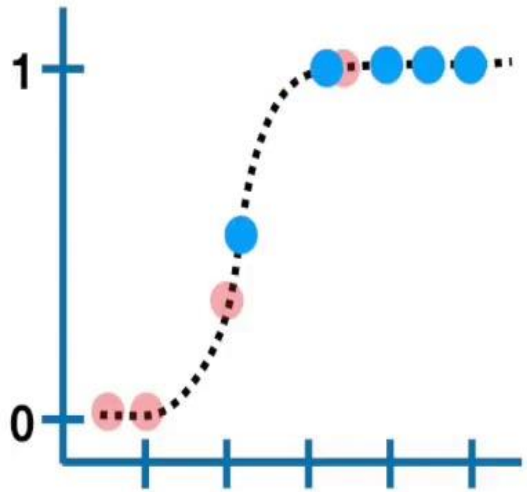
In this case, the probability is not calculated as the area under a curve, but instead is the y-axis value, and that's why it is the same as the likelihood.

Find Best Line



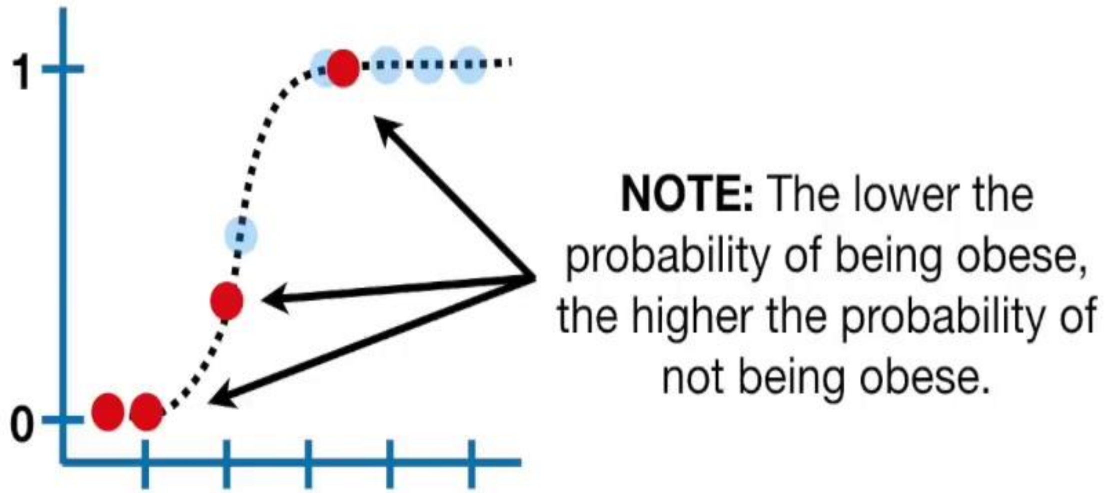
The likelihoods that these mice are obese are 0.91, 0.91 and 0.92

likelihood of data given the squiggle = 0.49 x 0.9 x 0.91 x 0.91 x 0.92 x …

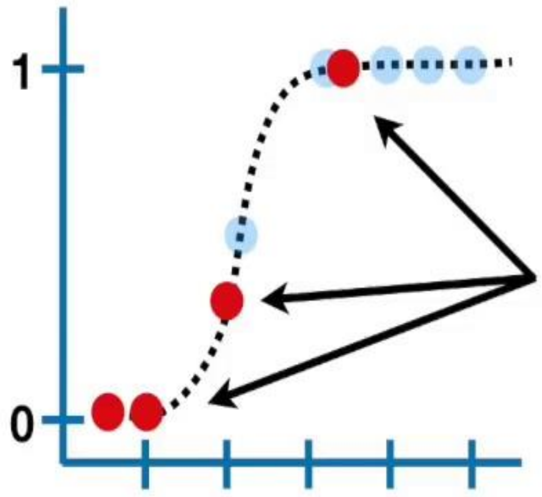


The likelihood for all of the **obese** mice is just the product of the individual likelihoods.

likelihood of data given the squiggle = $0.49 \times 0.9 \times 0.91 \times 0.91 \times 0.92 \times \ldots$

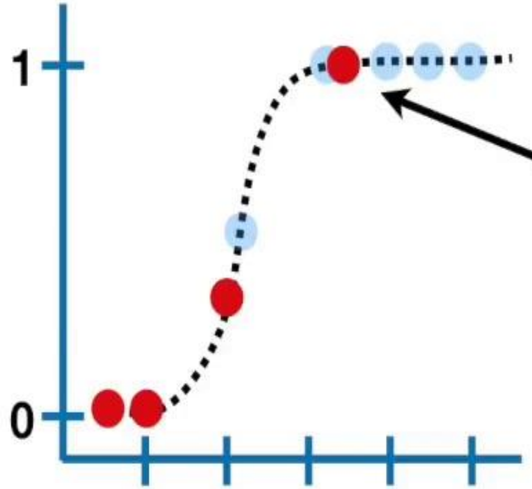


likelihood of data given the squiggle = 0.49 x 0.9 x 0.91 x 0.91 x 0.92 x …



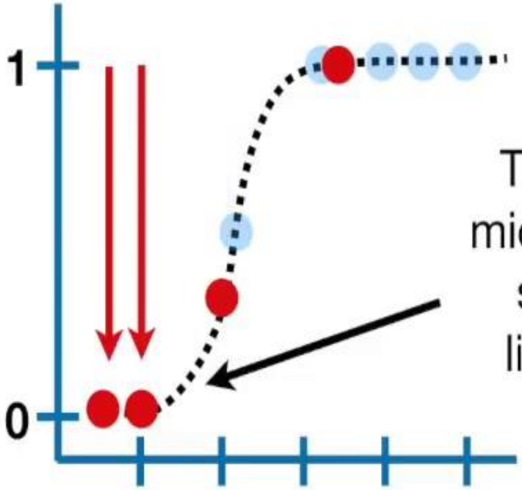
Thus, for these mice, the
likelihood = (1 - probability the mouse is obese)

