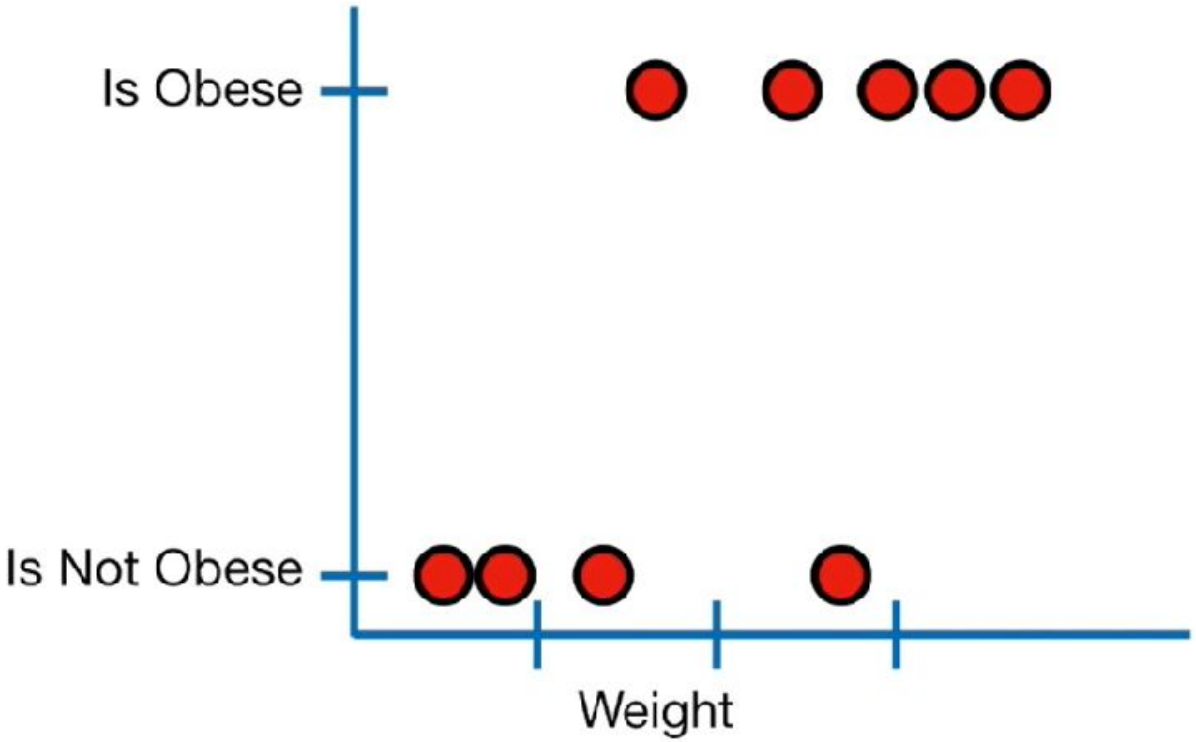
Logistic Regression

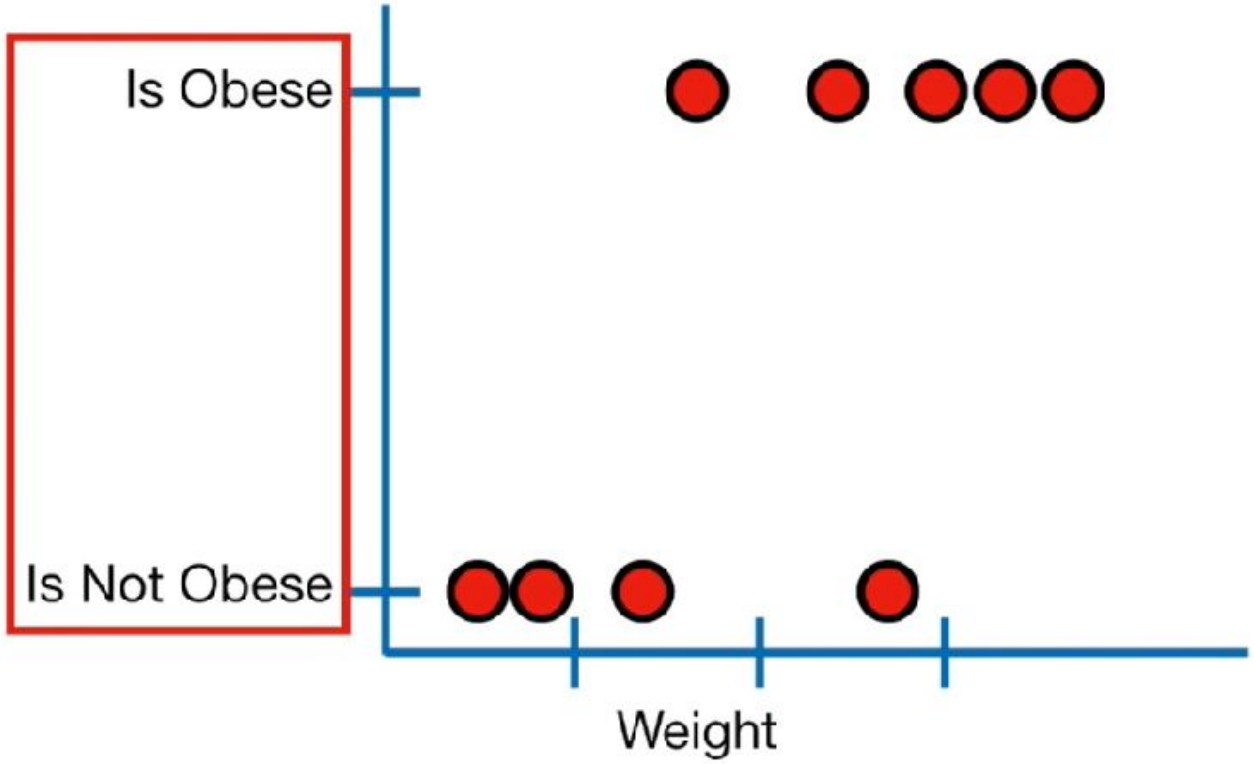
Logistic Regression

Logistic regression is similar to linear regression, except…

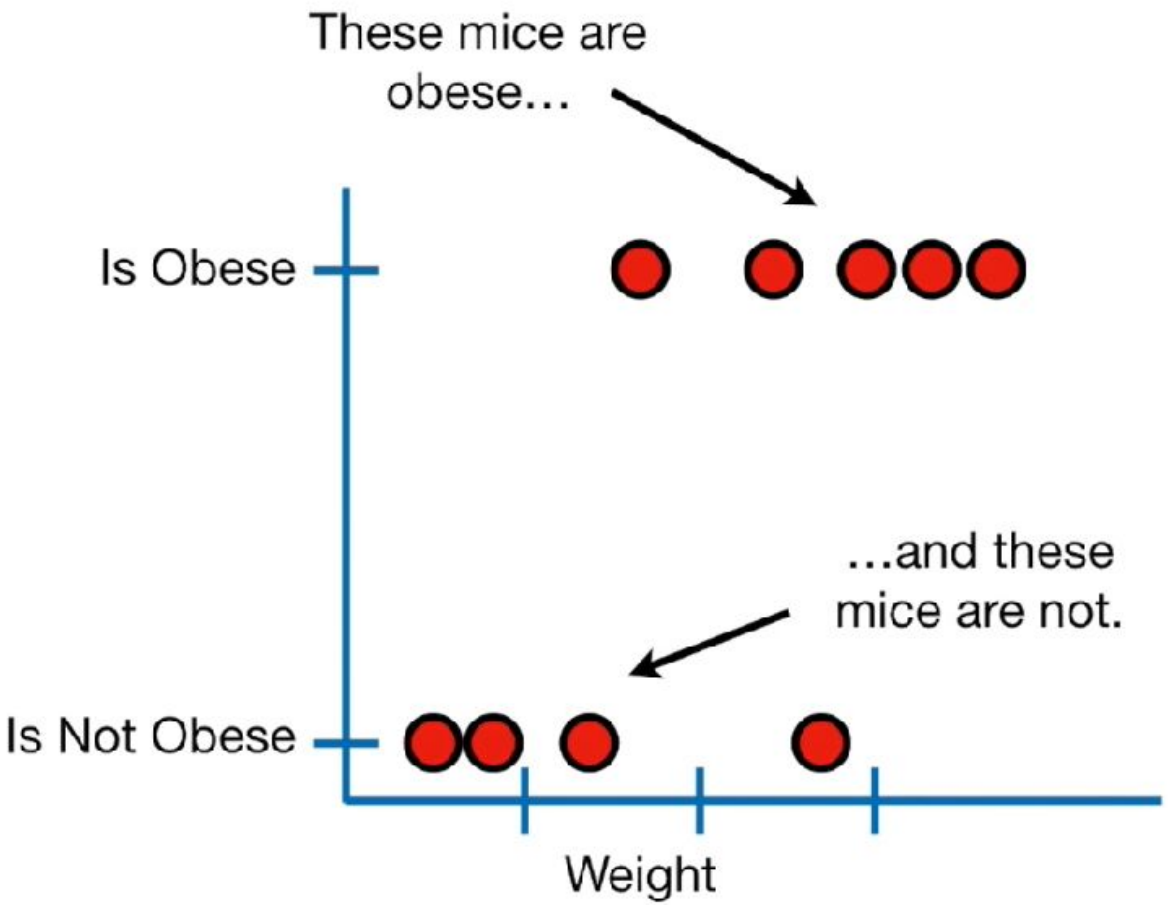


Logistic Regression

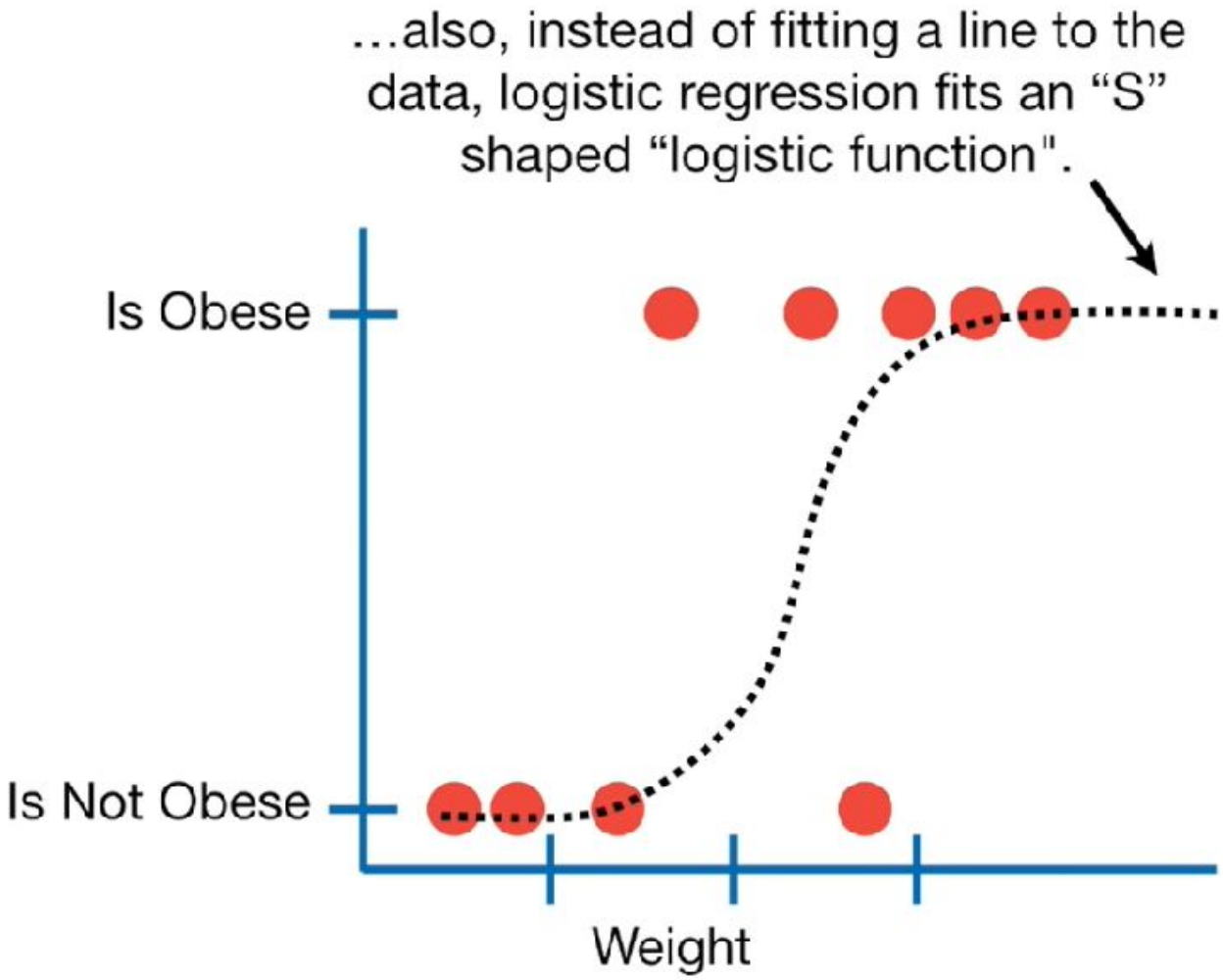
Logistic regression predicts whether something is ***True*** or ***False***, instead of predicting something continuous like ***size***.