likelihood of data given the squiggle = 0.49 x 0.9 x 0.91 x 0.91 x 0.92 x …

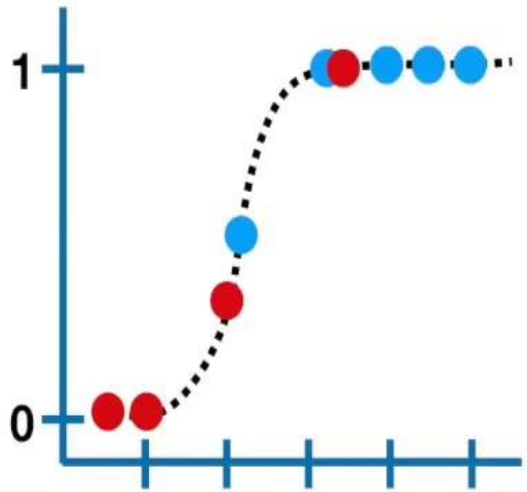


The probability that this mouse is **obese** is 0.9, so the probability and likelihood that it is **not obese** is (**1 - 0.9**)

likelihood of data given the squiggle = 0.49 x 0.9 x 0.91 x 0.91 x 0.92 x …

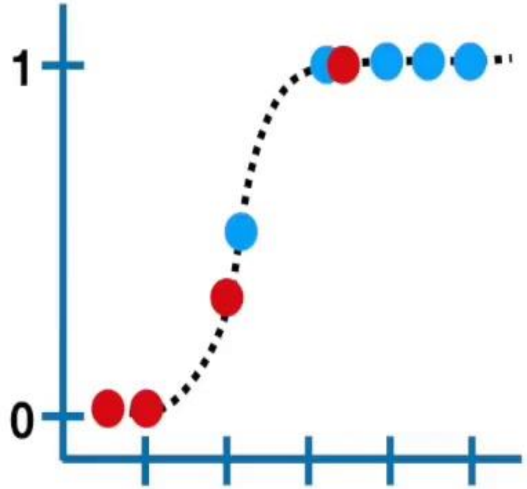


The probabilities that these mice are **obese** are both 0.01, so the probability and the likelihood that they are **not obese** is (**1 - 0.01**) likelihood of data given the squiggle = 0.49 × 0.9 × 0.91 × 0.91 × 0.92 × (1 - 0.9) × (1 - 0.3) × (1 - 0.01) × (1 - 0.01)



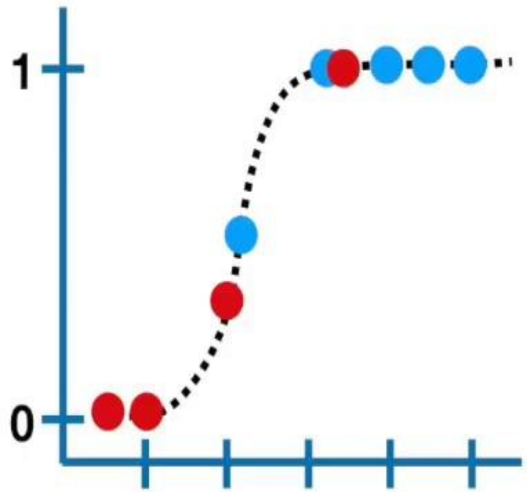
NOTE: Although it is possible to calculate the likelihood as the product of the individual likelihoods, statisticians prefer to calculate the **log of the likelihood** instead.

log(likelihood of data given the squiggle) = log(0.49) + log(0.9) + log(0.91) + log(0.91) + log(0.92) + log(1 - 0.9) + log(1 - 0.3) + log(1 - 0.01) + log(1 - 0.01)



With the log of the likelihood, or "log-likelihood" to those in the know, we **add the logs of the individual likelihoods** instead of multiplying the individual likelihoods…

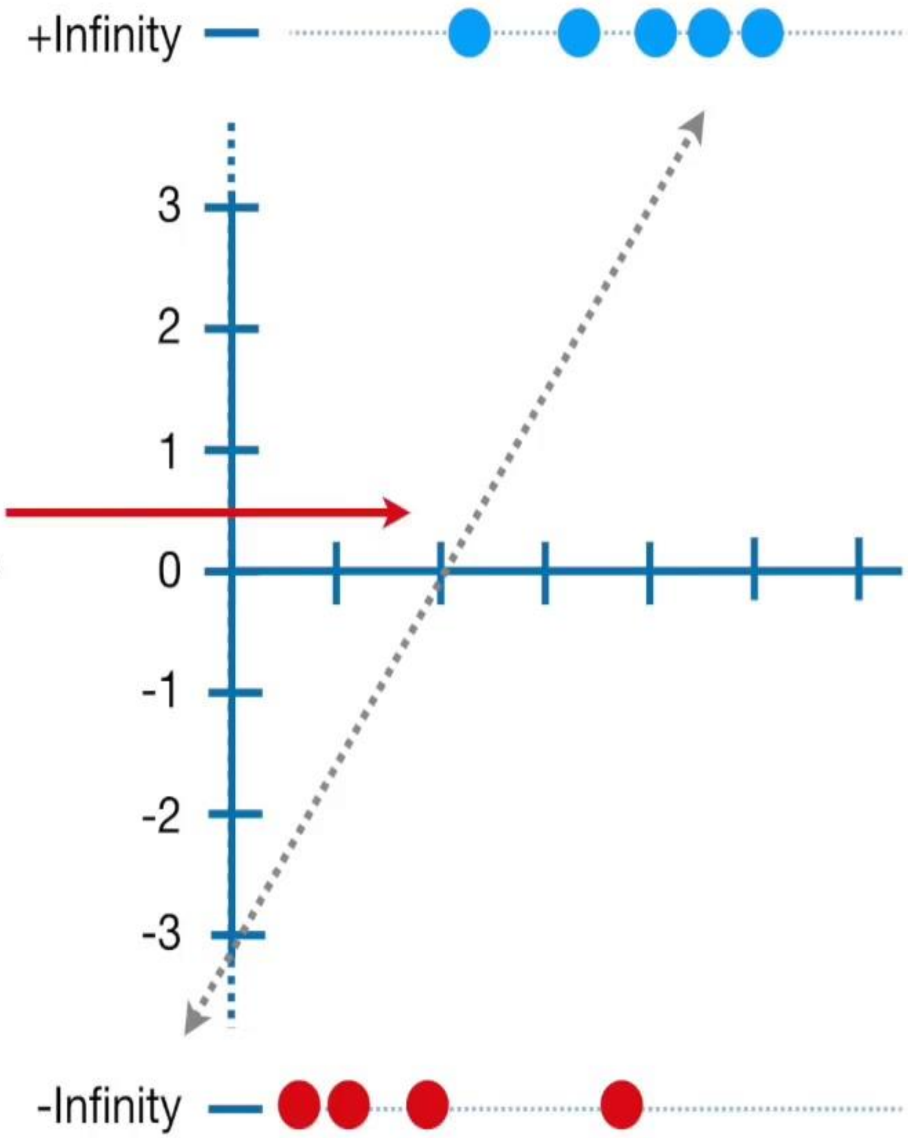
log(likelihood of data given the squiggle) = -3.77