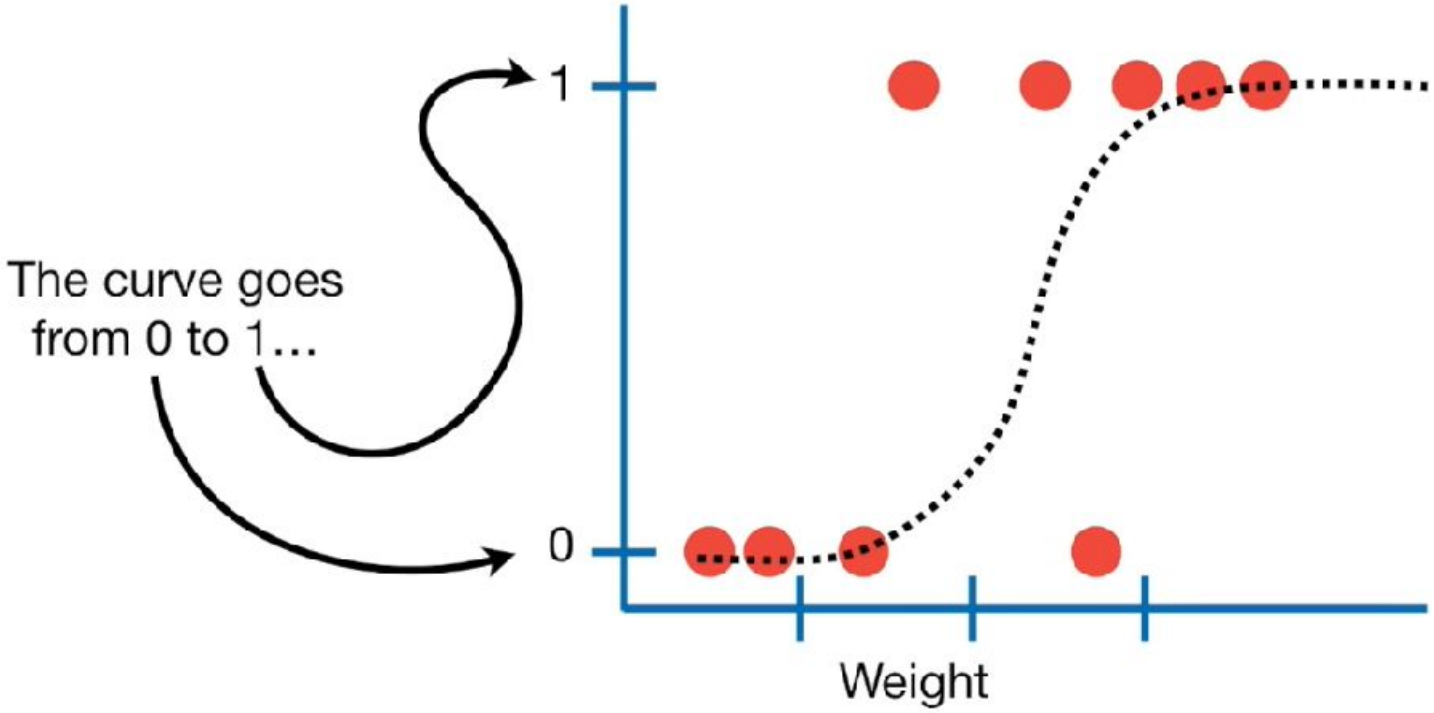
Logistic Regression



Logistic Regression

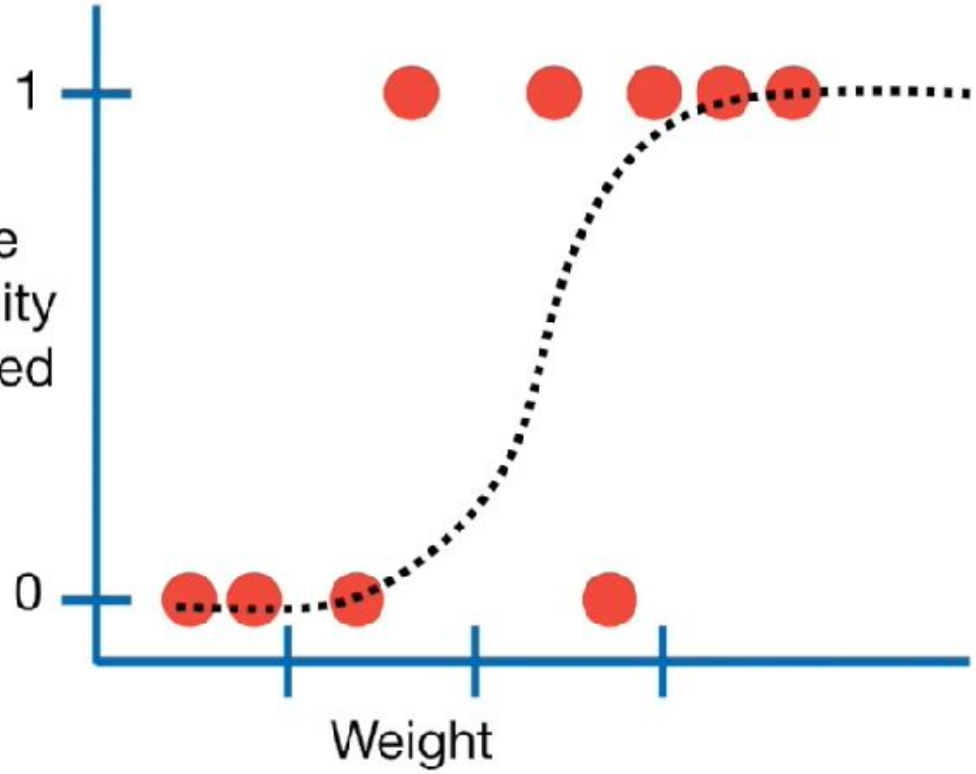


Logistic Regression

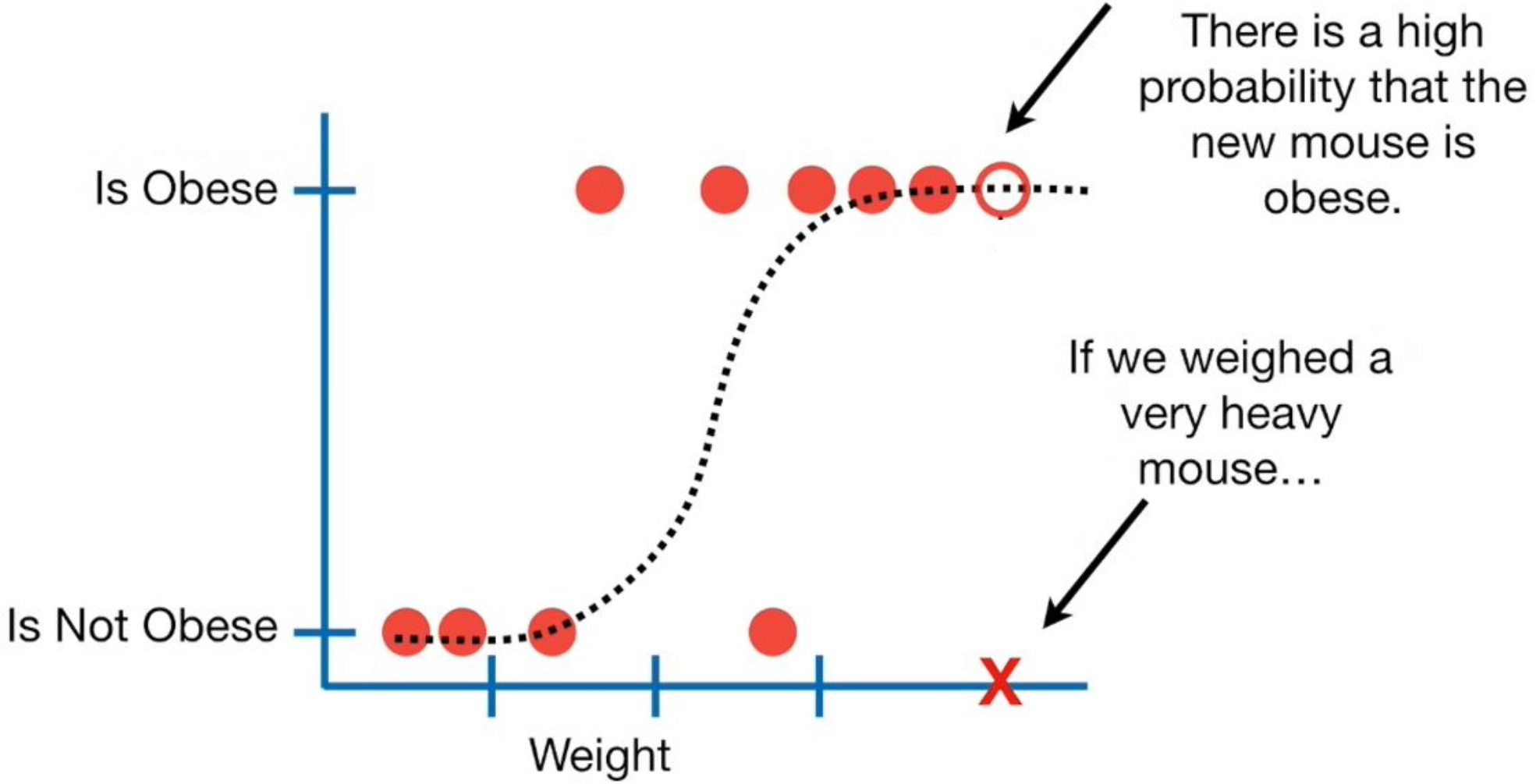


Logistic Regression

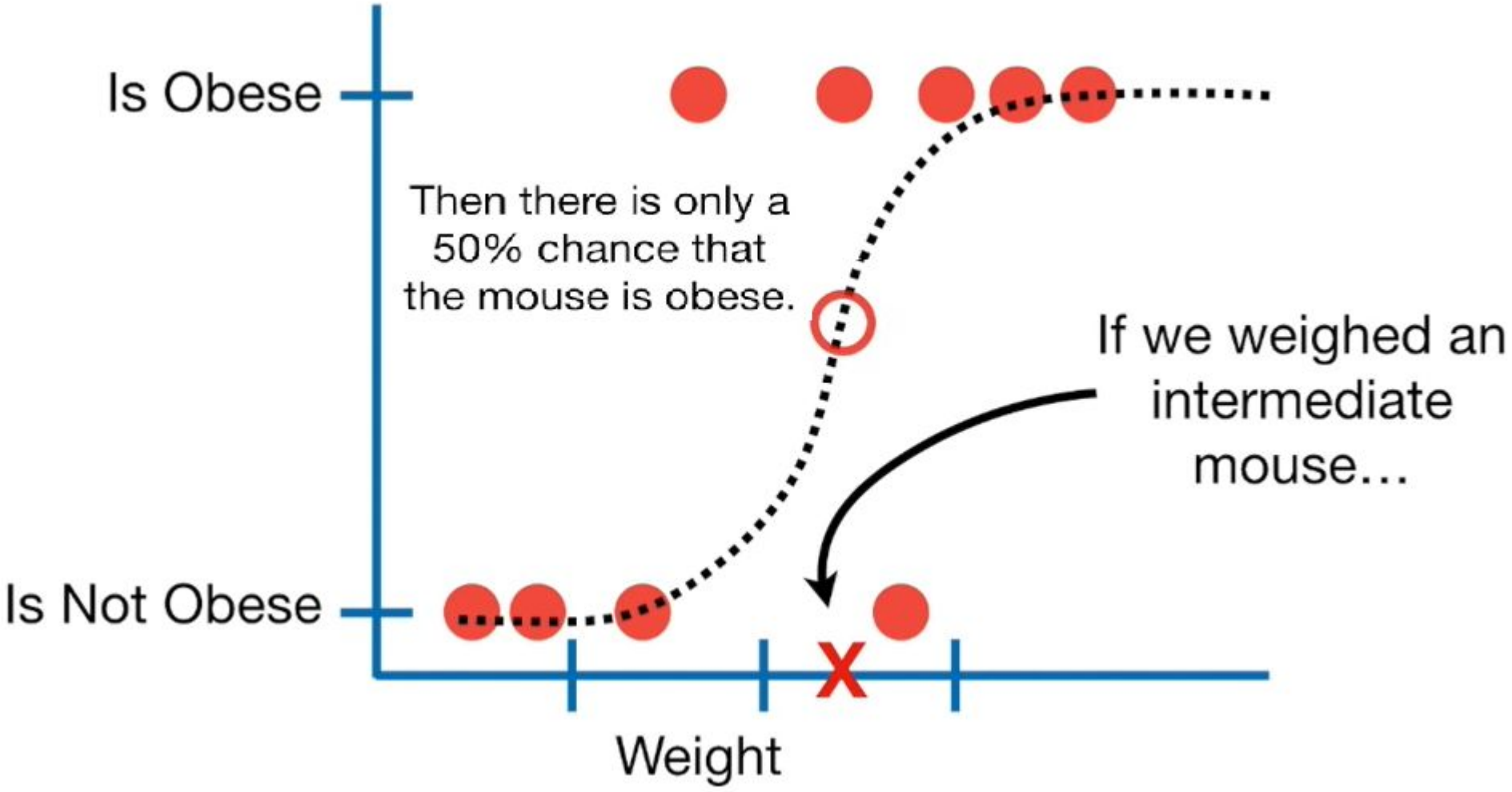
...and that means that the curve tells you the probability that a mouse is ***obese*** based on its ***weight***.



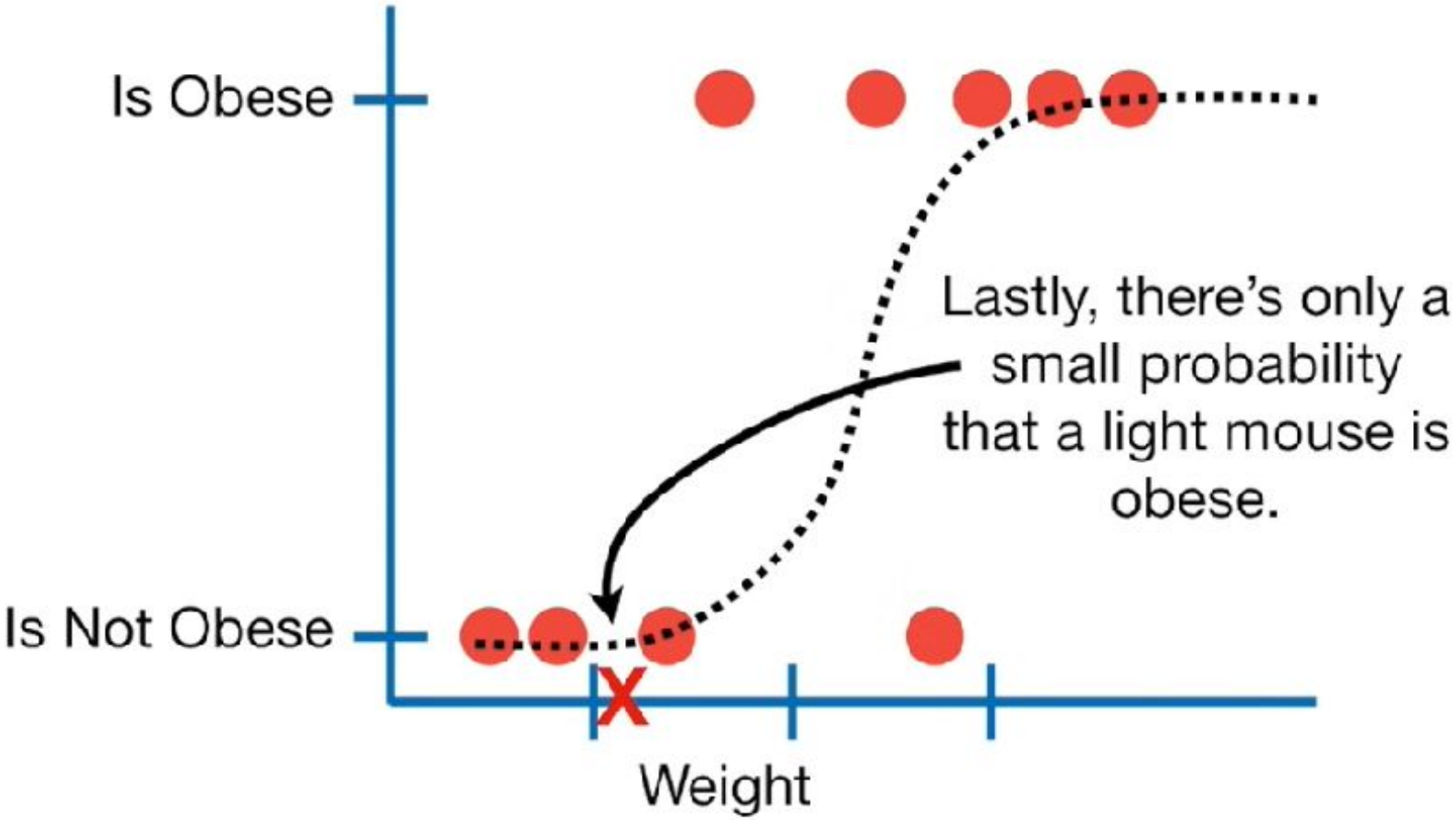
Logistic Regression



Logistic Regression

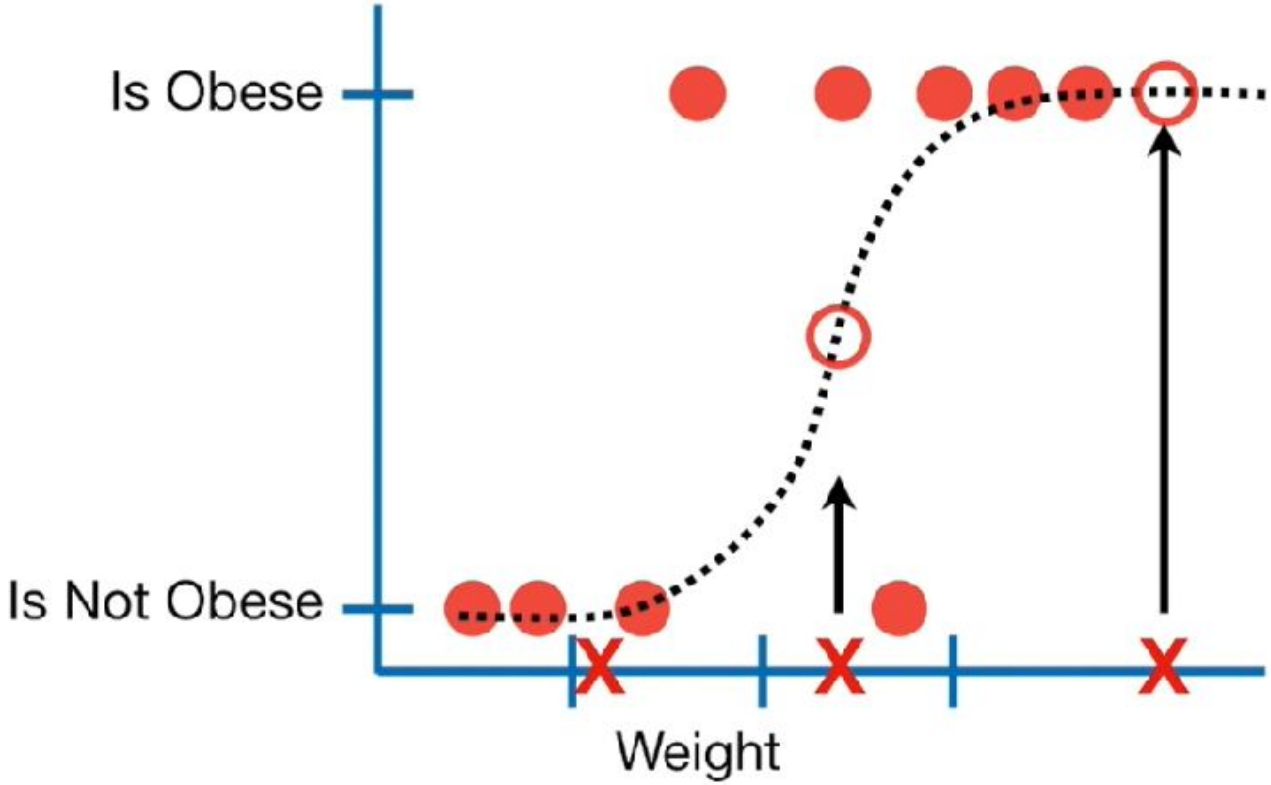


Logistic Regression



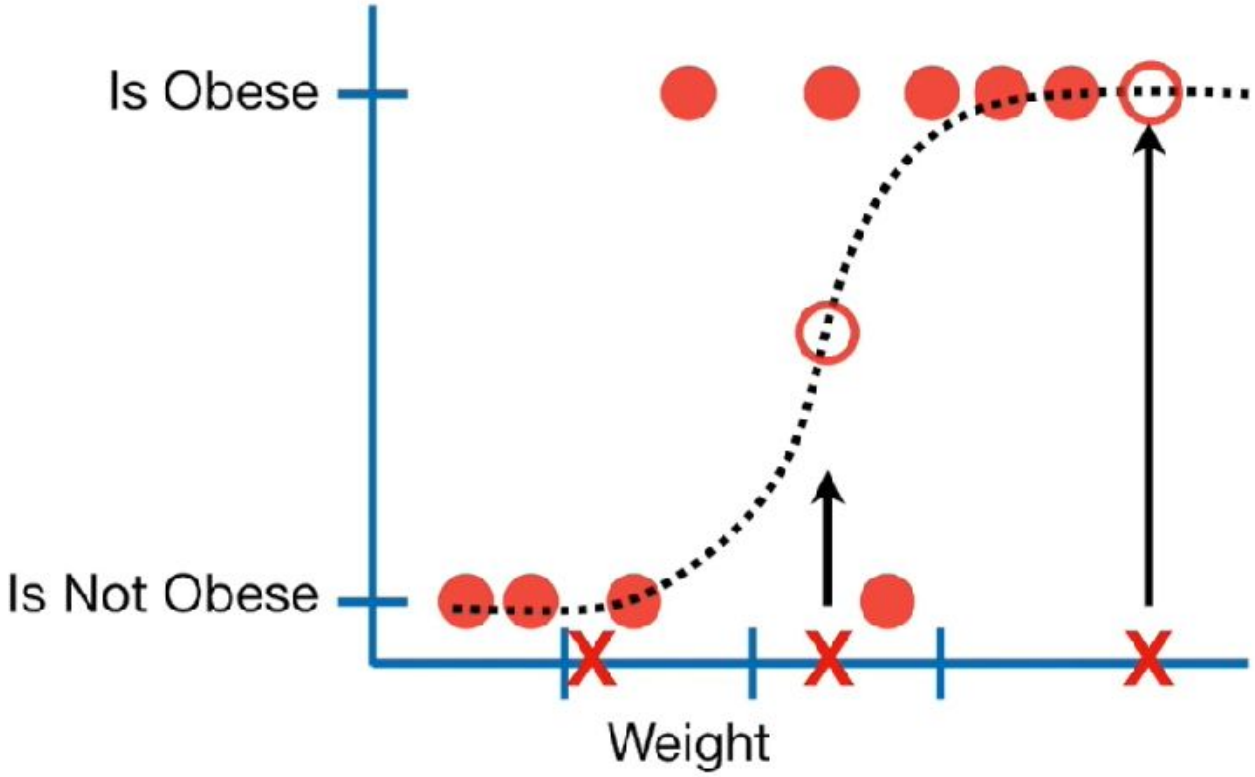
Logistic Regression

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



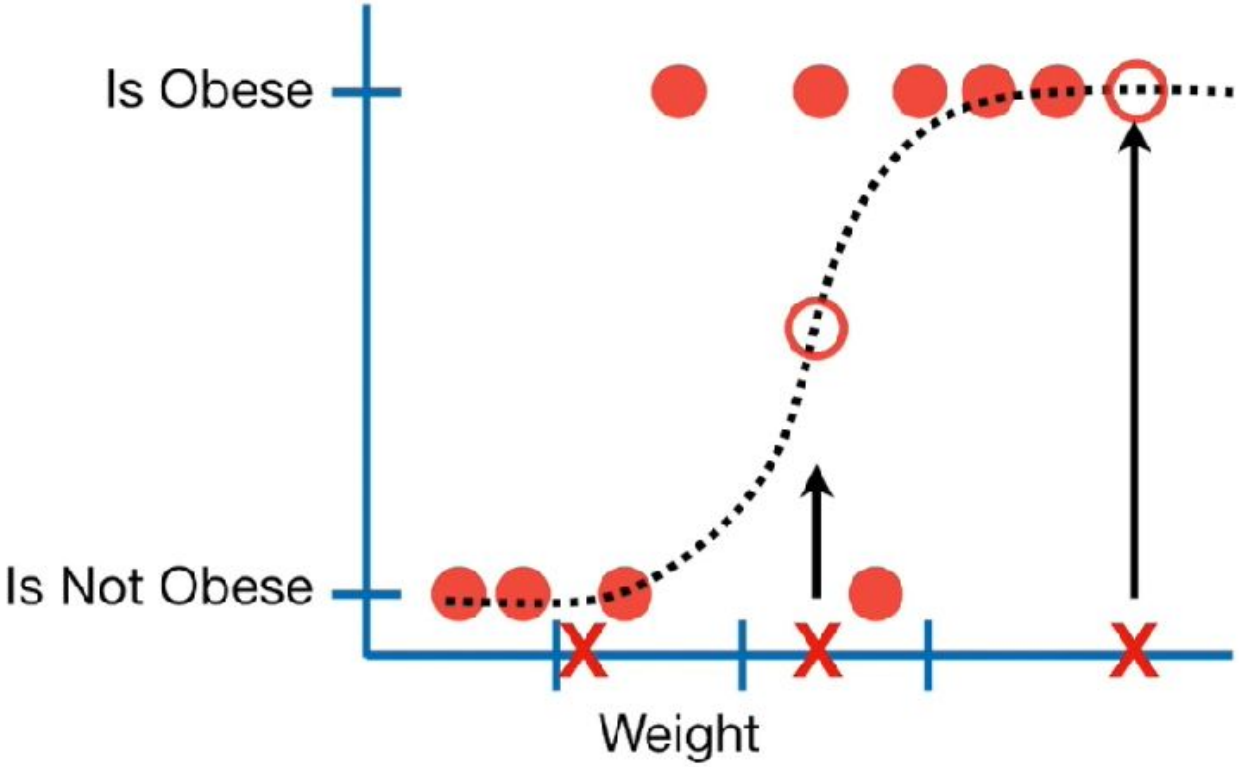
Logistic Regression

For example, if the probability a mouse is obese is > 50%, then we'll classify it as obese, otherwise we'll classify it as "not obese".

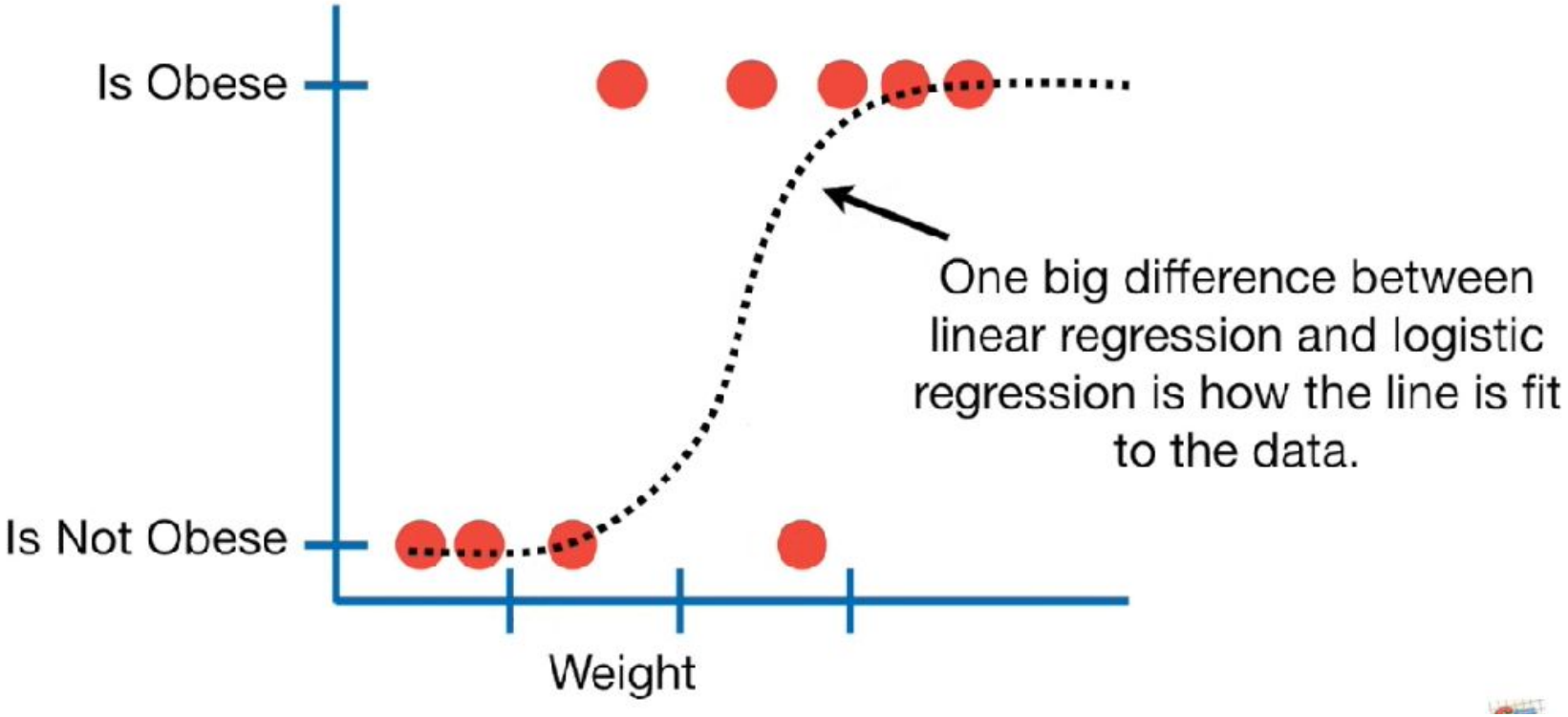


Logistic Regression

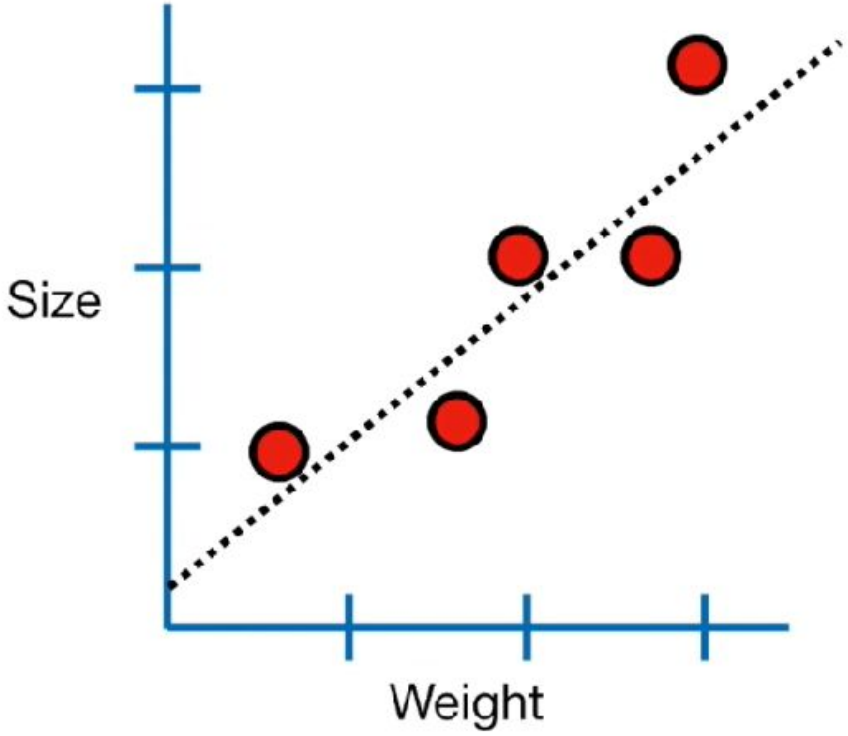
Logistic regression's ability to provide probabilities and classify new samples using continuous and discrete measurements makes it a popular machine learning method.



Logistic Regression

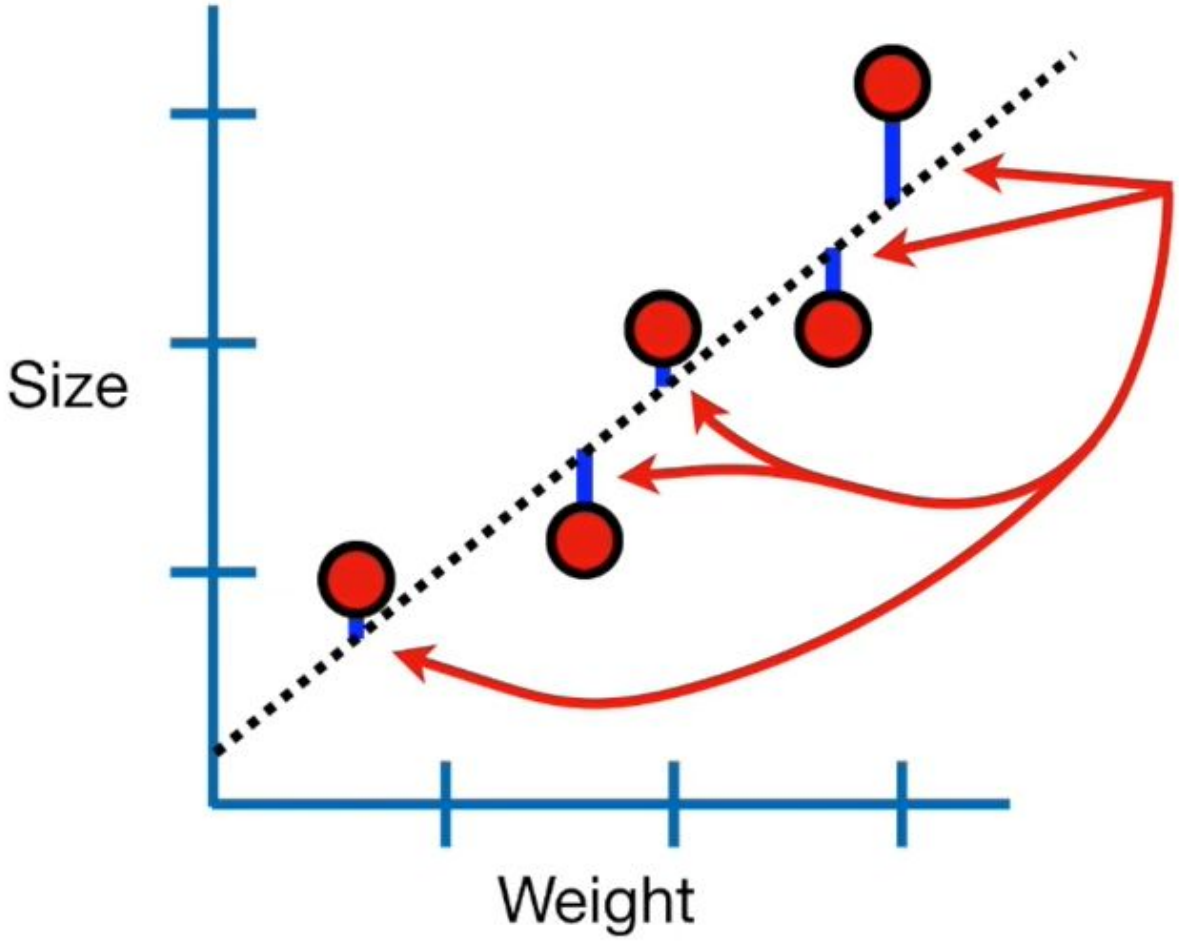


Logistic Regression



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

Logistic Regression

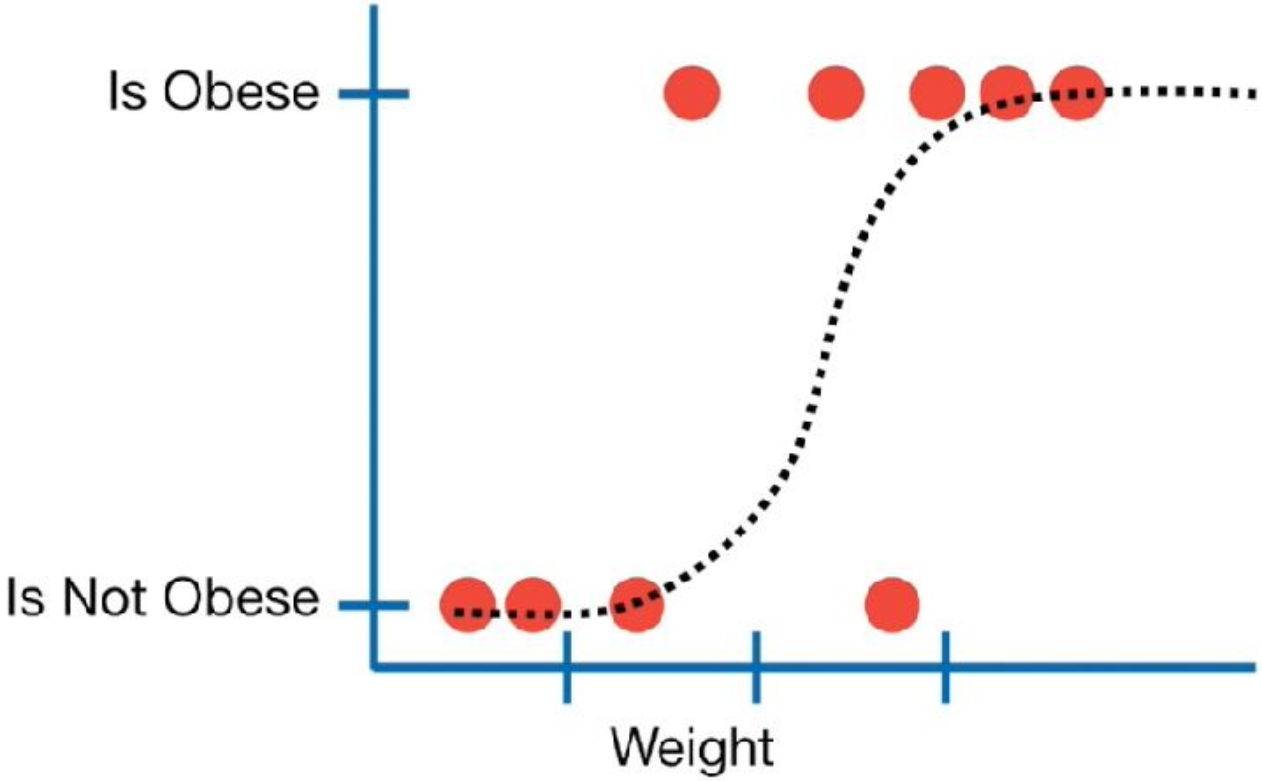


In other words, we find the line that minimizes the sum of the squares of these residuals.

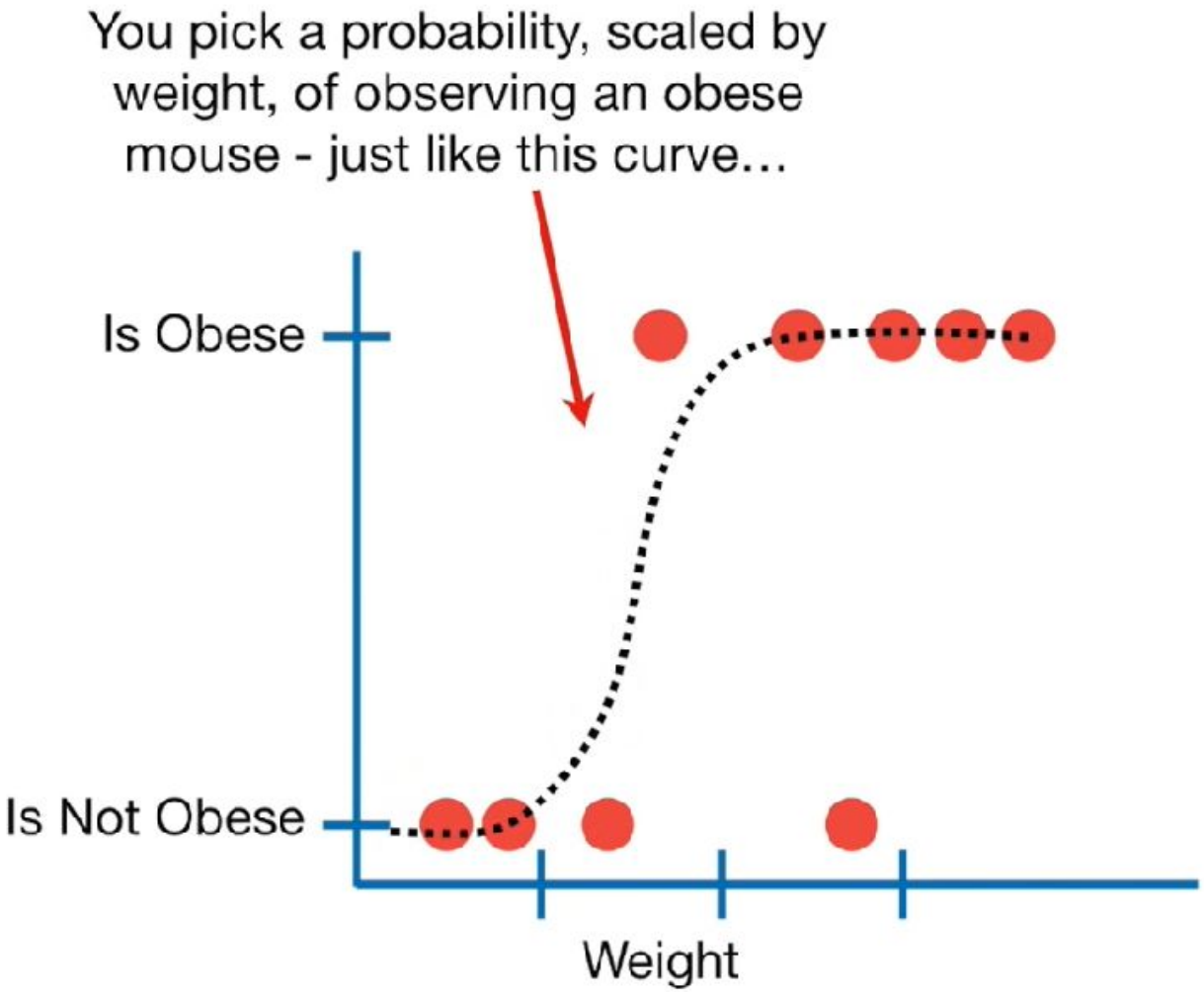
We also use the residuals to calculate R^2 and to compare simple models to complicated models.

Logistic Regression

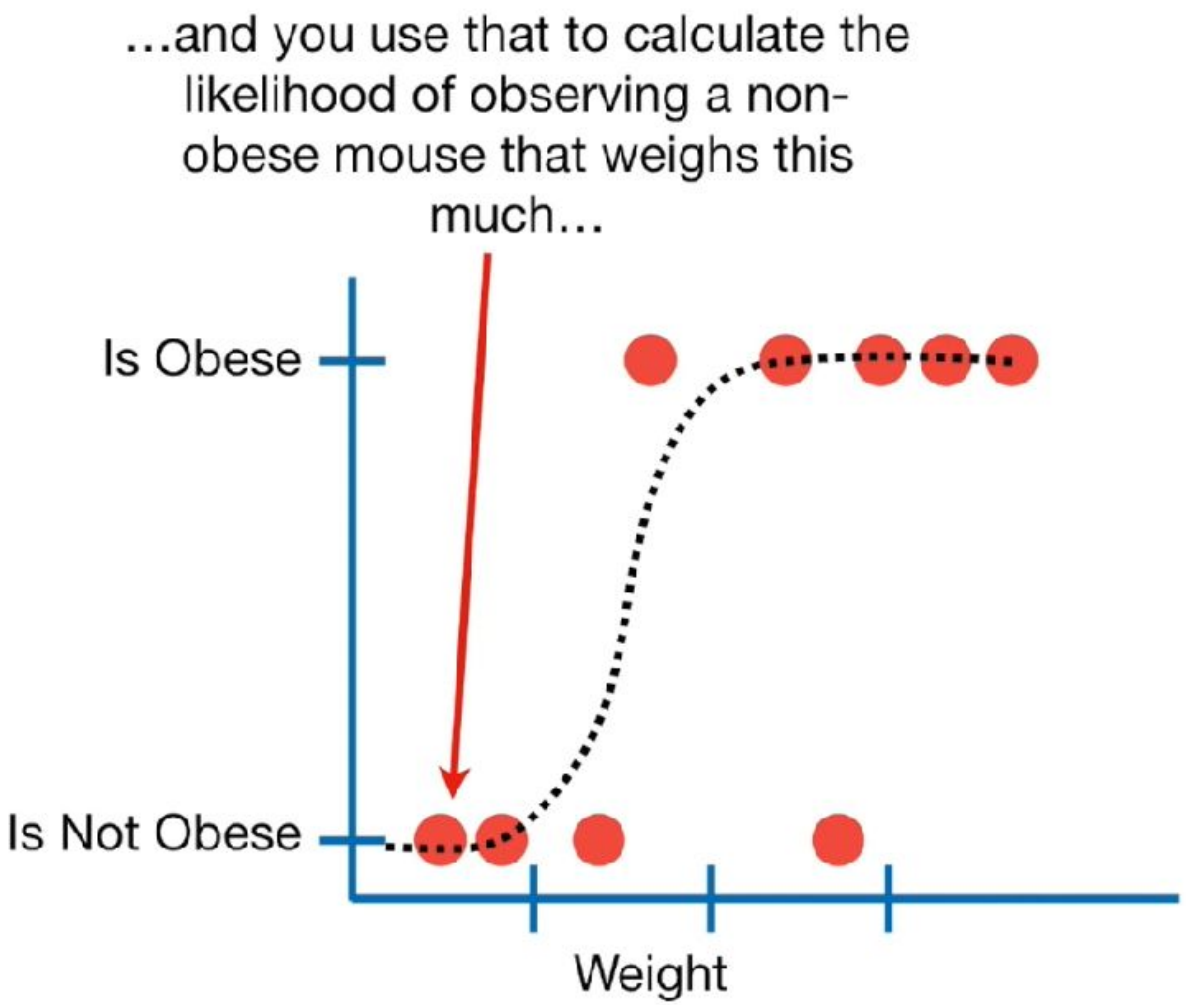
Logistic regression doesn't have the same concept of a "residual", so it can't use least squares and it can't calculate R^2.