Thus, the log-likelihood of the data given the squiggle is -3.77…

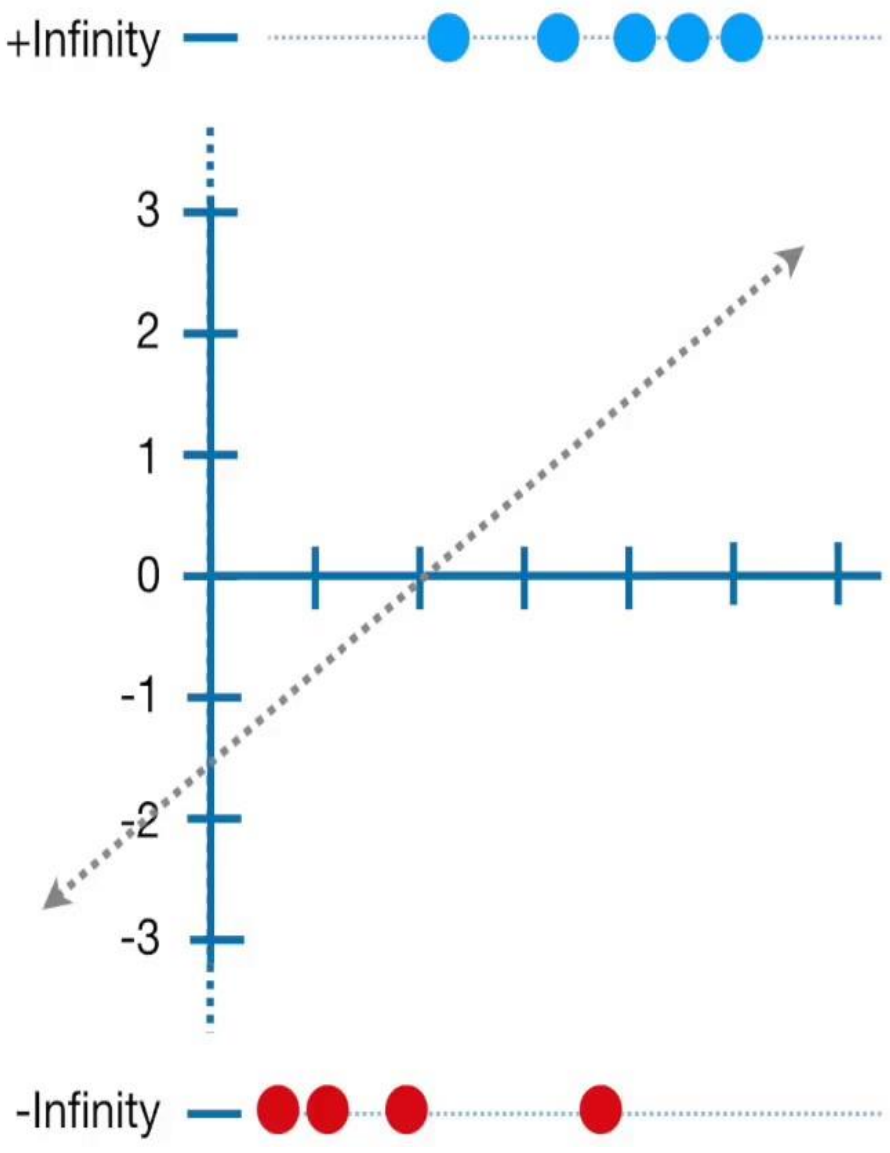
Find Best Line

...and this means that the log-likelihood of the original line is -3.77.