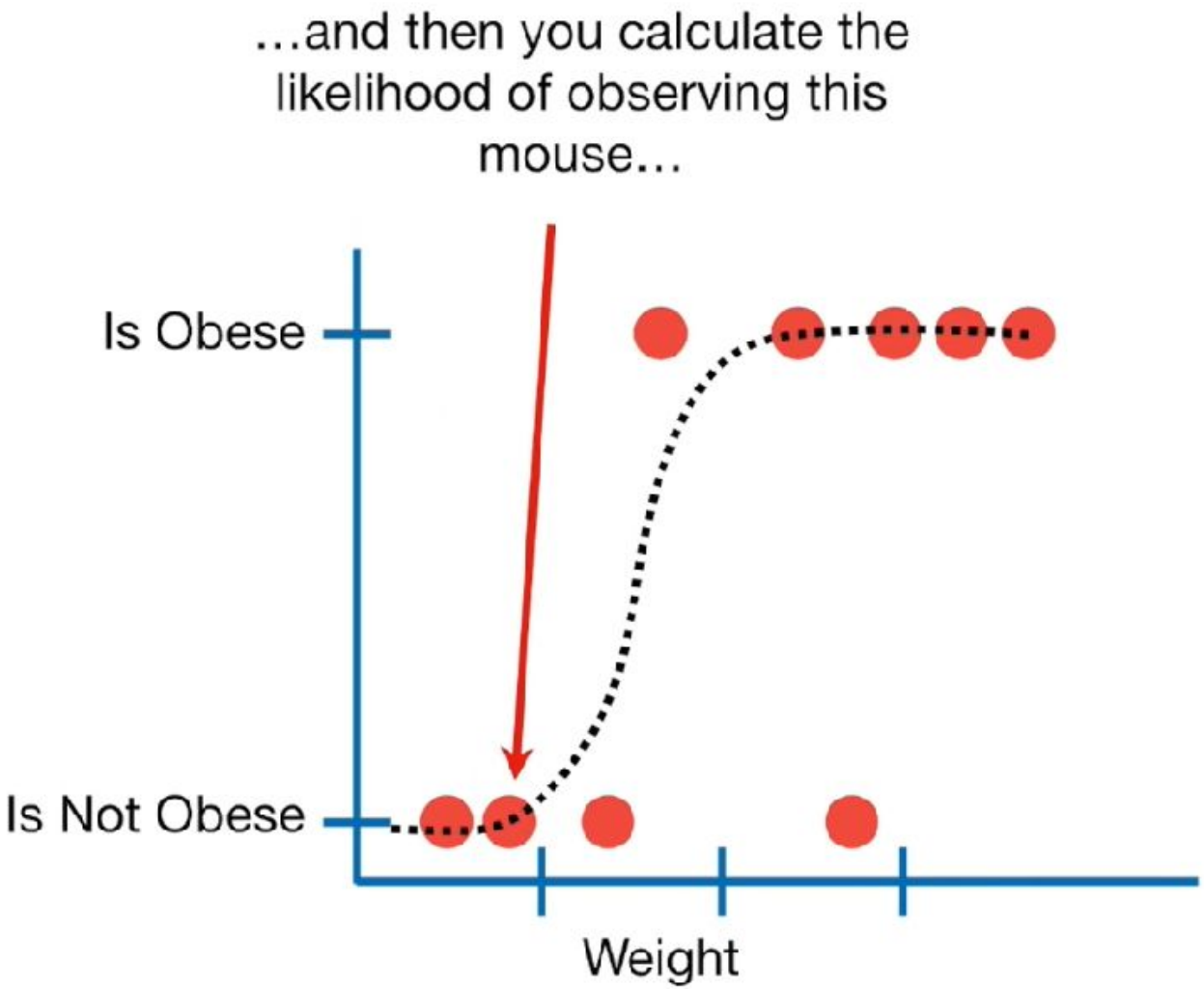
Logistic Regression



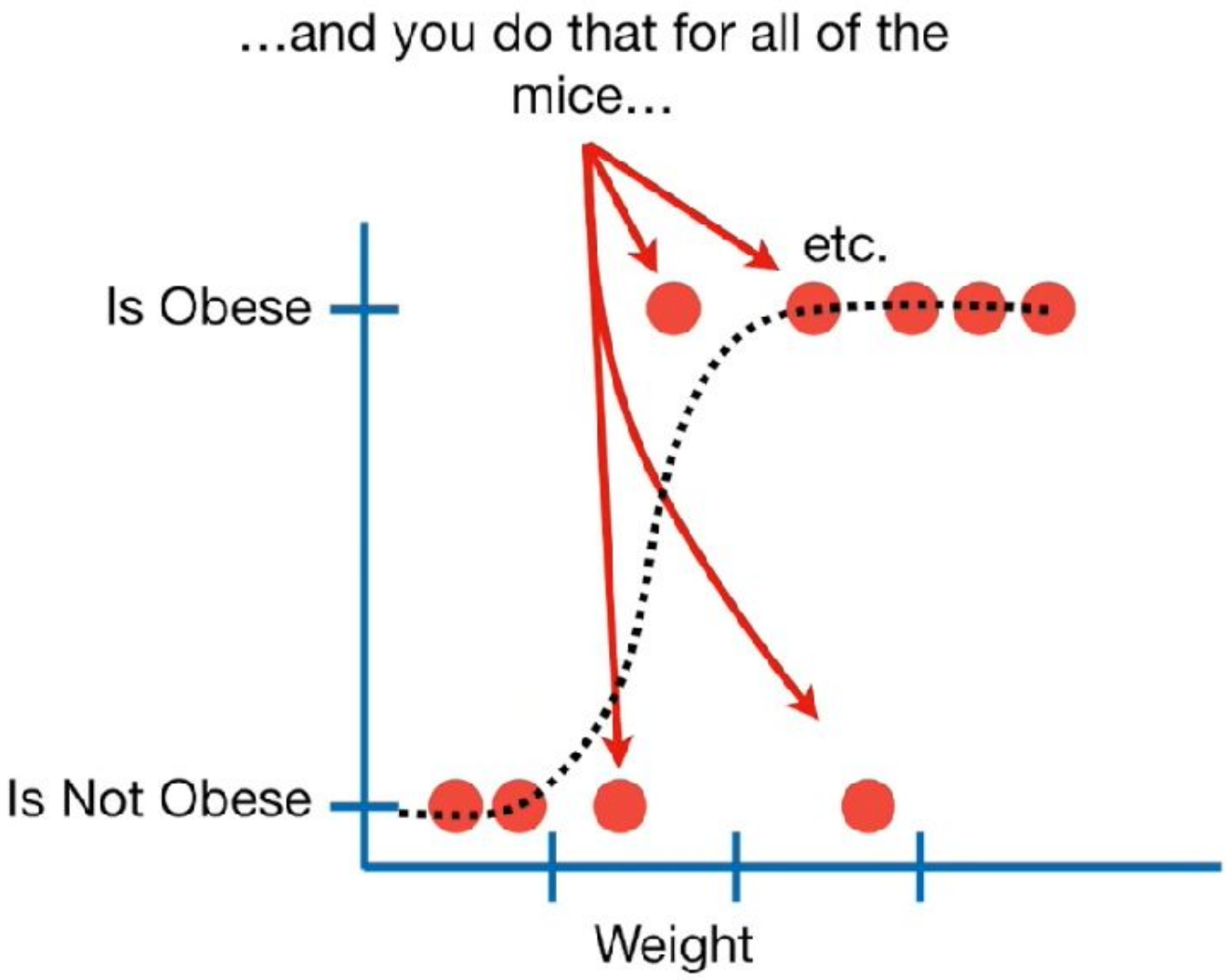
Logistic Regression



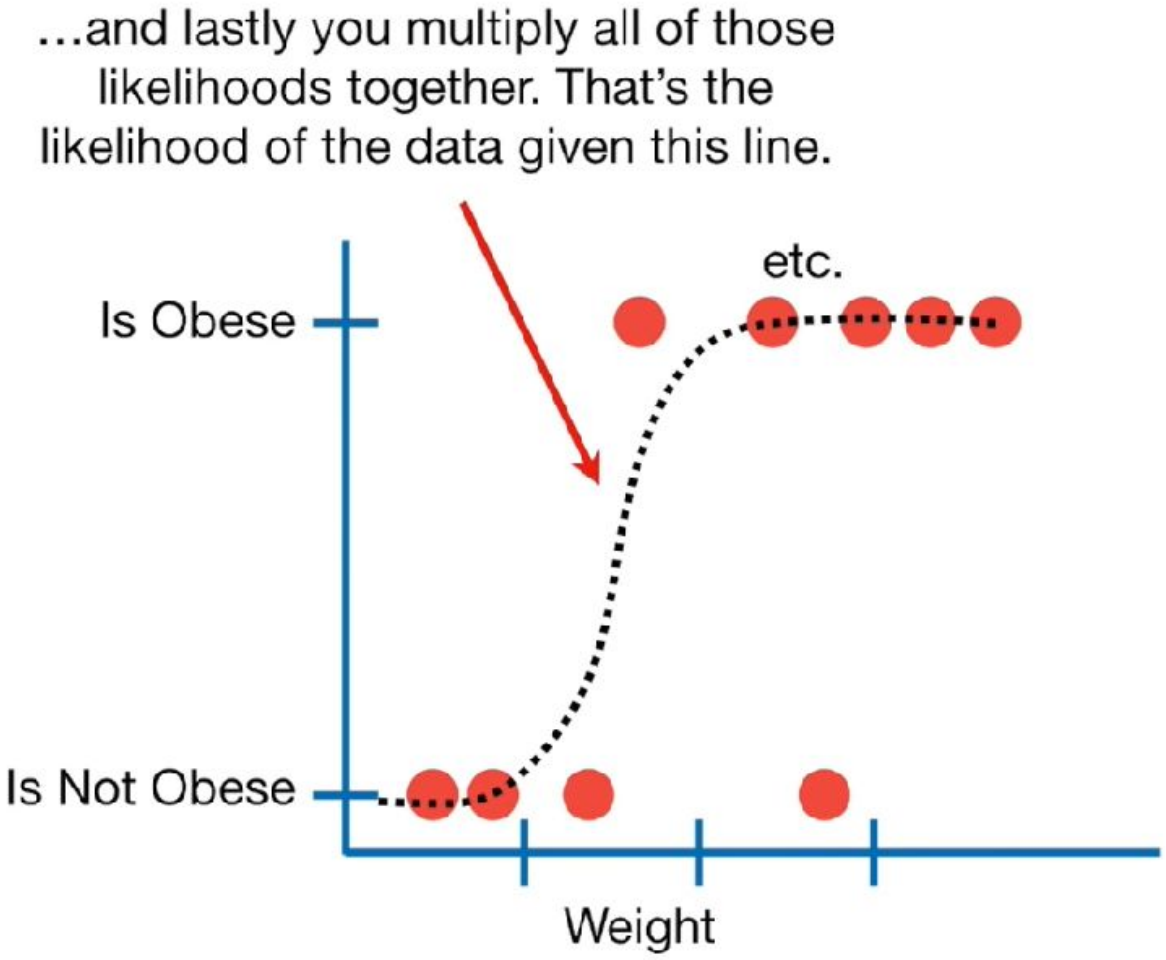
Logistic Regression



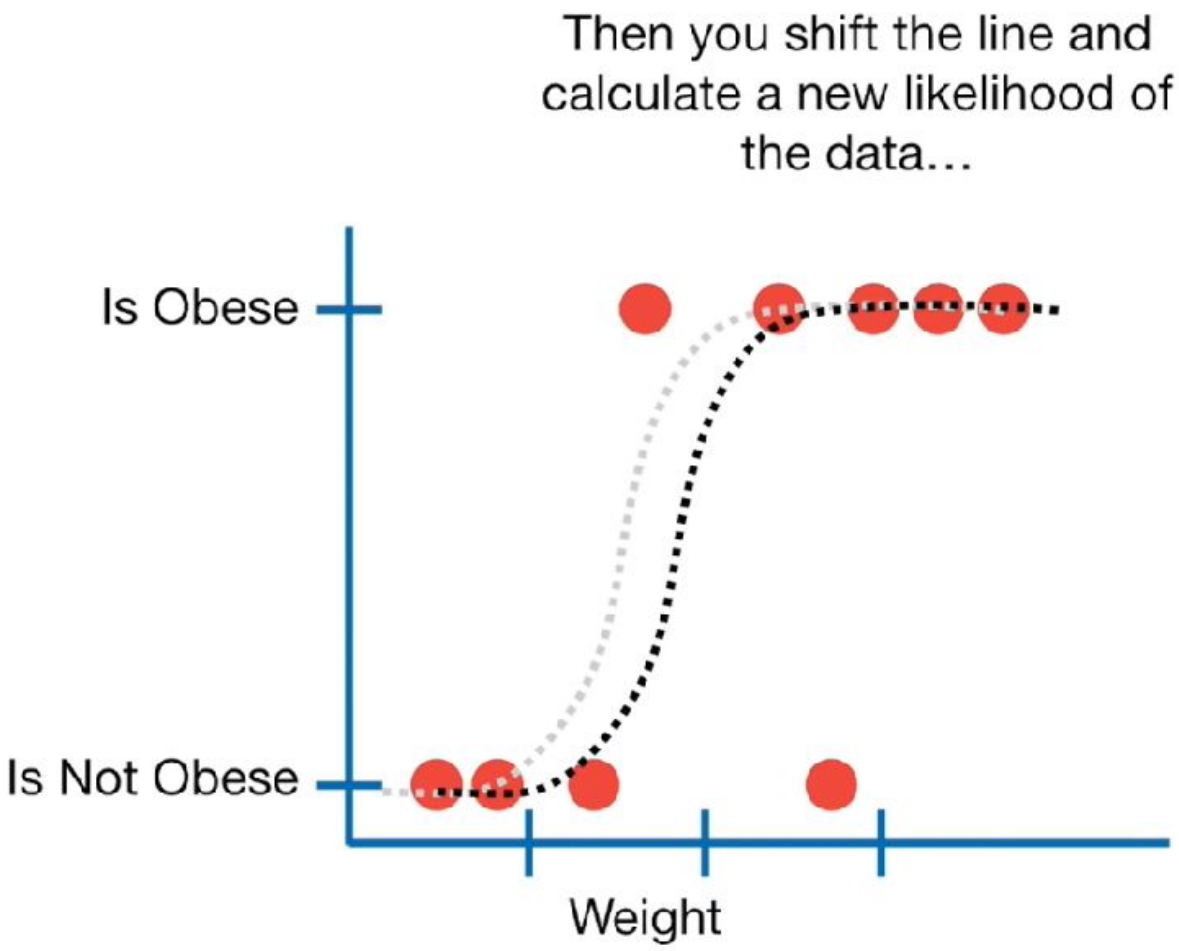
Logistic Regression



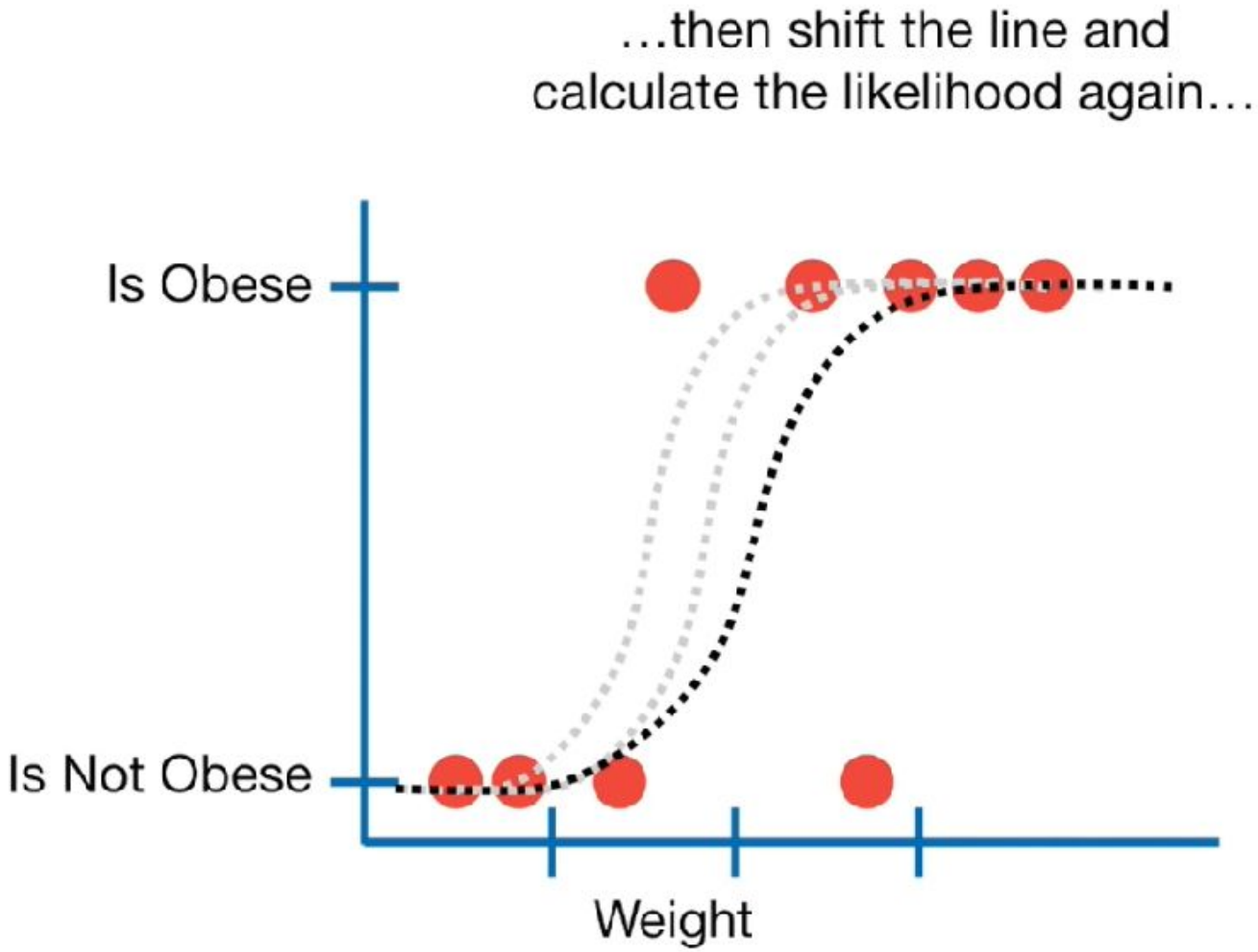
Logistic Regression



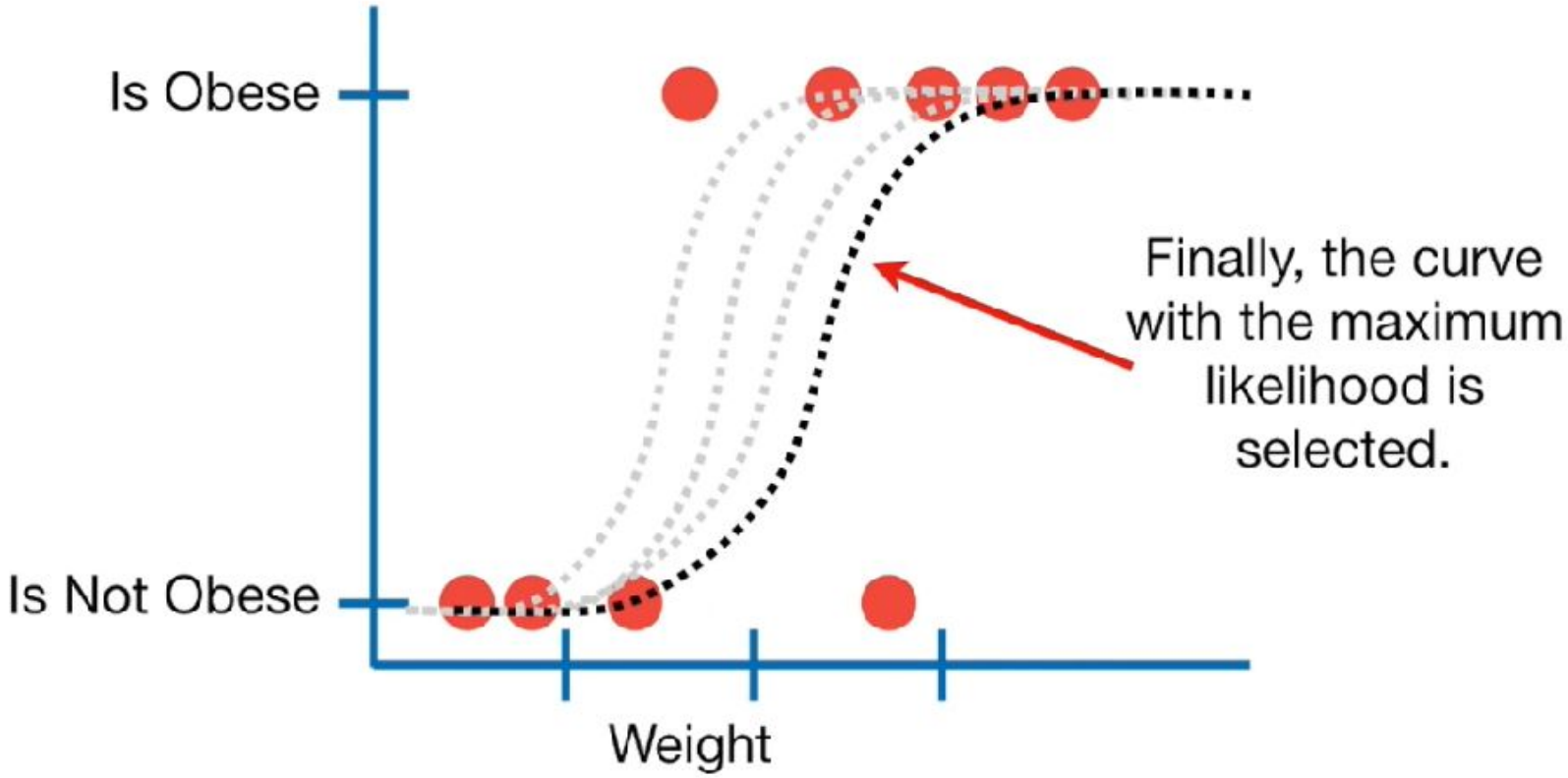
Logistic Regression



Logistic Regression

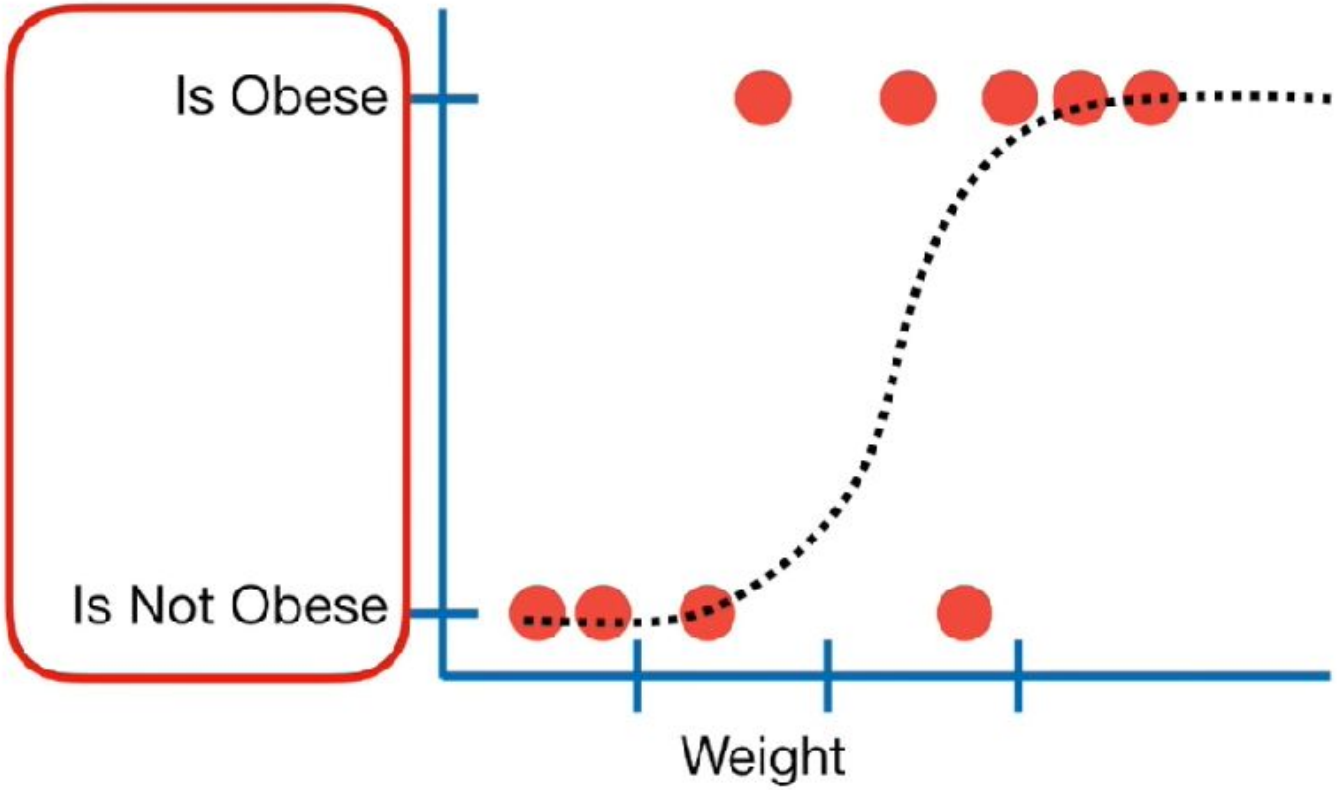


Logistic Regression

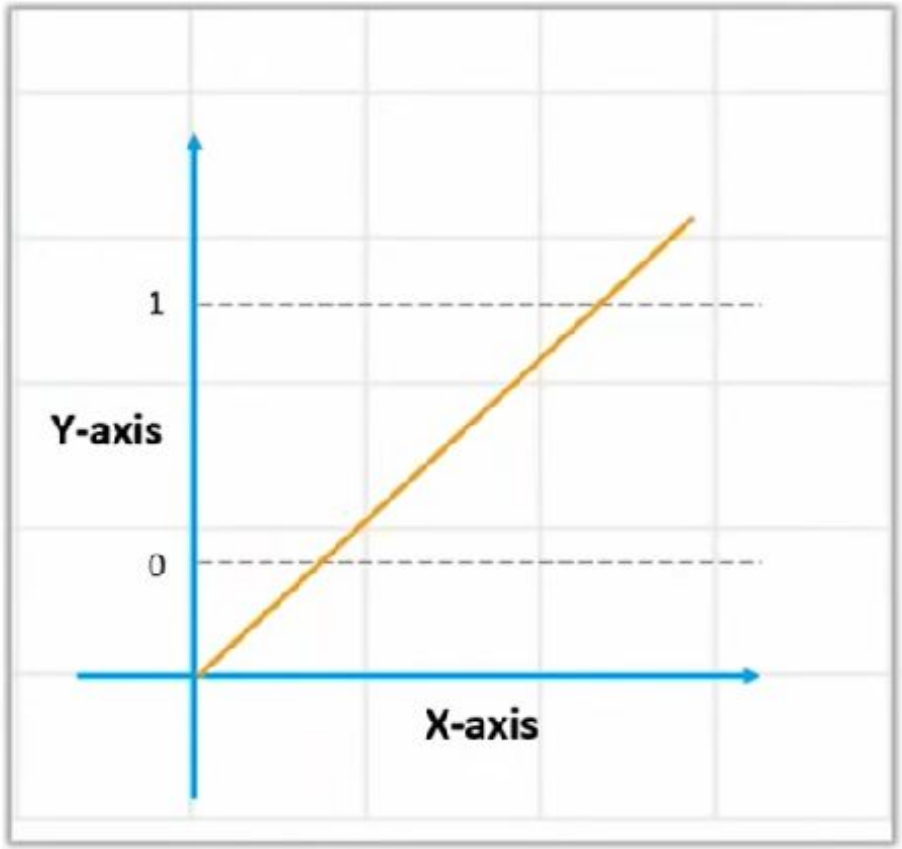


Logistic Regression

In summary… Logistic regression can be used to classify samples…

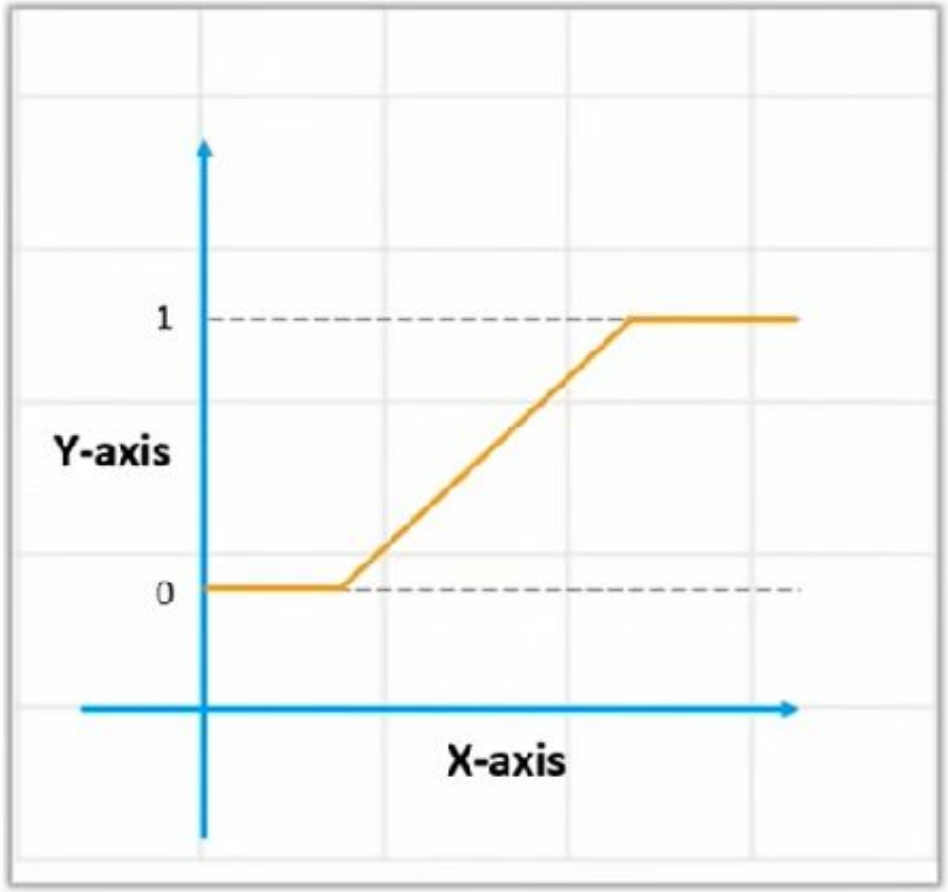


Logistic Regression



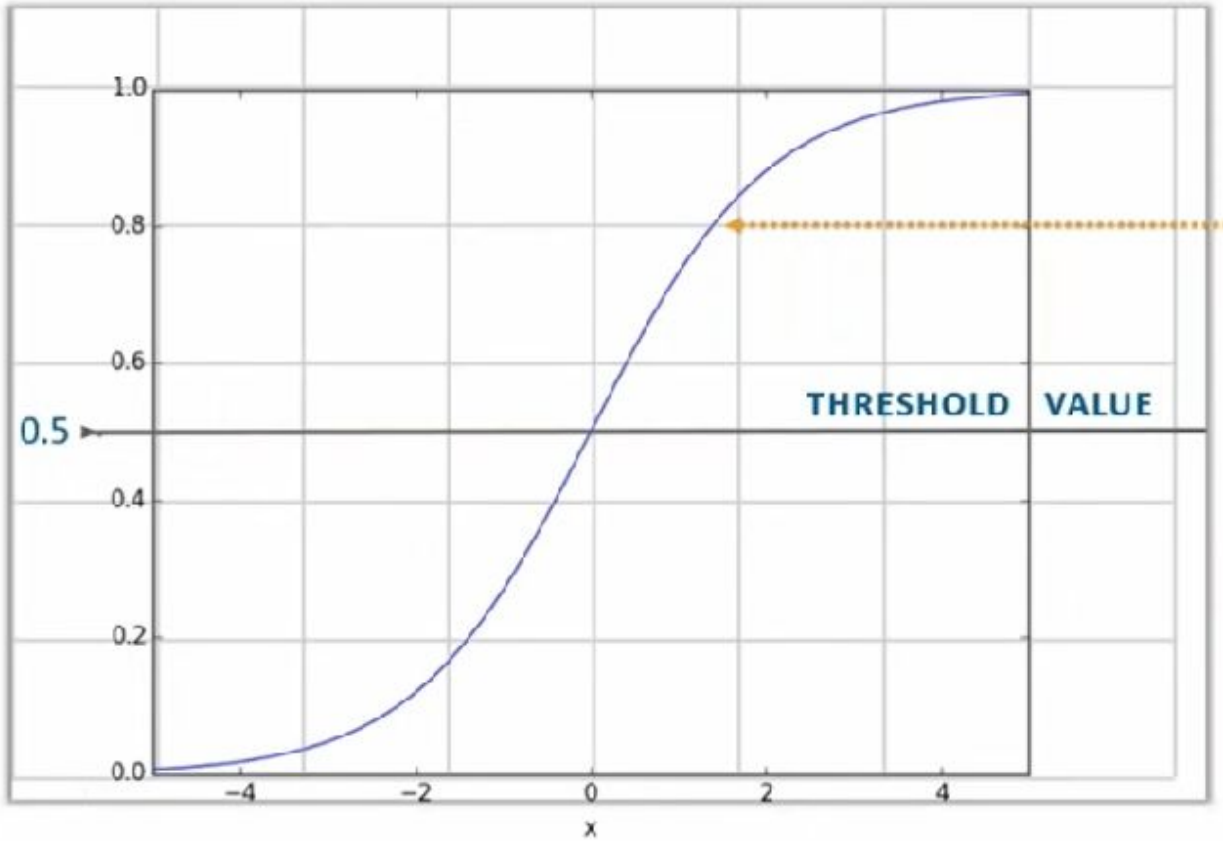
Since value of Y will be between 0 and 1, the linear line has to clipped at 0 and 1

Logistic Regression



With this, our resulting curve cannot be formulated into a single formula. Hence we came up with Logistic regression.

Logistic Regression



The Sigmoid “S” Curve

With this, the threshold value indicates the probability of wining or losing.

Logistic Regression

The Logistic Regression Equation is derived from the Straight Line Equation

Equation of a straight line

$Y = C + BIX1 + B2X2 + ...$ ——— Range is from –(infinity) to (infinity)

Let's try to reduce the Logistic Regression Equation from Straight Line Equation

$Y = C + BIX1 + B2X2 + ...$ ——— In Logistic equation Y can be only from 0 to 1

Now , to get the range of Y between 0 and infinity, let's transform Y

$\frac{Y}{1-Y}$ } Y= 0 then 0, Y= 1 then infinity ——— Now, the range is between 0 to infinity

Logistic Regression

The Logistic Regression Equation is derived from the Straight Line Equation

Equation of a straight line

$Y = C + BIX1 + B2X2 + ...$ ——— Range is from –(infinity) to (infinity)

Let's try to reduce the Logistic Regression Equation from Straight Line Equation

$Y = C + BIX1 + B2X2 + ...$ ——— In Logistic equation Y can be only from 0 to 1

Now , to get the range of Y between 0 and infinity, let's transform Y

Y	Y= 0 then 0
1-Y	Y= 1 then infinity

——— Now, the range is between 0 to infinity

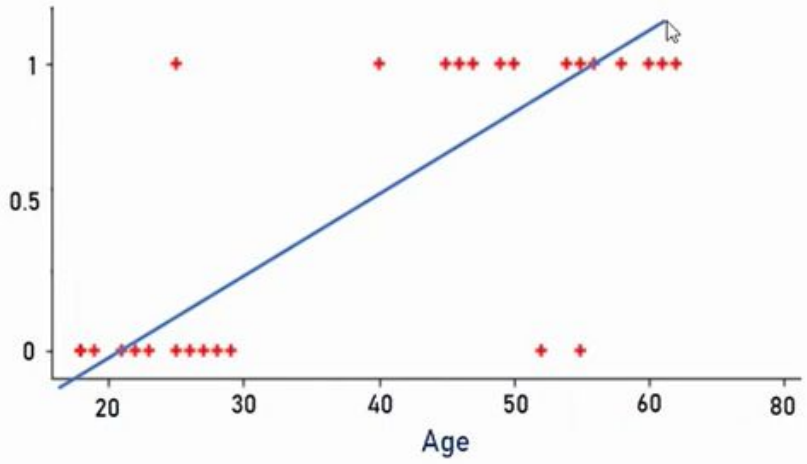
Let us transform it further, to get range between –(infinity) and (infinity)

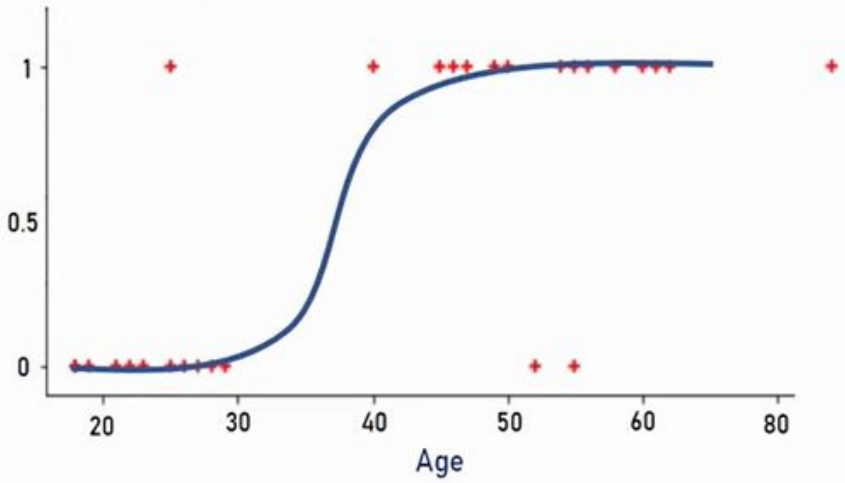
$\log\left[\frac{Y}{1-Y}\right] \Rightarrow Y = C + BIX1 + B2X2 + ...$ ——— Final Logistic Regression Equation

Logistic Regression

$$y = m * x + b$$

$$y = \frac{1}{1 + e^{-(m*x+b)}}$$





Logistic Regression

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
No	No	No	125	No
Yes	Yes	Yes	180	Yes
Yes	Yes	No	210	No
...	...	...	...	...

Imagine that we have this medical data…

Logistic Regression

We've got some clinical measurements…

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
No	No	No	125	No
Yes	Yes	Yes	180	Yes
Yes	Yes	No	210	No
…	…	…	…	…

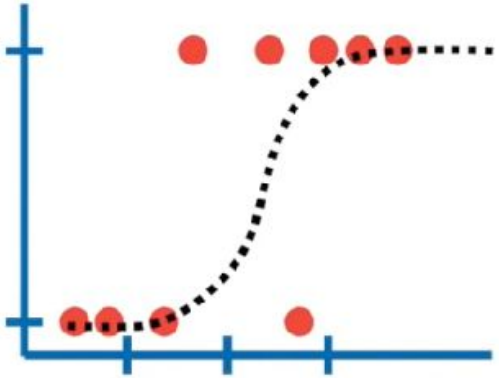
Logistic Regression

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
No	No	No	125	No
Yes	Yes	Yes	180	Yes
Yes	Yes	No	210	No
...	...	...	...	...

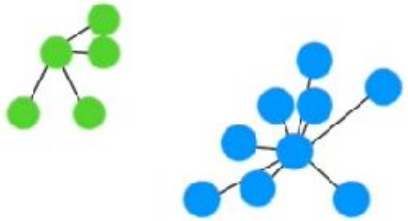
...and we want to apply a machine learning method to them to predict whether or not someone will develop heart disease.

Logistic Regression

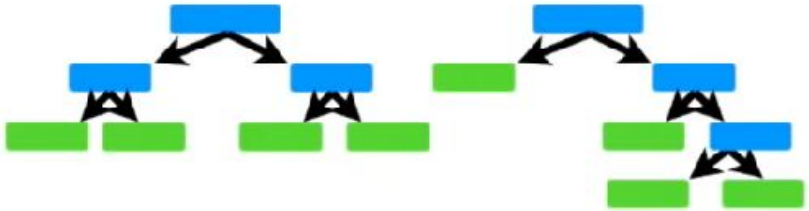
To do this, we could use **Logistic Regression**...



...or **K-Nearest Neighbors**...



...or a **Random Forest**...



...or some other method.
There are tons to choose from.

How do we decide which one works best with our data?

Logistic Regression

We start by dividing the data into **Training** and **Testing** sets…

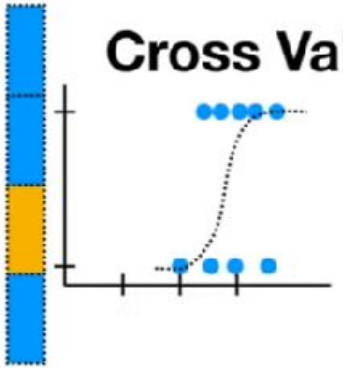
Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
No	No	No	125	No
…				…

Training Data

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
Yes	Yes	No	210	No
…	…	…	…	…

Testing Data

NOTE: This would be an excellent opportunity to use **Cross Validation**.



Cross Validation….

…it's no big deal!!!

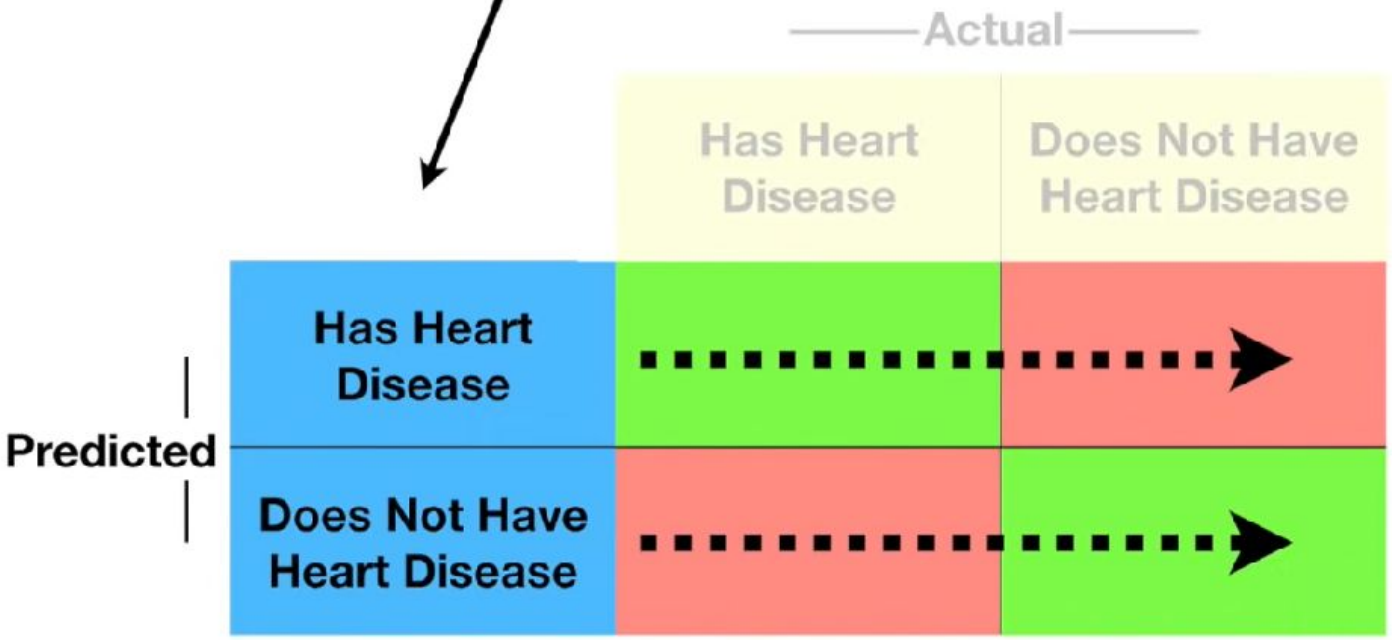
Logistic Regression

One way to do this is by creating a **Confusion Matrix** for each method.

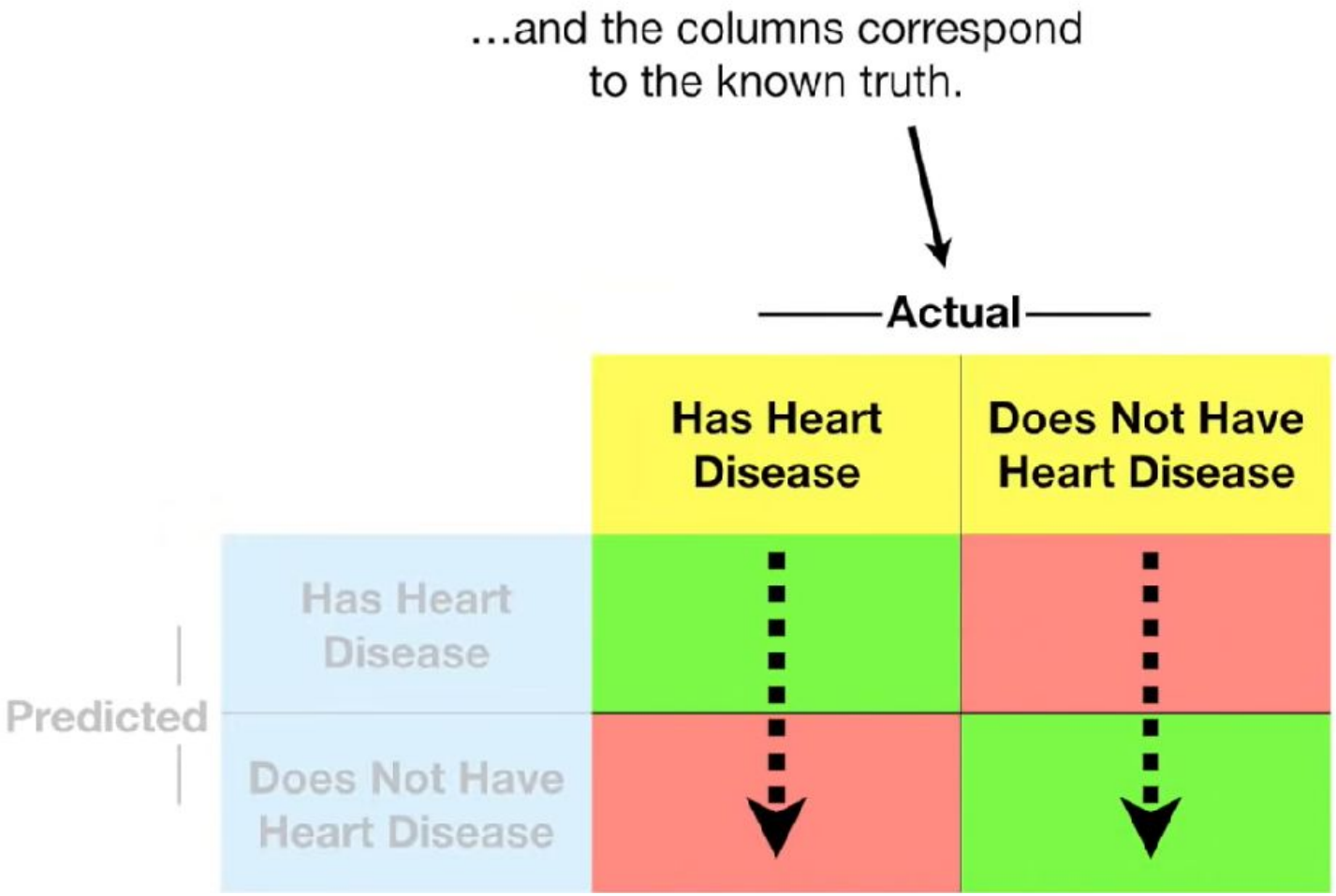
		Actual	
		Has Heart Disease	Does Not Have Heart Disease
Predicted	Has Heart Disease		
	Does Not Have Heart Disease		

Logistic Regression

The rows in a **Confusion Matrix** correspond to what the machine learning algorithm predicted…



Logistic Regression



Logistic Regression

Since there are only two categories to choose from:

		Actual	
		Has Heart Disease	Does Not Have Heart Disease
Predicted	Has Heart Disease		
	Does Not Have Heart Disease		

Logistic Regression

These are the patients that *had heart disease* that were correctly identified by the algorithm.

		Actual	
		Has Heart Disease	Does Not Have Heart Disease
Predicted	Has Heart Disease	True Positives	
	Does Not Have Heart Disease		

Logistic Regression

The **True Negatives** are in the bottom right-hand corner.

		Actual	
		Has Heart Disease	Does Not Have Heart Disease
Predicted	Has Heart Disease	True Positives	
	Does Not Have Heart Disease		True Negatives

Logistic Regression

These are the patients that *did not have heart disease* that were correctly identified by the algorithm.

		Actual	
		Has Heart Disease	Does Not Have Heart Disease
Predicted	Has Heart Disease	True Positives	
	Does Not Have Heart Disease		True Negatives

Logistic Regression

False Negatives are when a patient has heart disease, but the algorithm said they didn't.