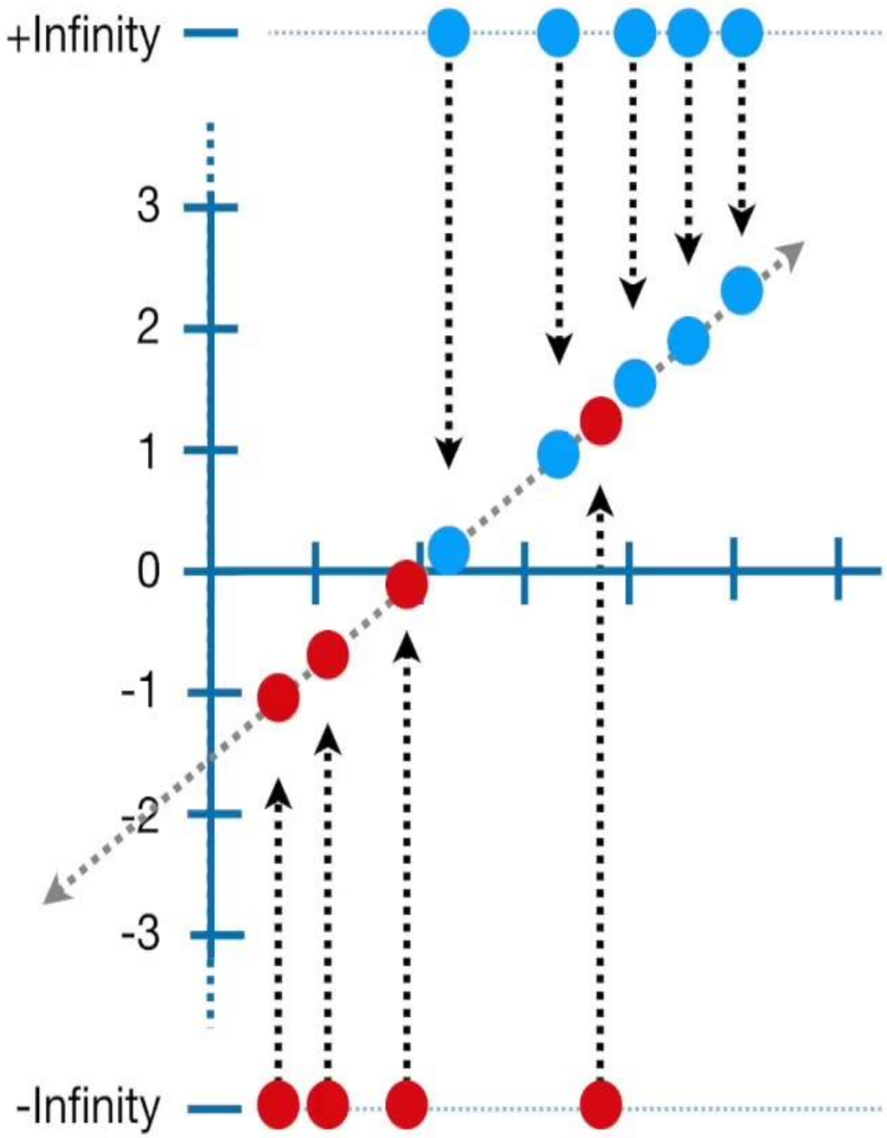
Find Best Line

Now we rotate the line…

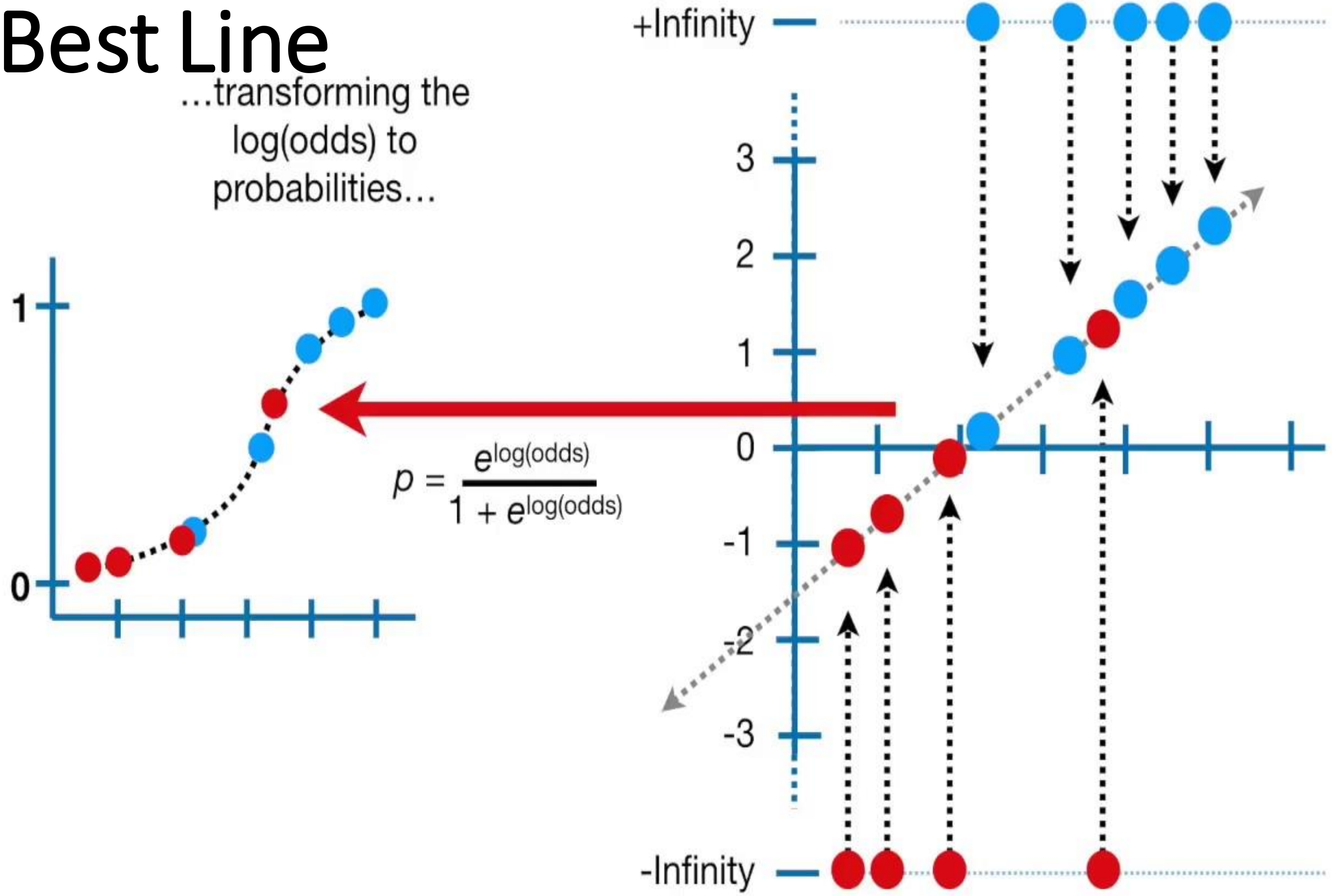


Find Best Line

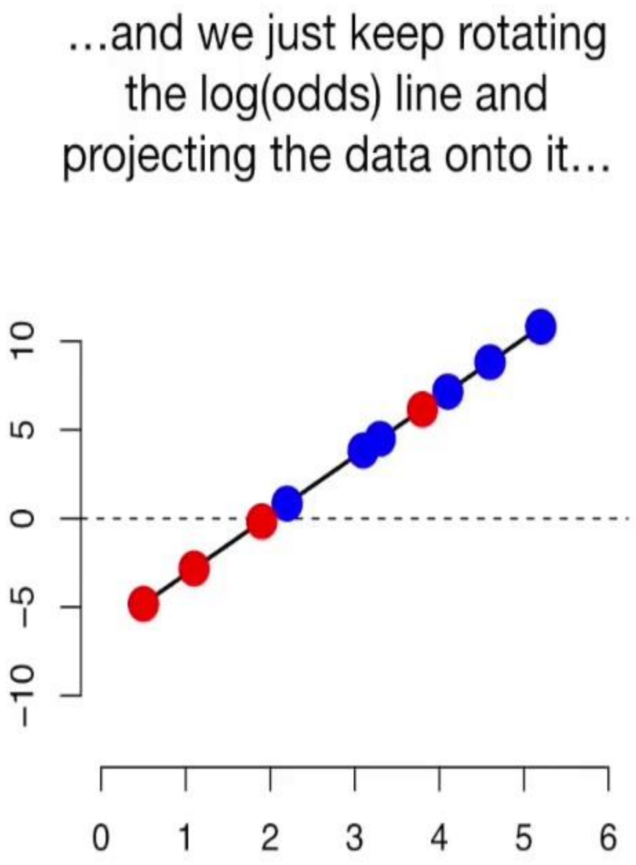
...and calculate its log-likelihood by projecting the data onto it...