		Actual	
		Has Heart Disease	Does Not Have Heart Disease
Predicted	Has Heart Disease	True Positives	
	Does Not Have Heart Disease	False Negatives	True Negatives

Logistic Regression

False Positives are patients that do not have heart disease, but the algorithm says they do.

		Actual	
		Has Heart Disease	Does Not Have Heart Disease
Predicted	Has Heart Disease	True Positives	False Positives
	Does Not Have Heart Disease	False Negatives	True Negatives

Logistic Regression

The numbers along the diagonal (the **Green Boxes**) tell us how many times the samples were correctly classified.

		Actual	
		Has Heart Disease	Does Not Have Heart Disease
Predicted	Has Heart Disease	142	22
	Does Not Have Heart Disease	29	110

Logistic Regression

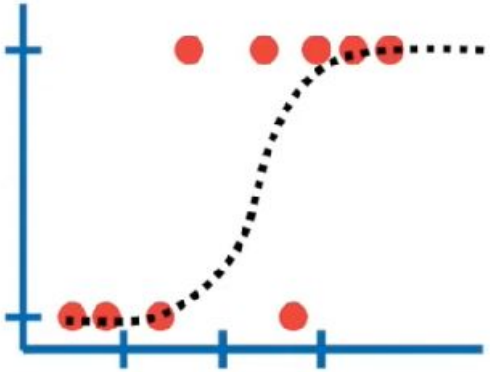
The numbers *not* on the diagonal (the **Red Boxes**) are samples the algorithm messed up.

		Actual	
		Has Heart Disease	Does Not Have Heart Disease
Predicted	Has Heart Disease	142	22
	Does Not Have Heart Disease	29	110

Logistic Regression

	Has Heart Disease	Does Not Have Heart Disease
Has Heart Disease	142	22
Does Not Have Heart Disease	29	110

Lastly, we can apply **Logistic Regression** to the **Testing Dataset** and create a **Confusion Matrix**.



	Has Heart Disease	Does Not Have Heart Disease
Has Heart Disease	139	20
Does Not Have Heart Disease	32	112

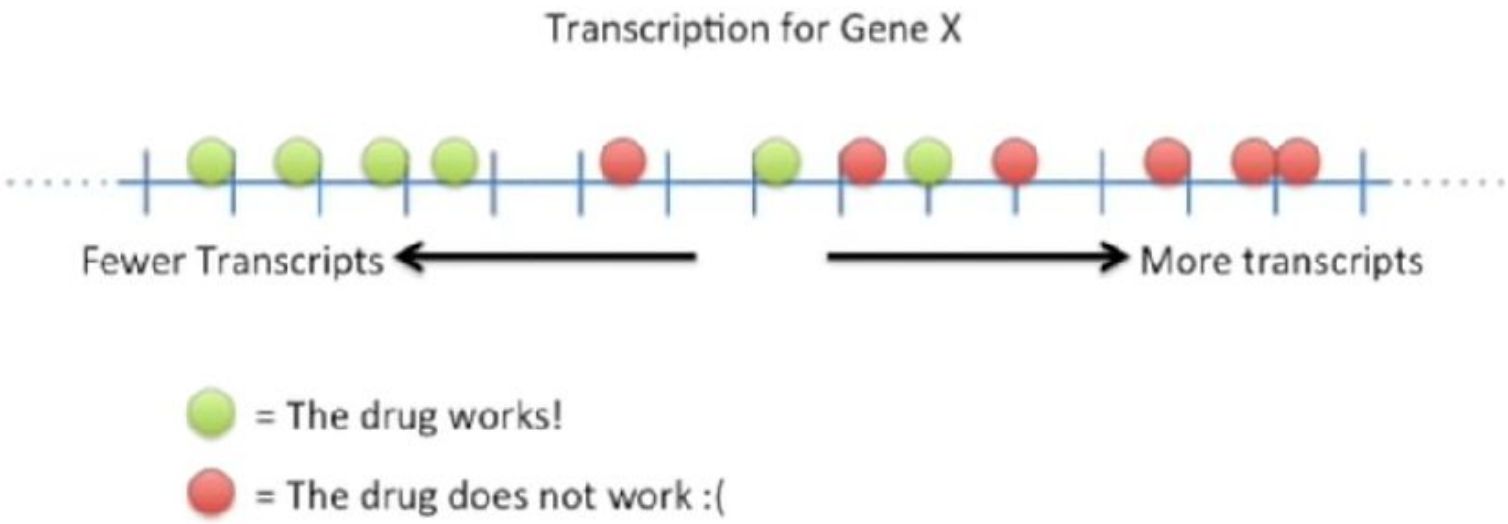
Linear Discriminant Analysis

The Problem…

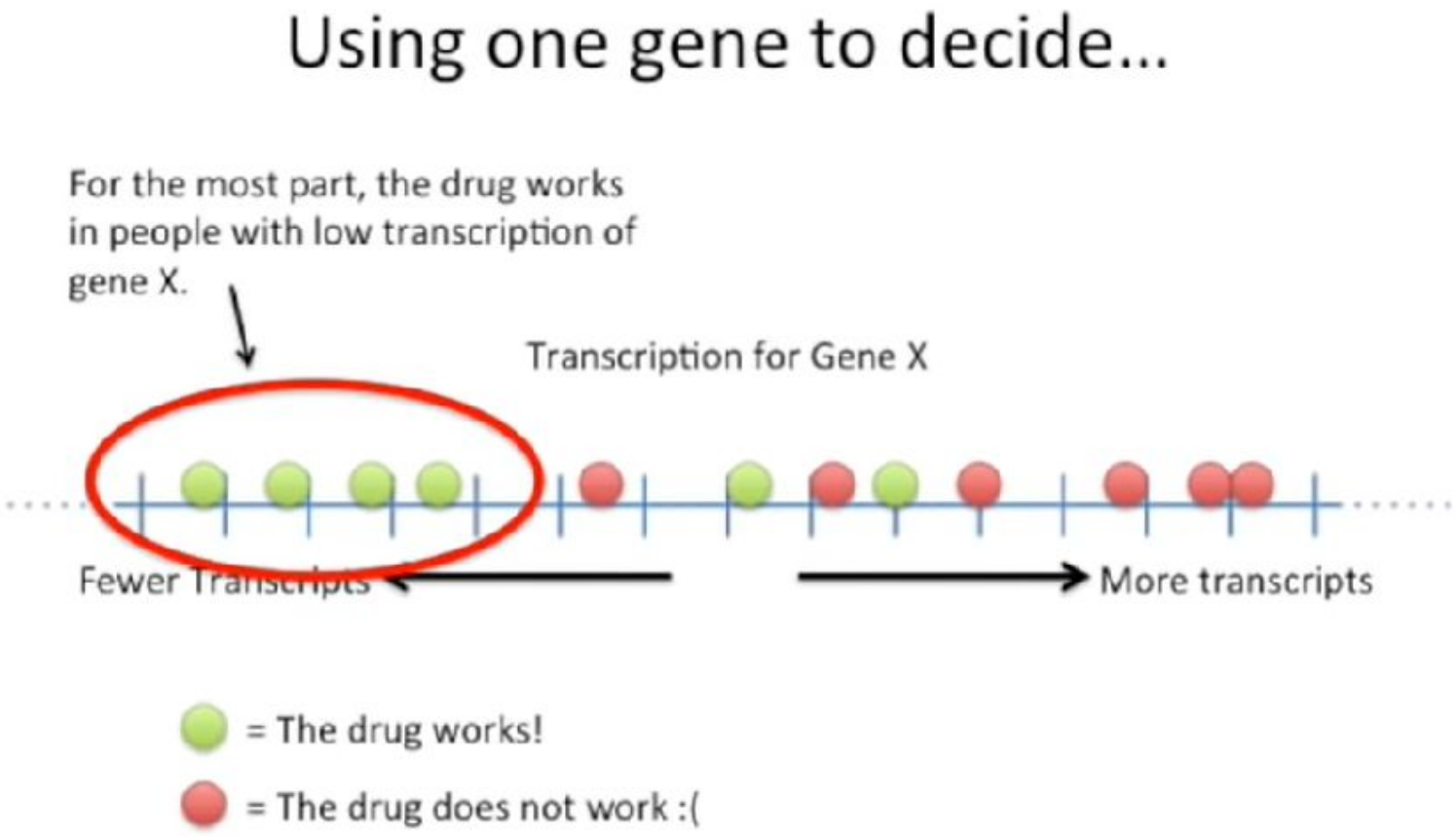
- We've got a cancer drug
  - It works great for some people…
  - … but it makes other people feel worse. :(
- How do we decide who to give the drug to?
  - Maybe gene expression can help us decide.

Linear Discriminant Analysis

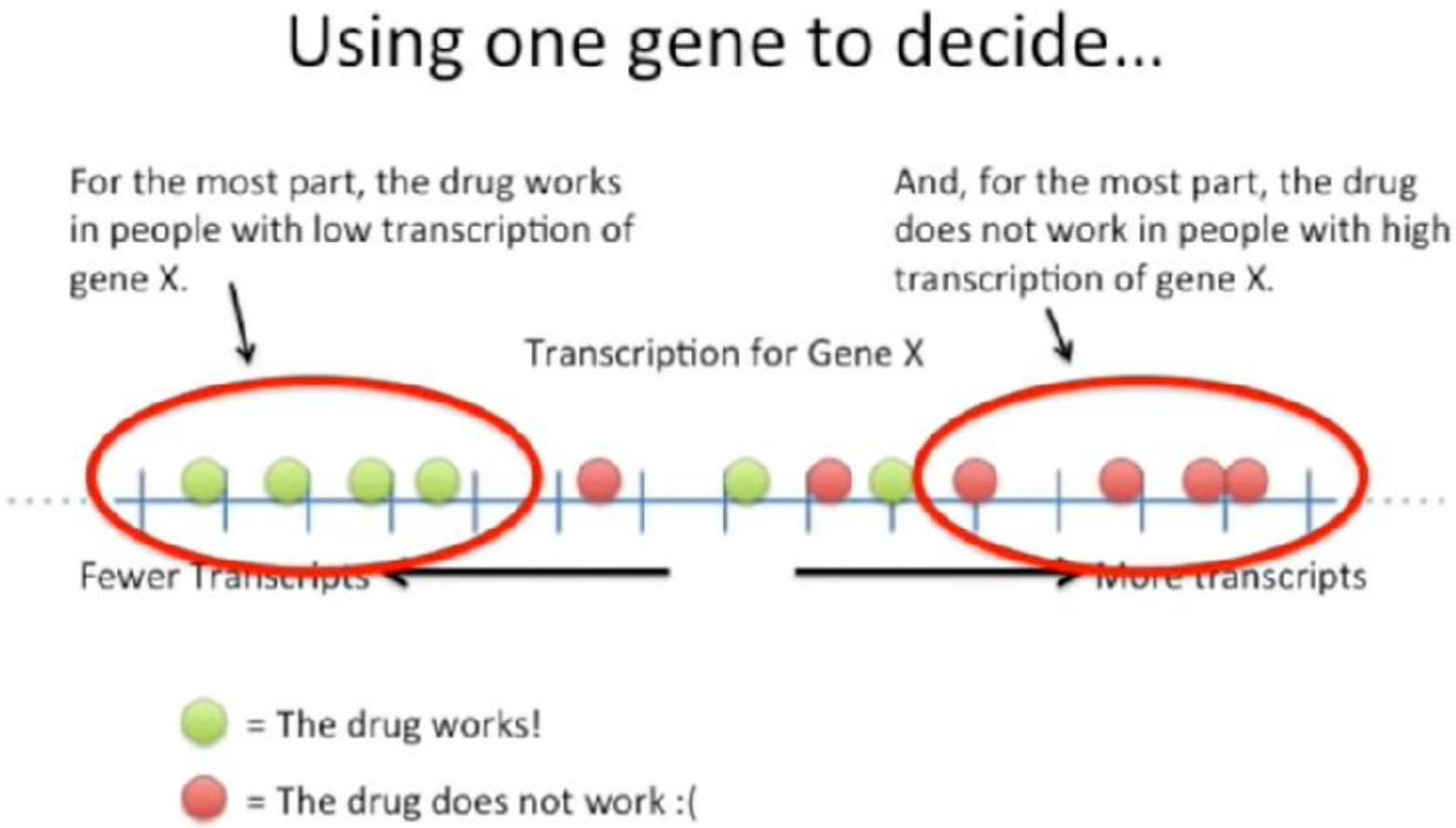
Using one gene to decide...



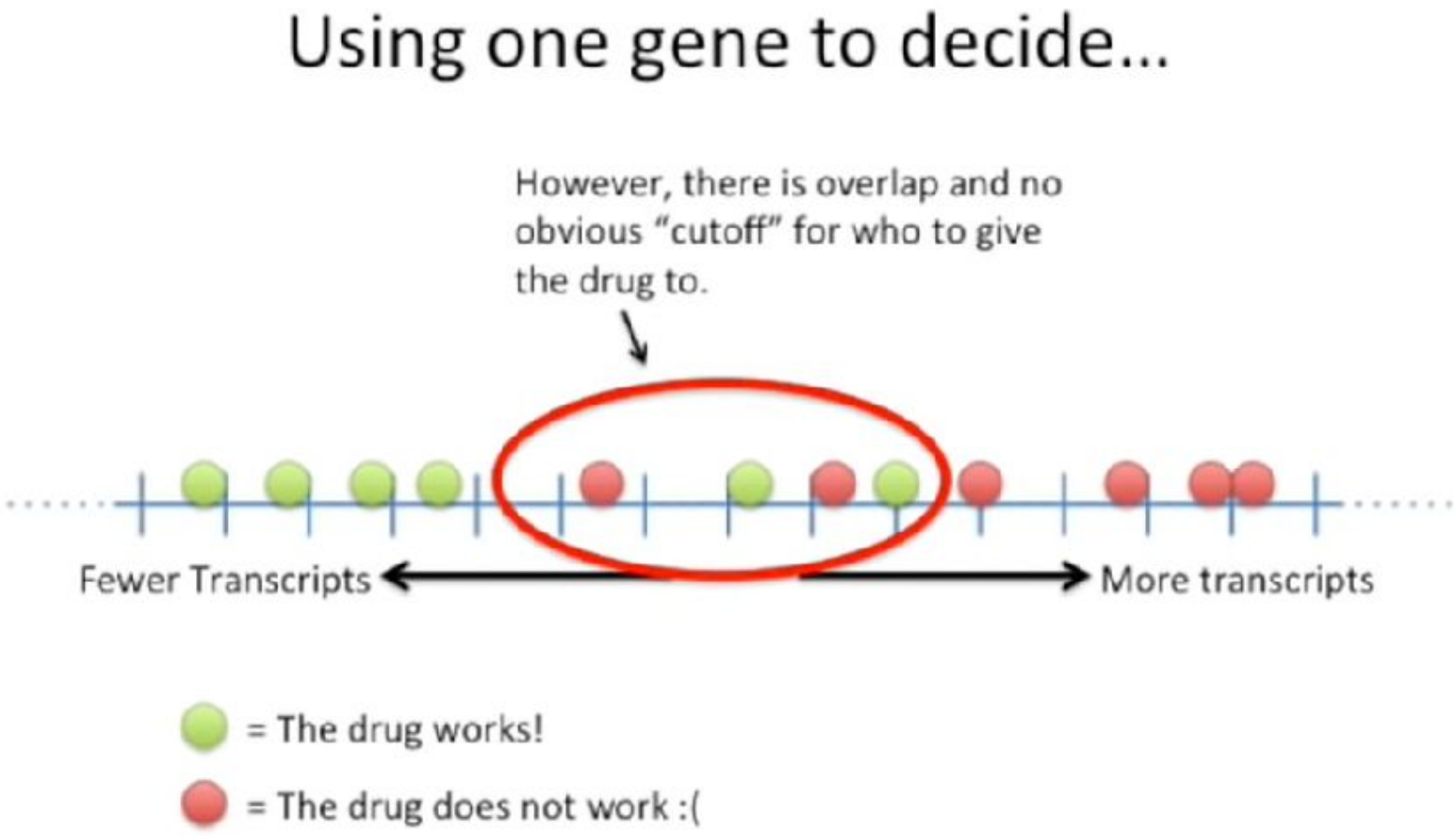
Linear Discriminant Analysis



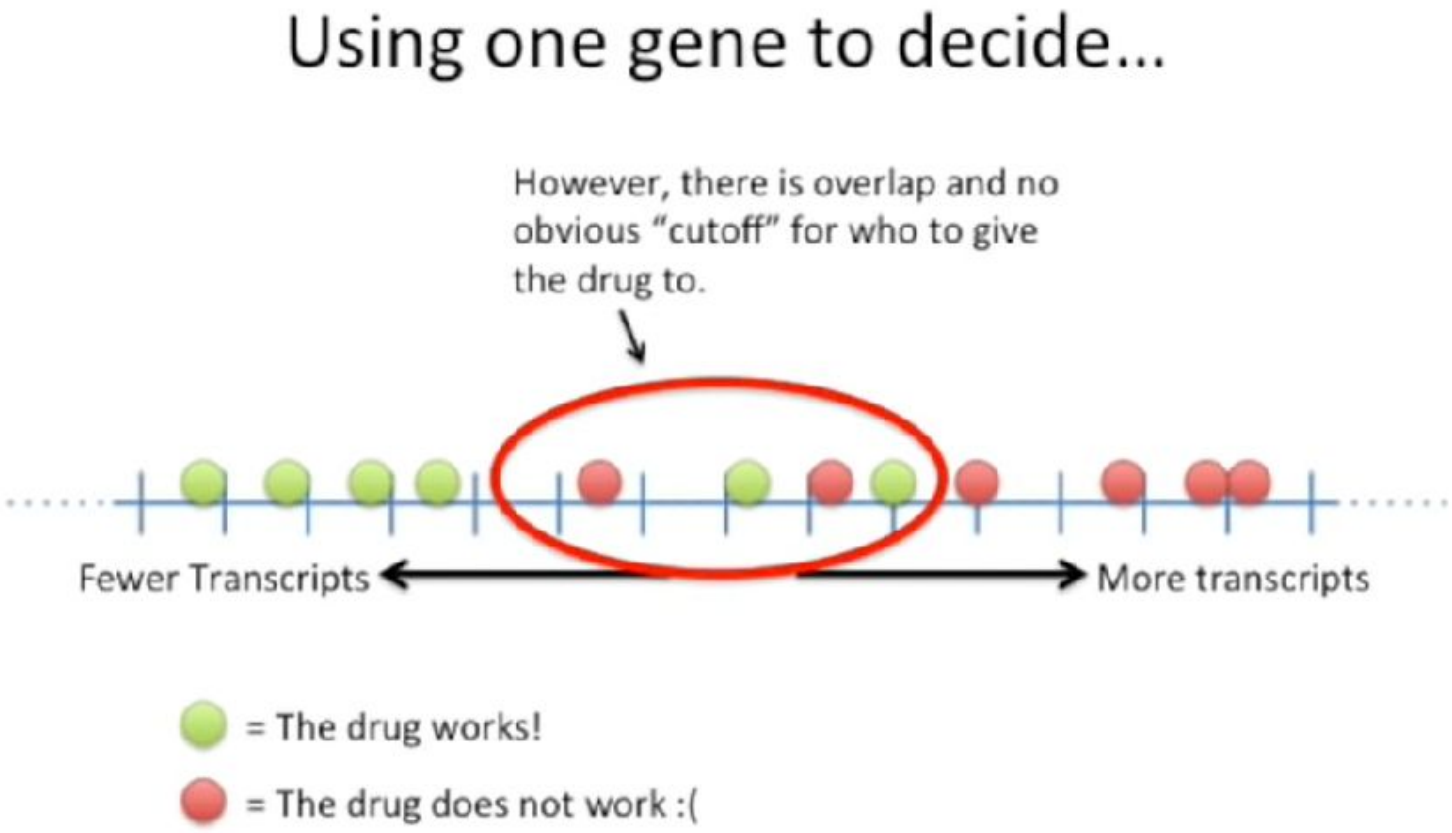
Linear Discriminant Analysis



Linear Discriminant Analysis

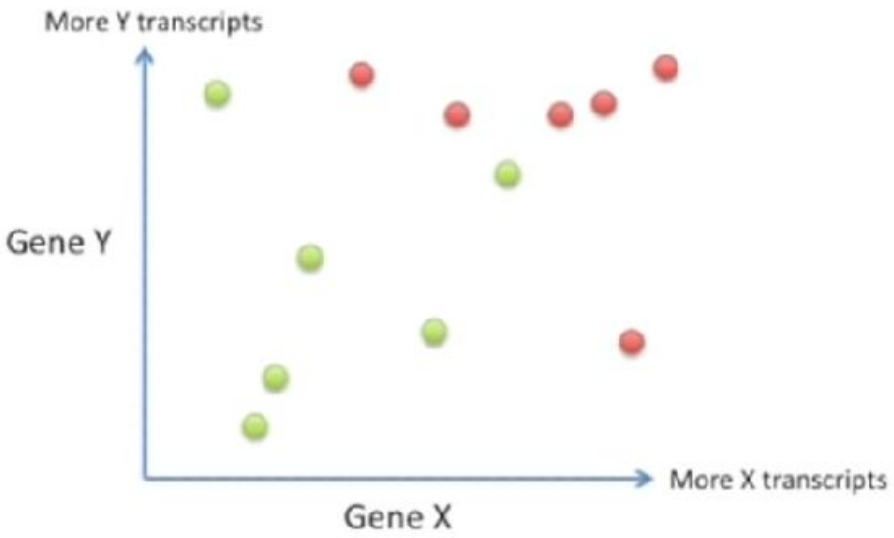


Linear Discriminant Analysis



Linear Discriminant Analysis

Using two genes to decide…

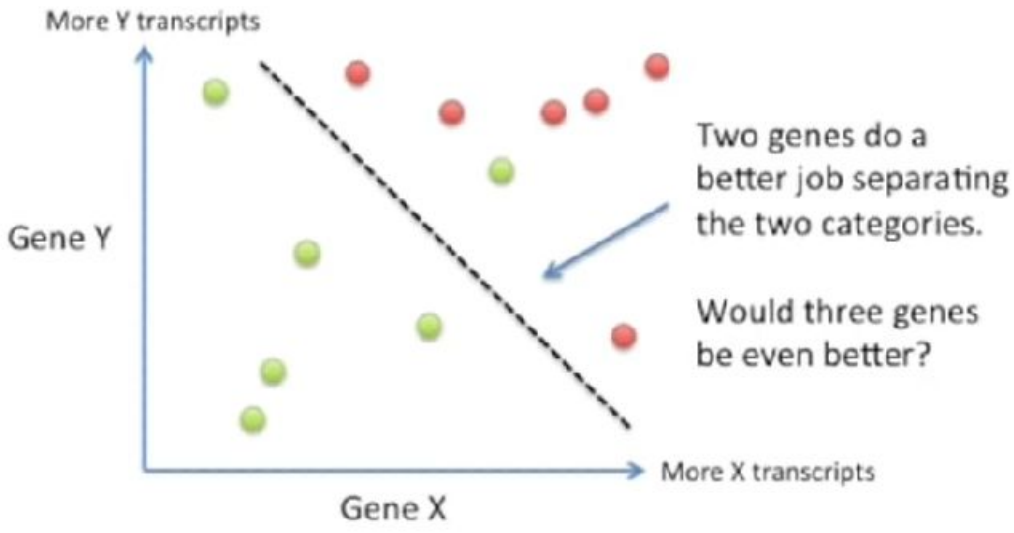


= The drug works!

= The drug does not work :(

Linear Discriminant Analysis

Using two genes to decide...

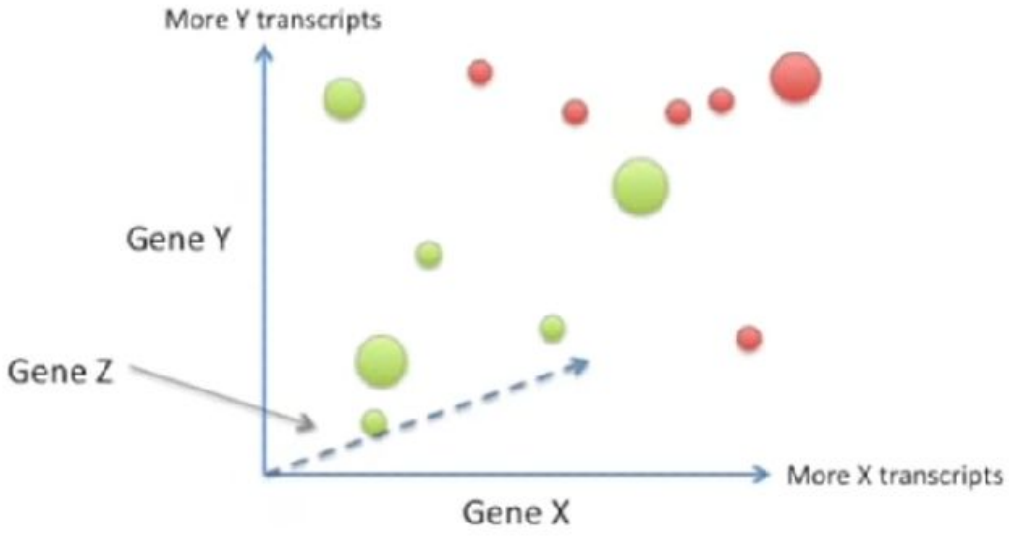


= The drug works!

= The drug does not work :(

Linear Discriminant Analysis

Using three genes to decide...

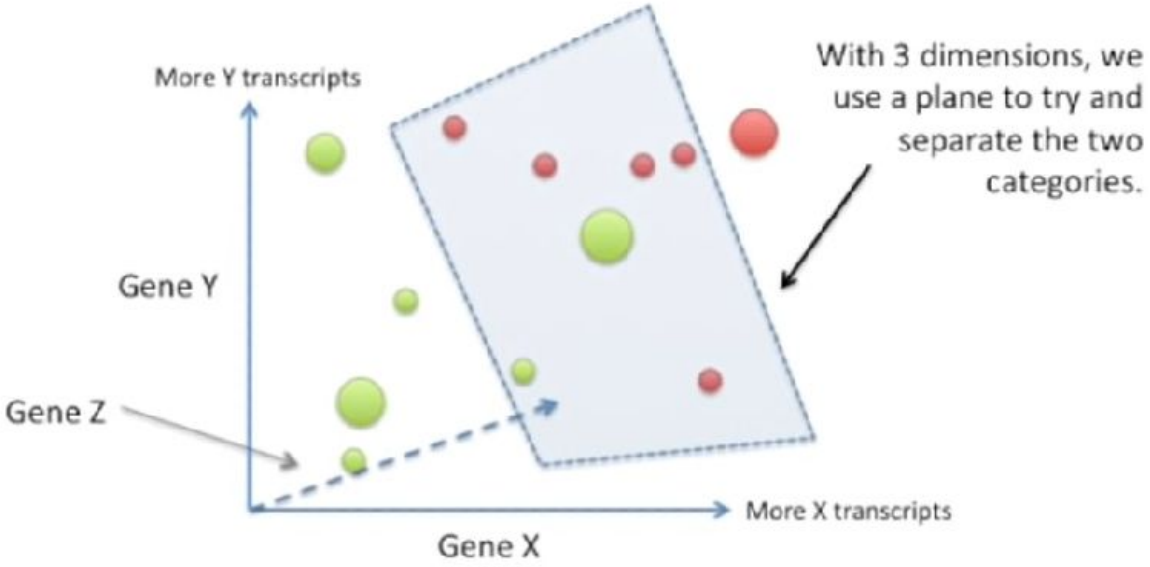


= The drug works!

= The drug does not work :(

Linear Discriminant Analysis

Using three genes to decide...



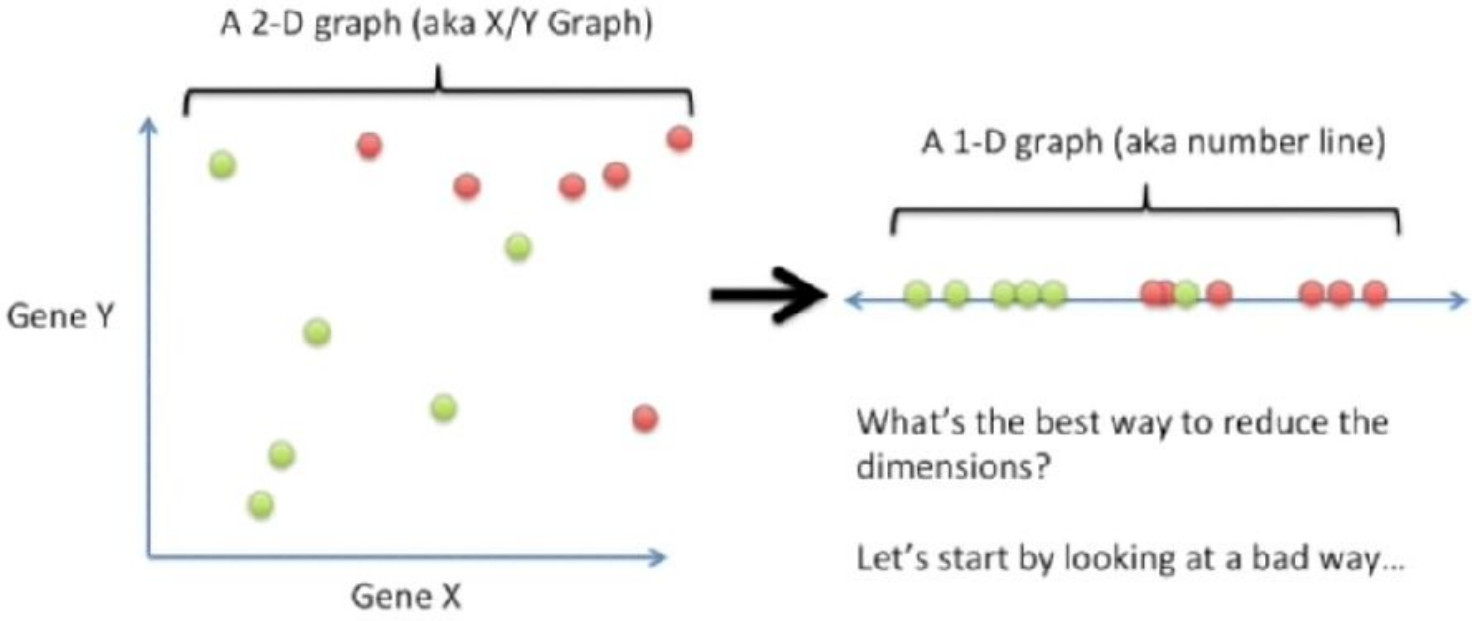
= The drug works!

= The drug does not work :(

Linear Discriminant Analysis

A super simple example

Reducing a 2-D graph to a 1-D graph

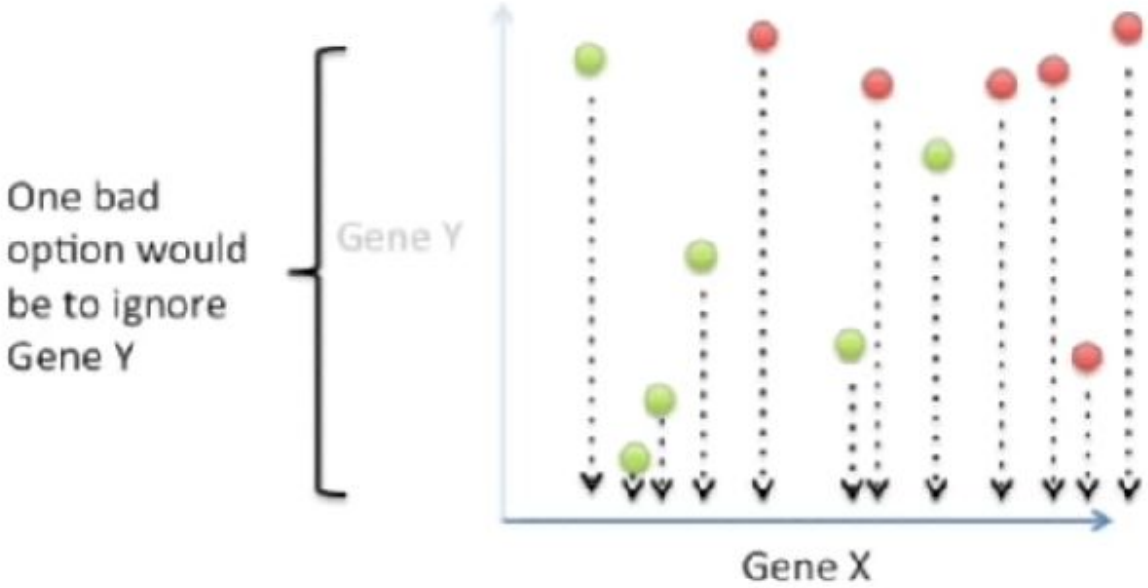


What's the best way to reduce the dimensions?

Let's start by looking at a bad way...

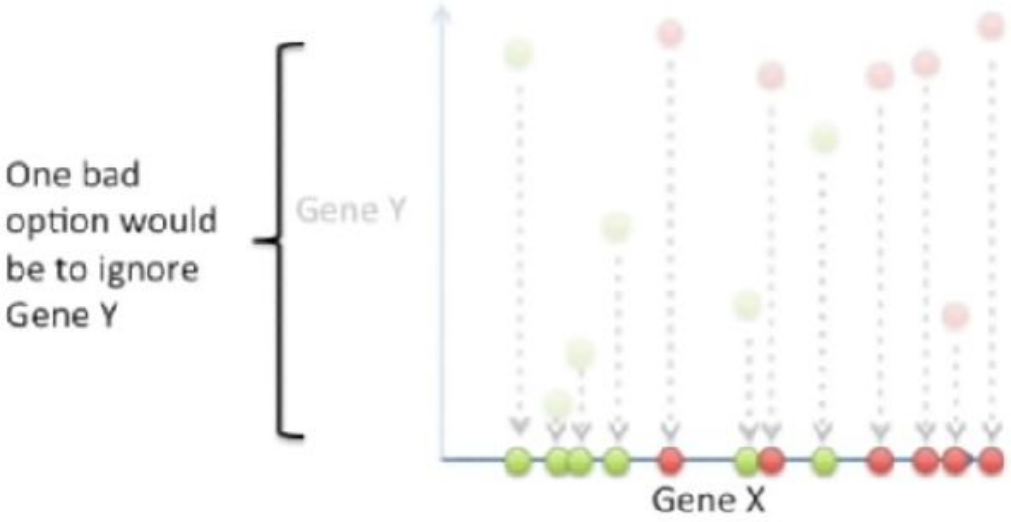
Linear Discriminant Analysis

Reducing a 2-D graph to a 1-D graph



Linear Discriminant Analysis

Reducing a 2-D graph to a 1-D graph

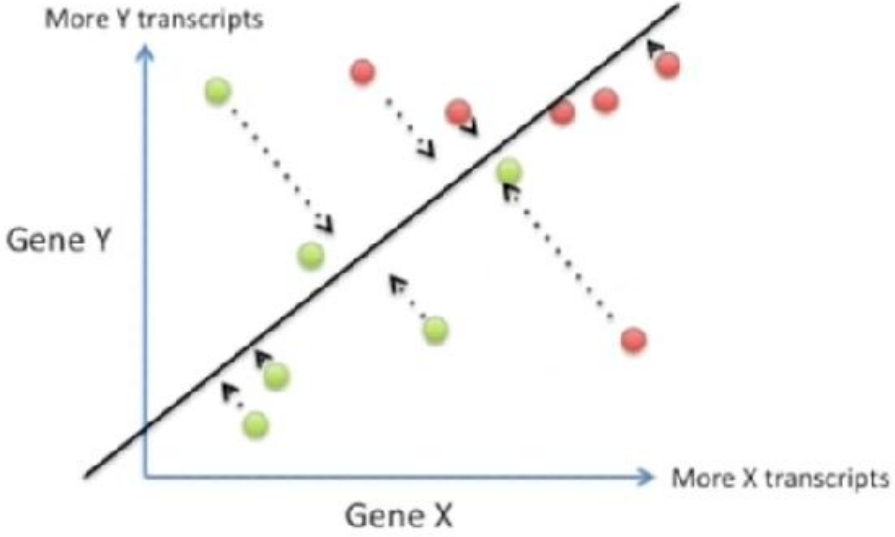


This way is bad because it ignores the useful information that Gene Y provides...

Projecting the genes onto the Y-axis (i.e. ignoring Gene X) isn't any better

Linear Discriminant Analysis

Reducing a 2-D graph to a 1-D graph with LDA

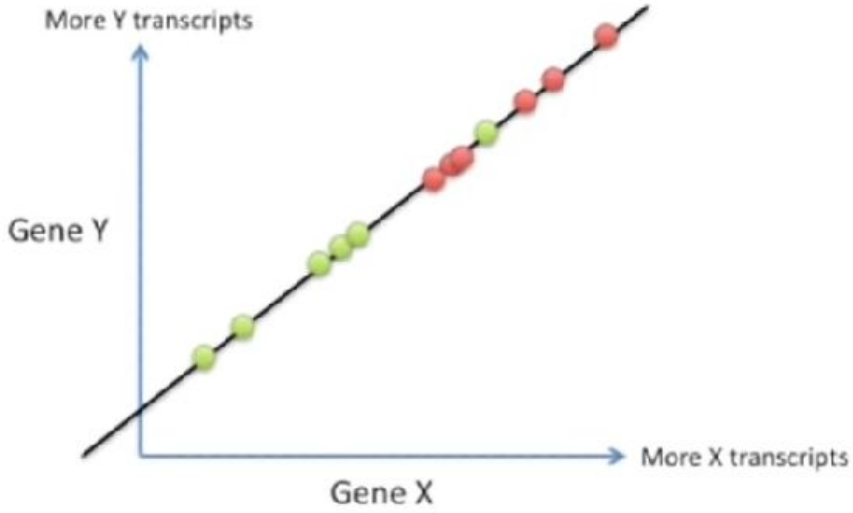


LDA uses both genes to create a new axis...

...and projects the data onto this new axis in a way to maximize the separation of the two categories.

Linear Discriminant Analysis

Reducing a 2-D graph to a 1-D graph with LDA

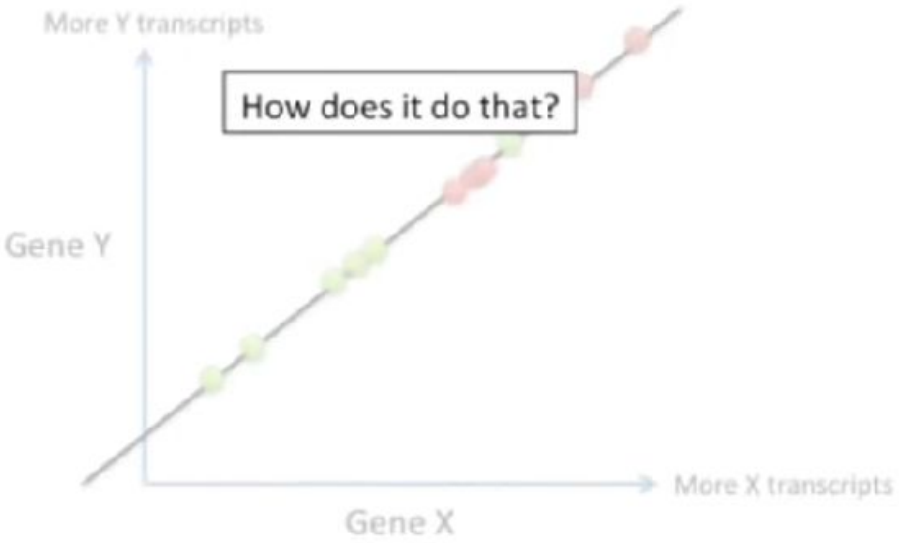


LDA uses both genes to create a new axis...

...and projects the data onto this new axis in a way to maximize the separation of the two categories.

Linear Discriminant Analysis

Reducing a 2-D graph to a 1-D graph with LDA



LDA uses both genes to create a new axis...

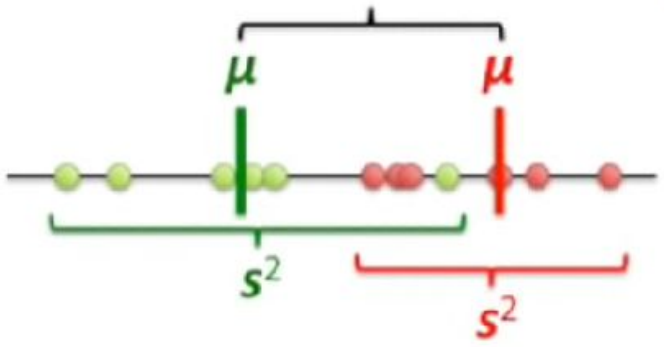
...and projects the data onto this new axis in a way to maximize the separation of the two categories.

Linear Discriminant Analysis

How LDA creates a new axis…

The new axis is created according to two criteria (considered simultaneously):

1) Maximize the distance between means.



Let's call $(\mu - \mu)$ ***d*** for ***d***istance.

$$\frac{(\mu - \mu)^2}{s^2 + s^2} \rightarrow \frac{\text{Ideally large}}{\text{Ideally small}}$$

2) Minimize the variation (which LDA calls "scatter" and is represented by s^2) within each category.

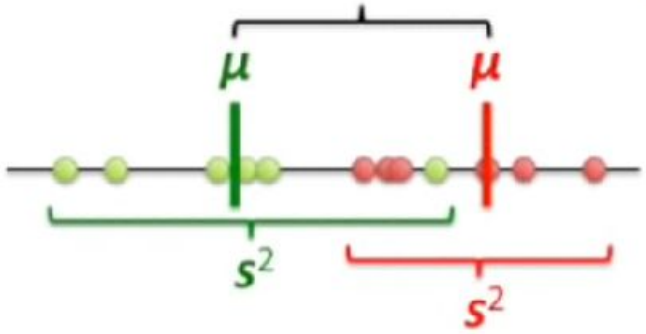
Linear Discriminant Analysis

How LDA creates a new axis...

The new axis is created according to two criteria (considered simultaneously):

1) Maximize the distance between means.

Let's call ($\mu - \mu$) ***d*** for ***d***istance.

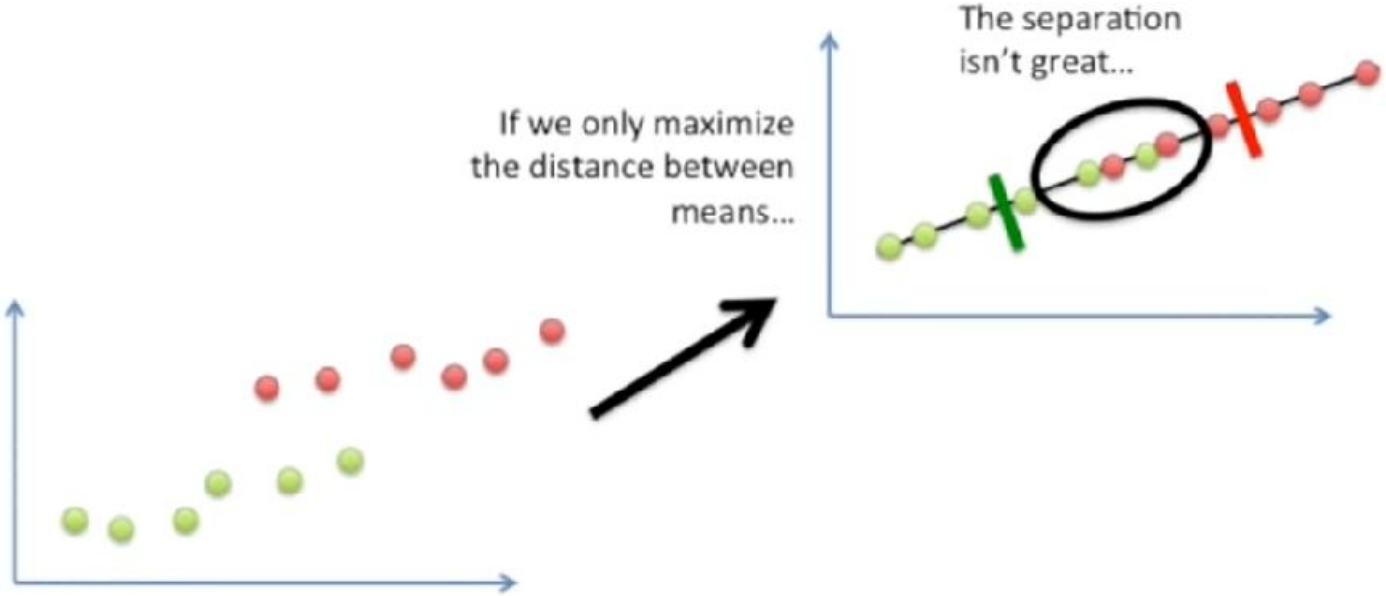


$$\frac{d^2}{s^2 + s^2} \rightarrow \frac{\text{Ideally large}}{\text{Ideally small}}$$

2) Minimize the variation (which LDA calls "scatter" and is represented by s^2) within each category.