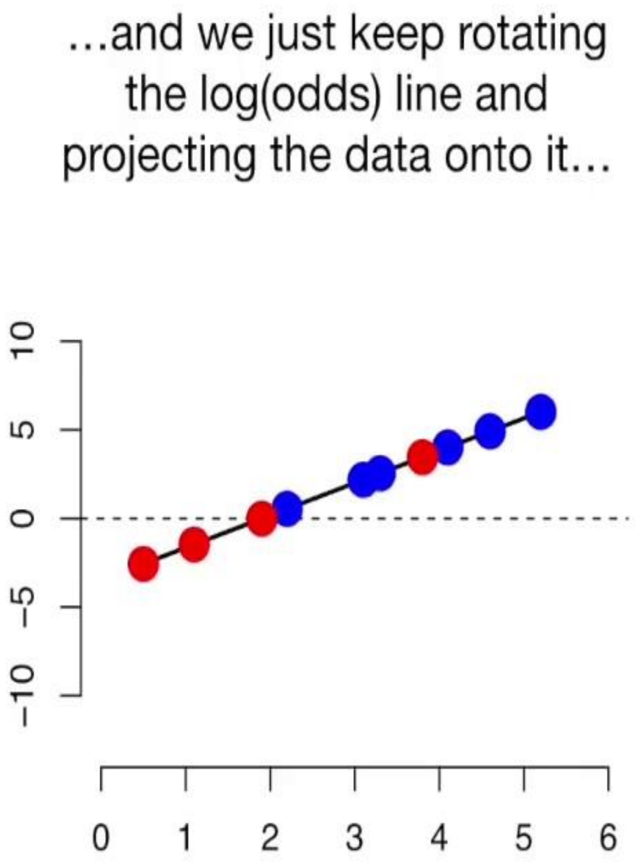
Find Best Line



Find Best Line

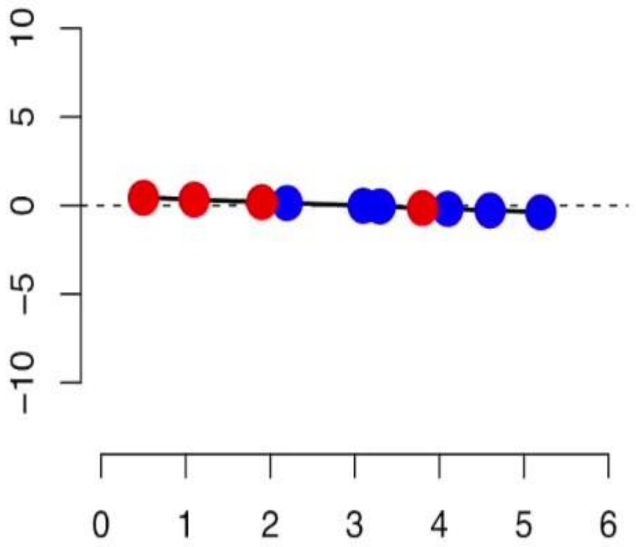


Find Best Line



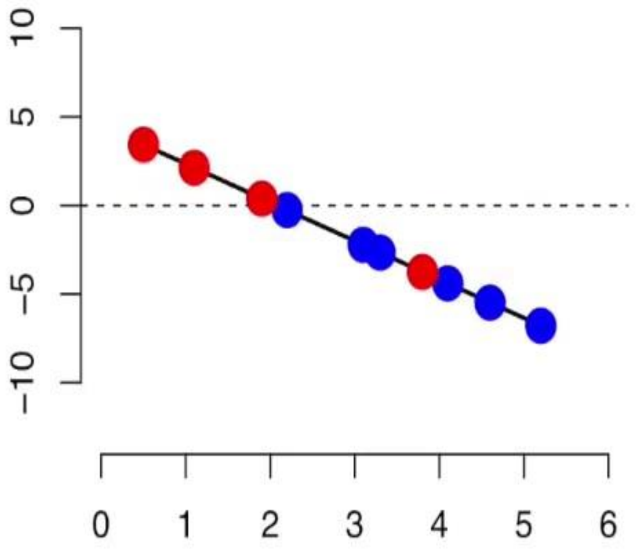
Find Best Line

...and we just keep rotating the log(odds) line and projecting the data onto it...

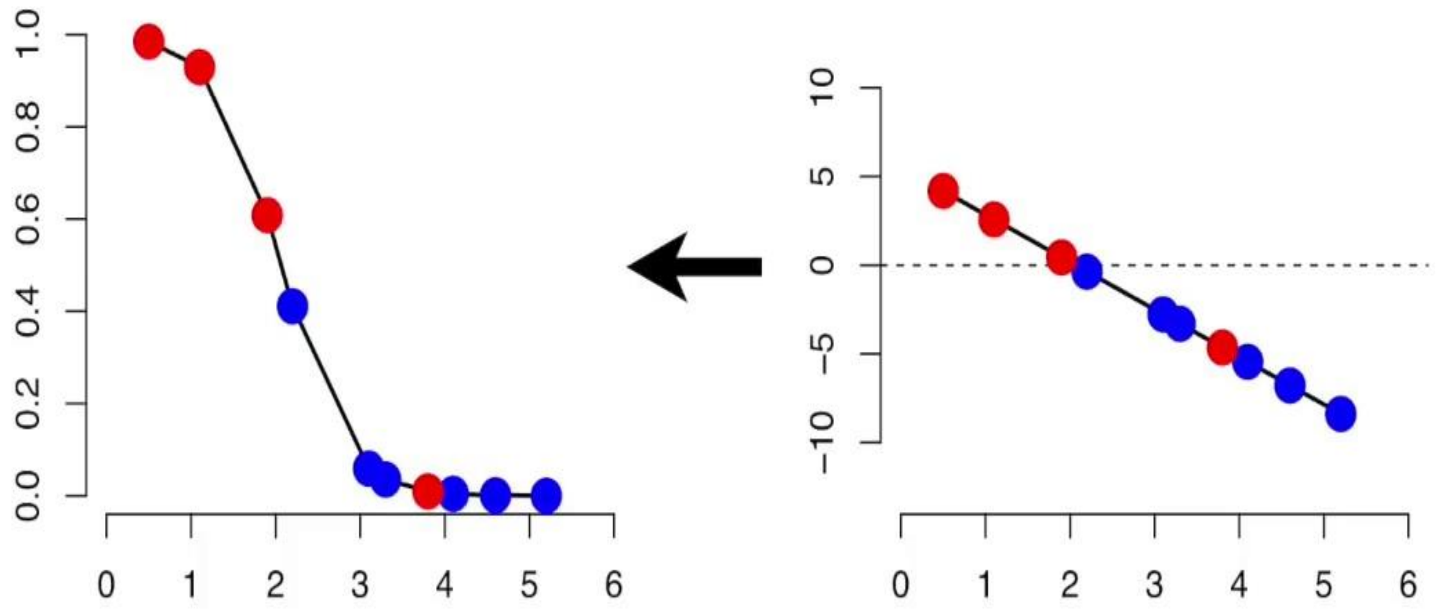


Find Best Line

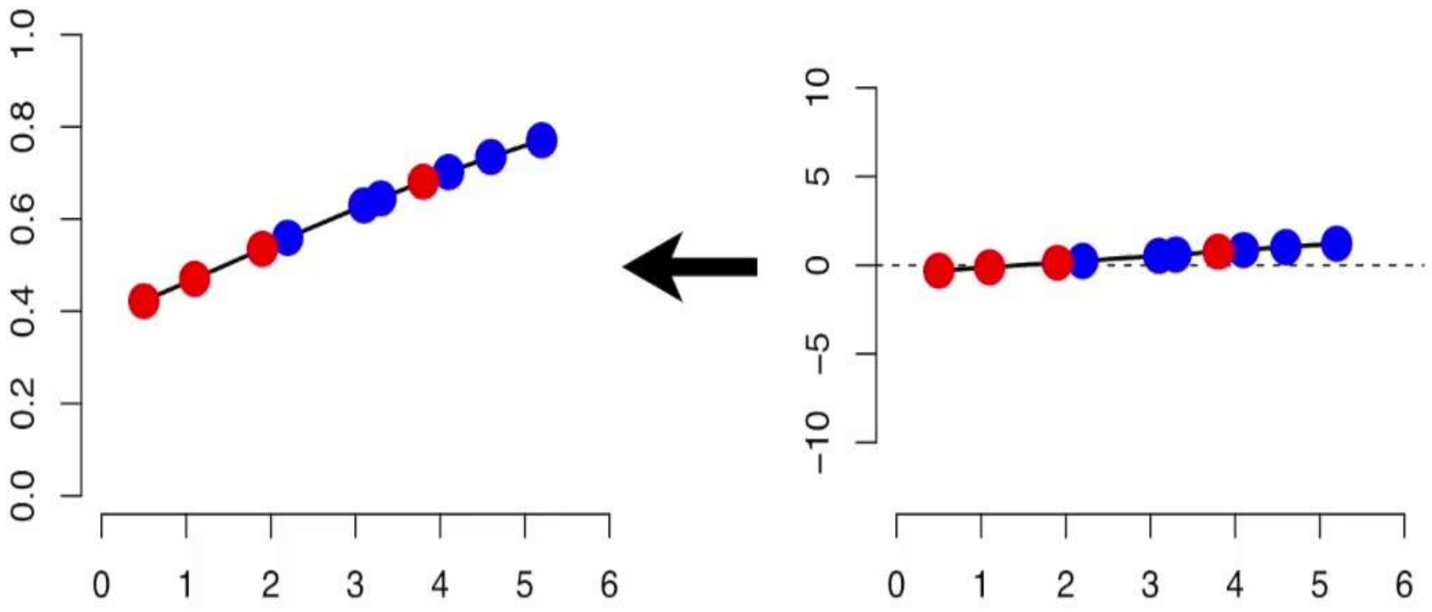
...and we just keep rotating the log(odds) line and projecting the data onto it...



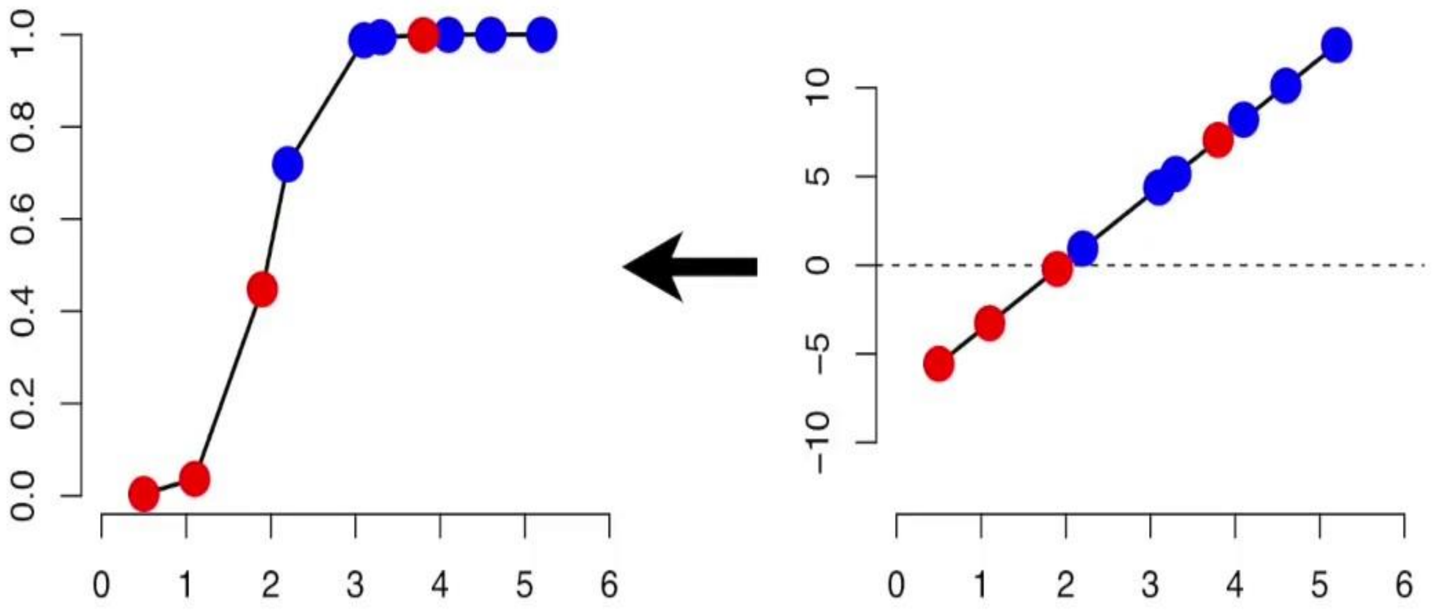
...and transforming it to probabilities and calculating the log-likelihood.