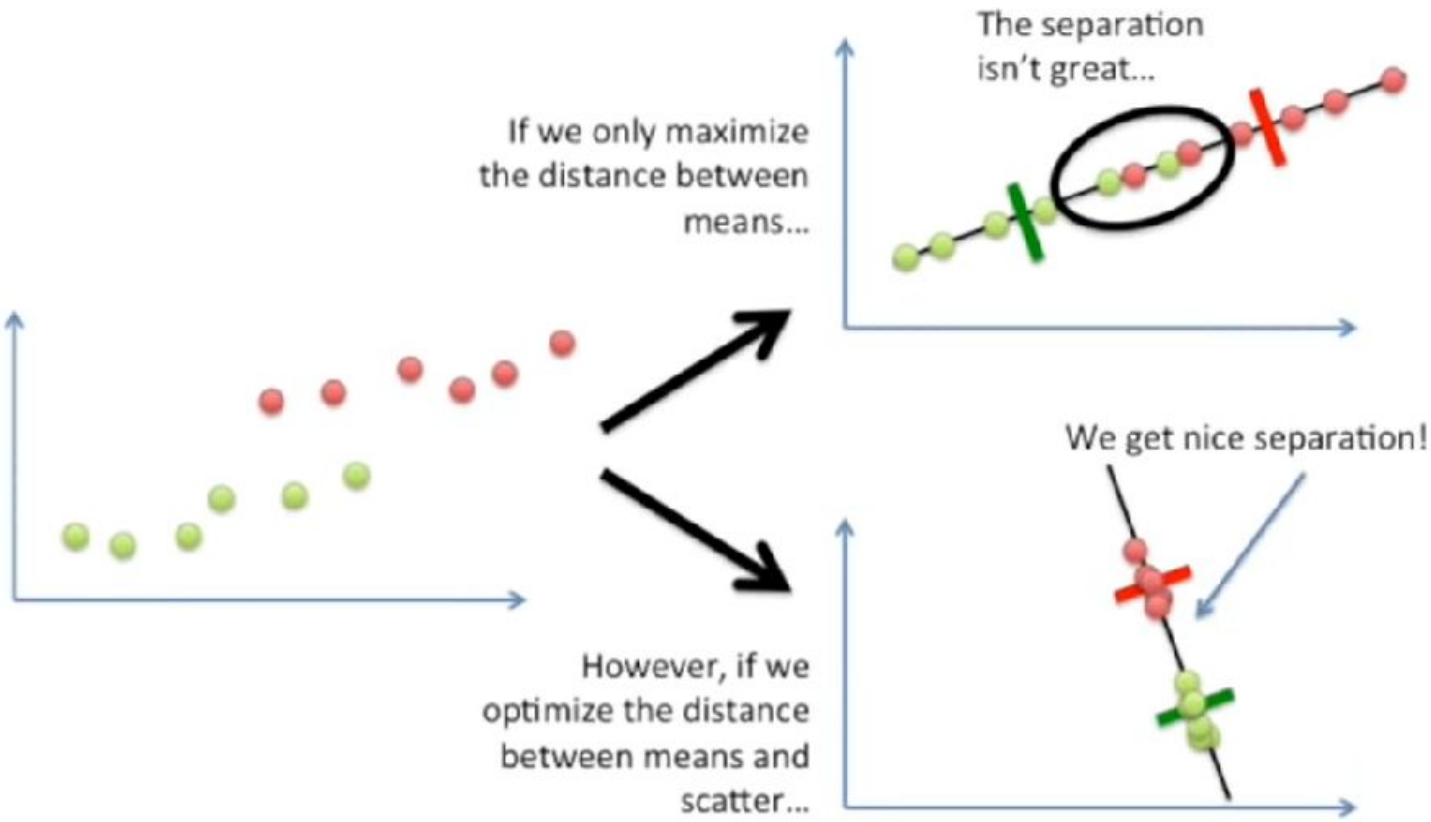
Linear Discriminant Analysis

An example showing why both distance and scatter are important.



Linear Discriminant Analysis

An example showing why both distance and scatter are important.

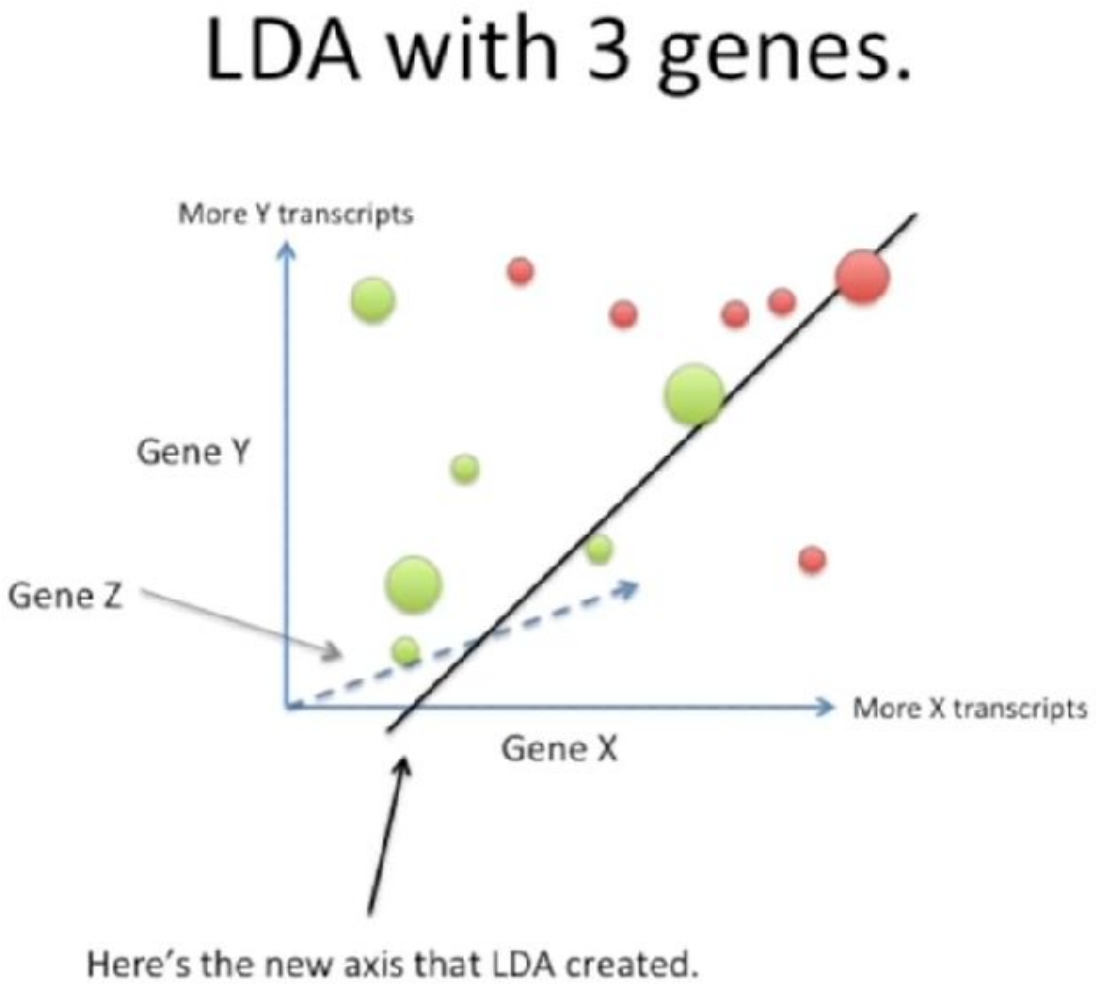


Linear Discriminant Analysis

What if we have more than 2 genes (more than 2 dimensions)?

- The process is the same:
  - Create an axis that maximizes the distance between the means for the two categories while minimizing the scatter.

Linear Discriminant Analysis



Linear Discriminant Analysis

LDA with 3 genes.

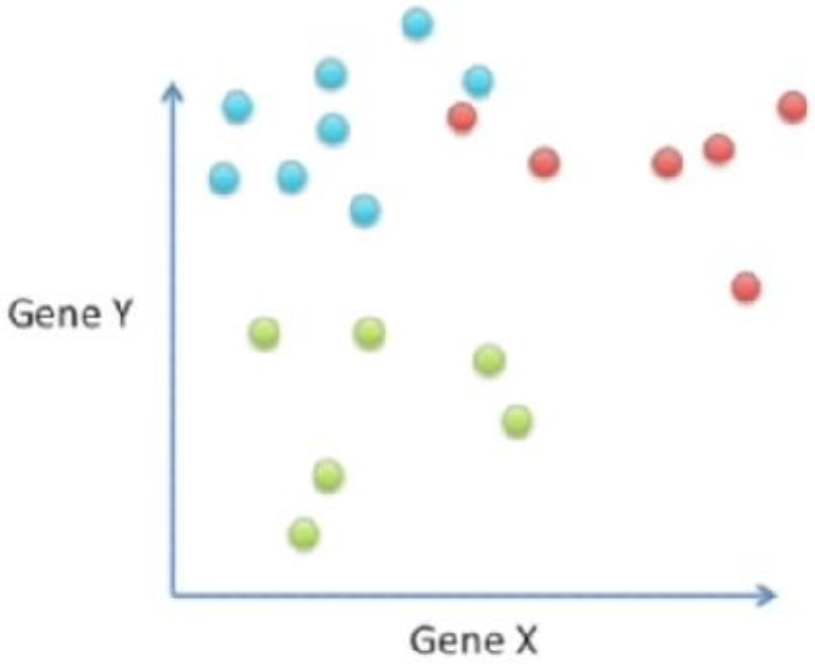
The data are projected onto the new axis.

The axis was chosen to maximize the distance between the two means (between the two categories) while minimizing the "scatter".

Linear Discriminant Analysis

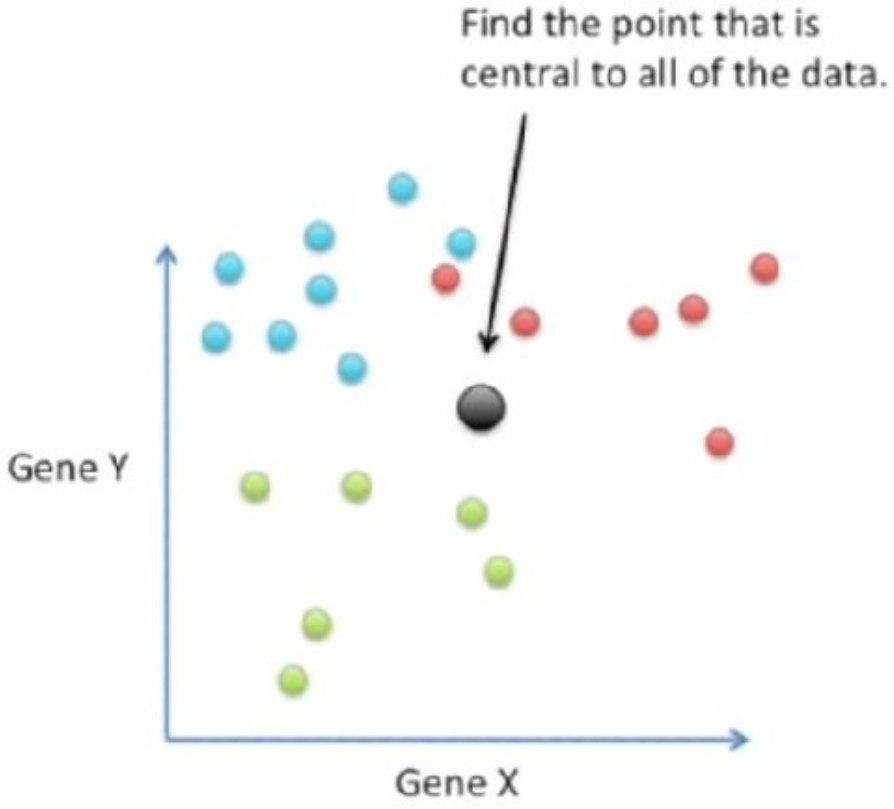
LDA for 3 categories

The first difference is how you measure the distances among the means.



Linear Discriminant Analysis

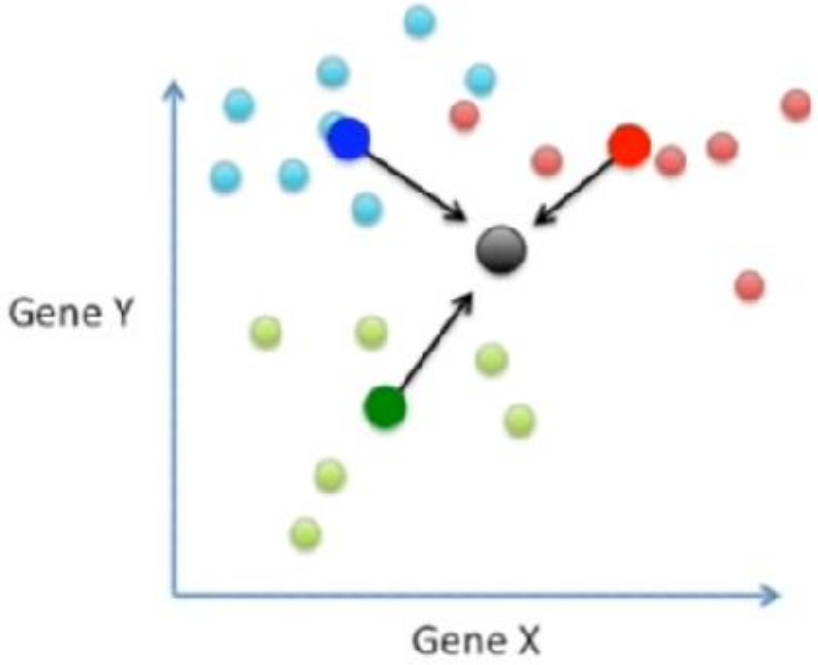
LDA for 3 categories



Linear Discriminant Analysis

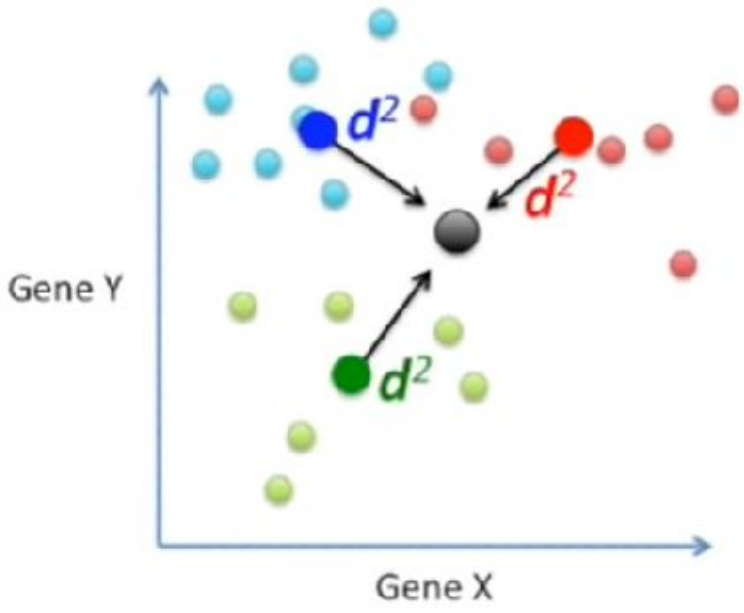
LDA for 3 categories

Then measure the distances between a point that is central in each category and the main central point.



Linear Discriminant Analysis

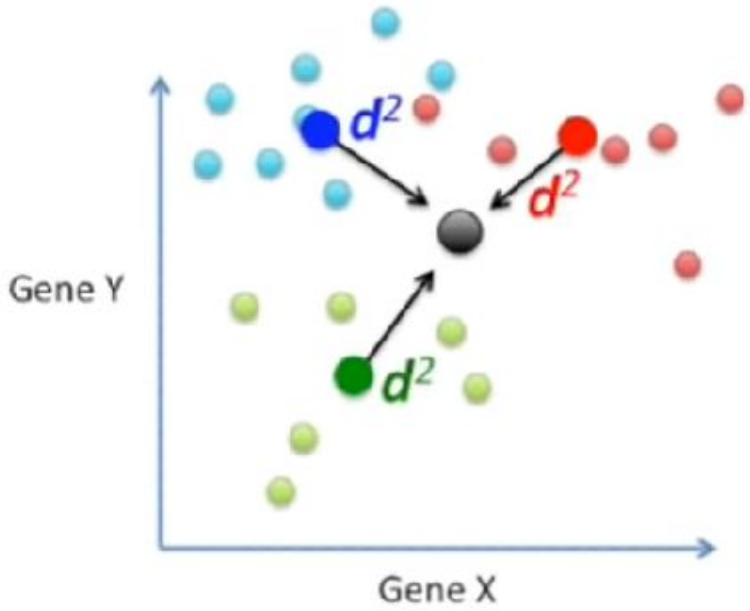
LDA for 3 categories



Now maximize the distance between each category and the central point while minimizing the scatter for each category.

Linear Discriminant Analysis

LDA for 3 categories



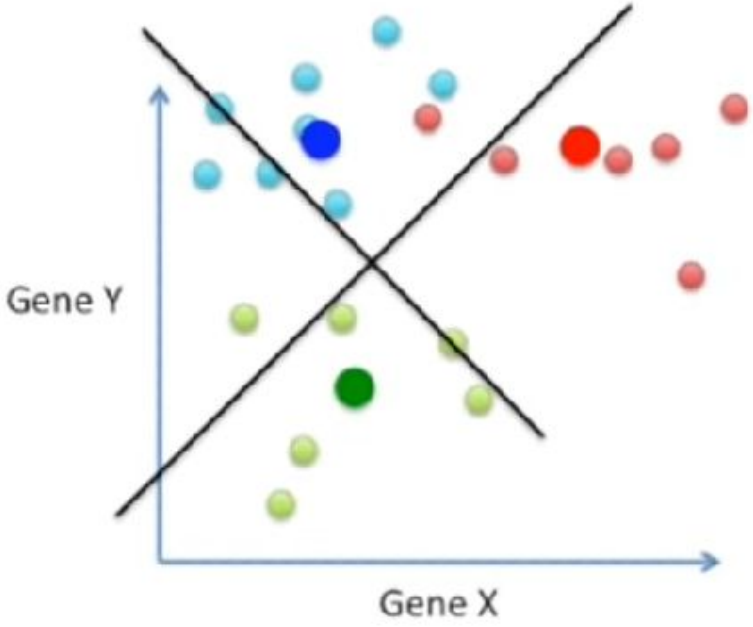
$$\frac{d^2 + d^2 + d^2}{s^2 + s^2 + s^2}$$

This is the same equation as before, but now there are terms for the **blue** category.

Linear Discriminant Analysis

LDA for 3 categories

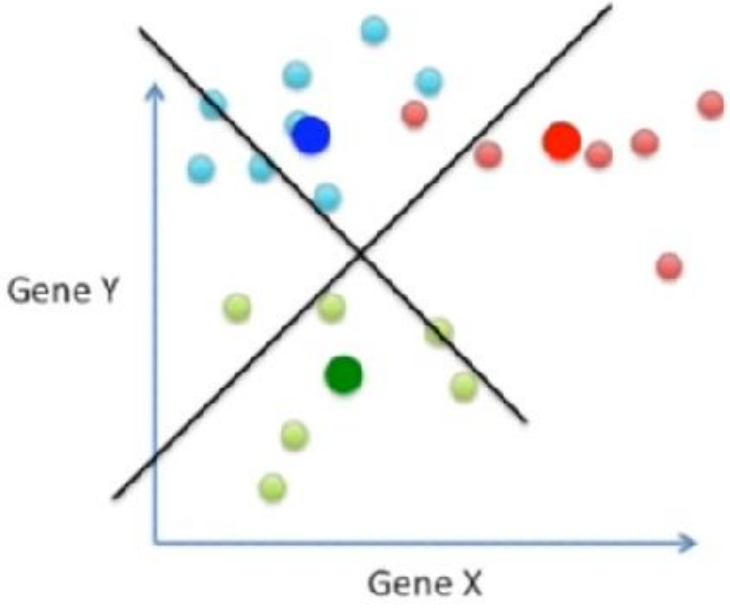
The second difference is LDA creates 2 axes to separate the data.
This is because the 3 central points for each category define a plane.
(Remember from high school: 2 points define a line, 3 points define a plane...)



Linear Discriminant Analysis

LDA for 3 categories

The second difference is LDA creates 2 axes to separate the data.
This is because the 3 central points for each category define a plane.
(Remember from high school: 2 points define a line, 3 points define a plane...)



With three points, we can draw two lines to optimize separation.