...and transforming it to probabilities and calculating the log-likelihood.

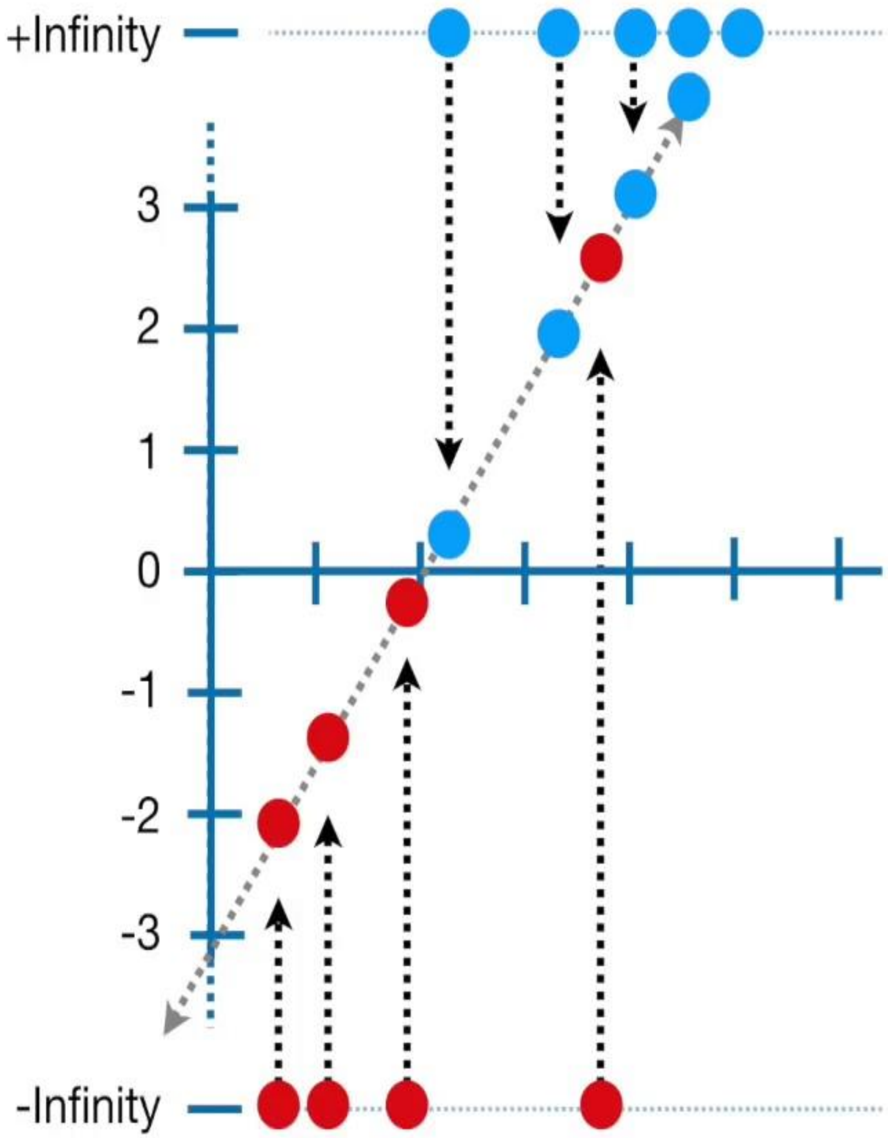


...and transforming it to probabilities and calculating the log-likelihood.

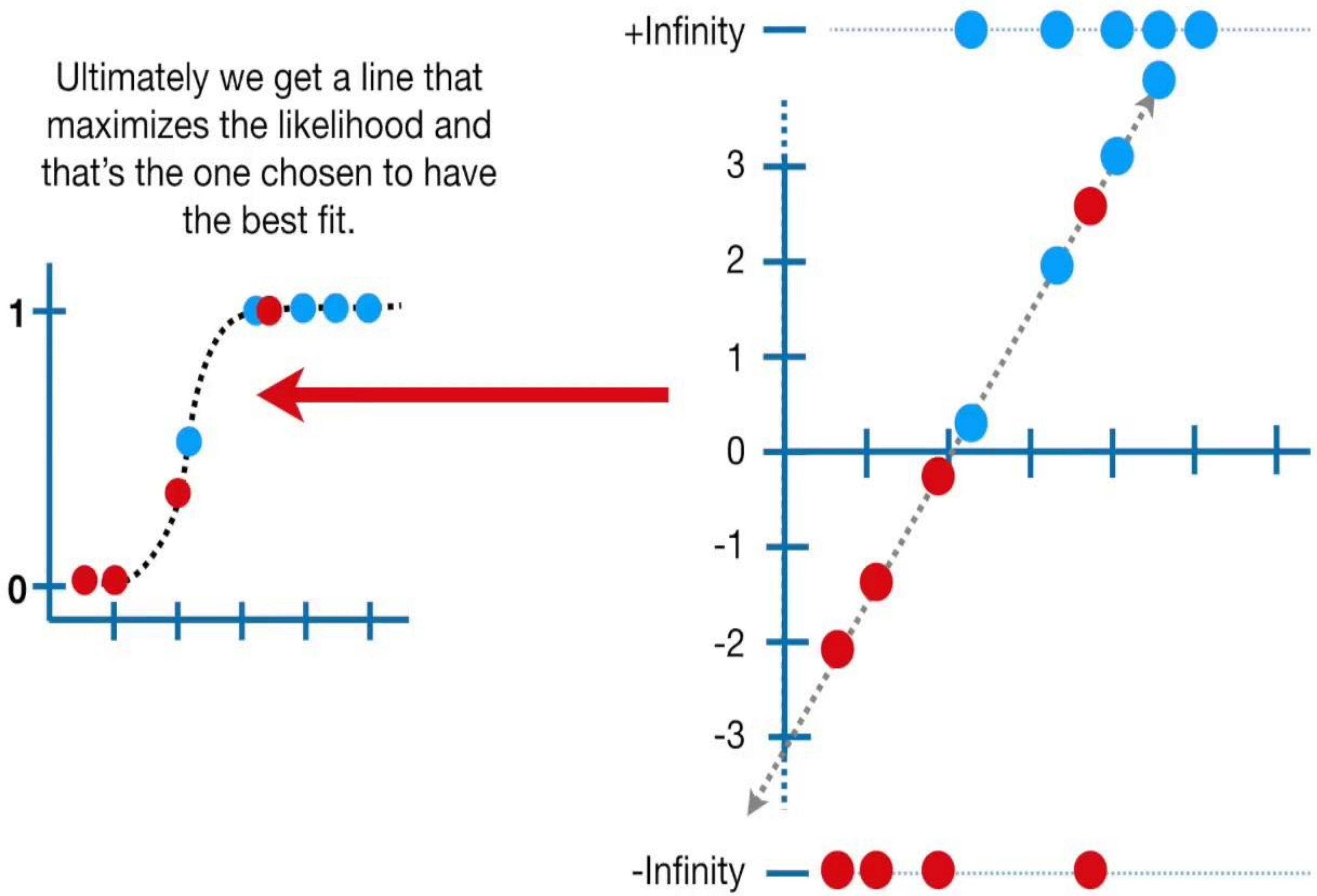


Find Best Line

NOTE: The algorithm that finds the line with the maximum likelihood is pretty smart - each time it rotates the line, it does so in a way that increases the log-likelihood. Thus, the algorithm can find the optimal fit after a few rotations.



Ultimately we get a line that maximizes the likelihood and that's the one chosen to have the best fit.

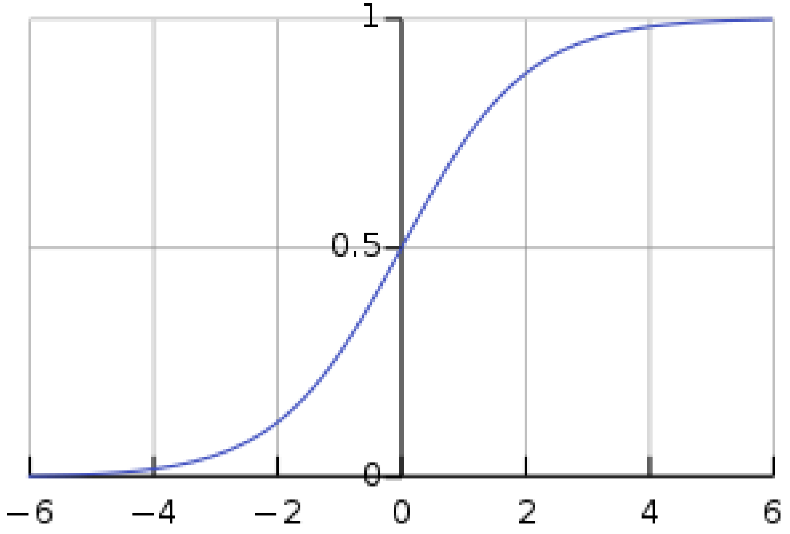


Recovering Probabilities from Log Odds

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X$$

$$\Leftrightarrow \frac{p}{1-p} = e^{\beta_0 + \beta_1 X}$$

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



- which gives *p* as the sigmoid function!

Logistic Regression

- In Logistic Regression we seek a model

$$Y = logit(p) = log(p/(1-p))$$

- That is, the **log odds, i.e., the logit,** is assumed to be linearly related to the independent variable X
- In this way it is possible to solve an ordinary (linear) regression.

Interpretation of Beta1

- Let:
  - odds1 = odds for value X (p/(1–p))
  - odds2 = odds for value X + 1 unit
- Then:

$$\frac{odds2}{odds1} = \frac{e^{b_0 + b_1(X+1)}}{e^{b_0 + b_1 X}}$$

If the odds ratio of two consecutive value is large it means that an increment on X has a large impact in the prediction of Y.

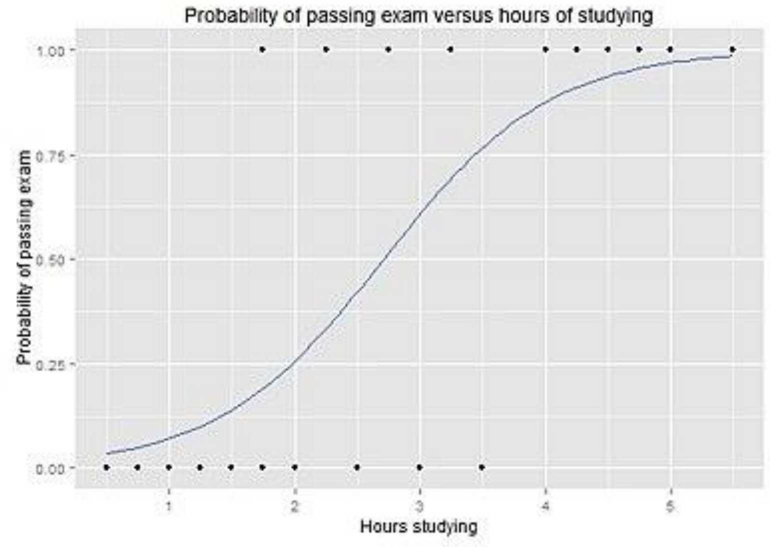
$$= \frac{e^{(b_0 + b_1 X) + b_1}}{e^{b_0 + b_1 X}} = \frac{e^{(b_0 + b_1 X)} e^{b_1}}{e^{b_0 + b_1 X}} = e^{b_1}$$

- The exponent of the slope describes the proportionate rate at which the predicted odds ratio changes with each successive unit of *X*

Example

- Hours: 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 3.25, 3.50, 4.00, 4.25, 4.50, 4.75, 5.00, 5.50
- Pass: 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1

Beta0 = −4.0777, Beta1 = 1.5046



One additional hour of study is estimated to increase log-odds by 1.5046, so multiplying odds by $e^{1.5046}$ = 4.5.

For example, for a student who studies 2 hours we have an estimated probability of passing the exam of 0.26.

Similarly, for a student who studies 4 hours, the estimated probability of passing the exam is 0.87.

References

- Regression. Appendix D. Introduction to Data Mining.