Machine Learning

Lecture 5 – Odds and Logistic Regression

Dr SHI Lei

Last Lecture

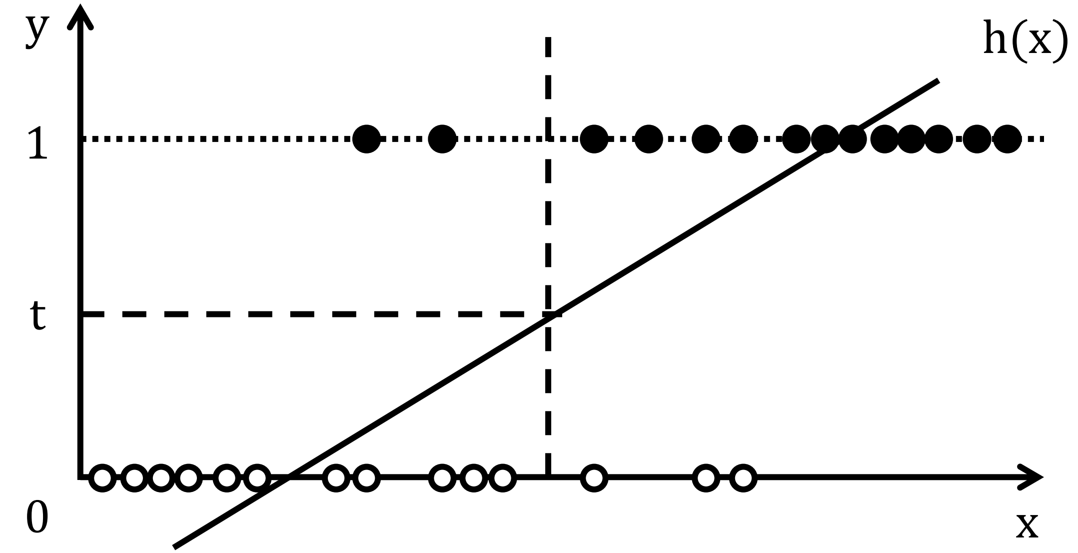
- Binary Classifier
- Performance Measures

Last Lecture

Binary Classifier

- Observed response **y** takes only two possible values **+** and **–**
- Define relationship between **h(x)** and **y**
- Use the decision rule: $\hat{y} = \begin{cases} +, & h(x) \geq t \\ -, & otherwise \end{cases}$

	X					y	h(x)	$\hat{y}$
1						+	h_1	+
2						+	h_2	−
3						−	h_3	+
4						−	h_4	−
…				…			…	
n						−	h_n	+
…				…			…	



Last Lecture

Performance Measures

- Prediction Success (Confusion Matrix)

predicted \ actual	+	−
+	true positives	false positives
−	false negatives	true negatives

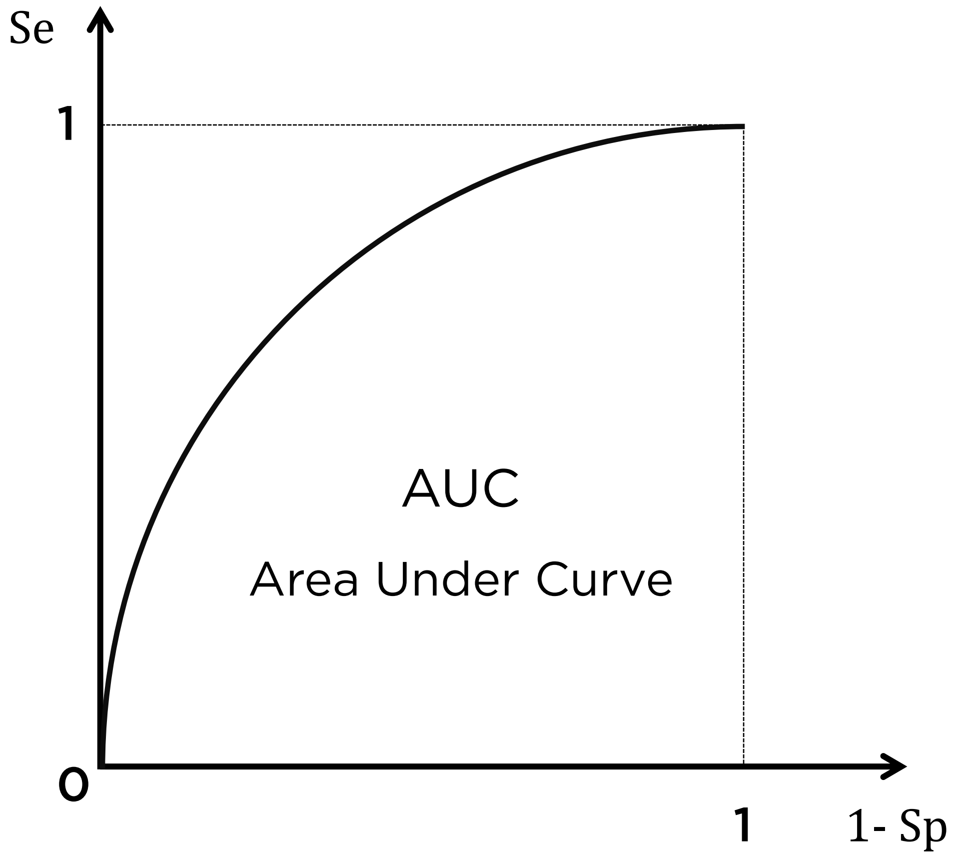
- Precision, Sensitivity (Recall), Specificity

$$Pr = \frac{tp}{tp + fp} \qquad Se = \frac{tp}{tp + fn} \qquad Sp = \frac{tn}{tn + fp}$$

Last Lecture

Performance Measures

- ROC Curve (receiver operating characteristic curve)



It tells how much model is capable of distinguishing between classes.

Today

- Odds
- Logistic Regression

Odds

Odds

Odds, a numerical expression, expressed as a pair of numbers.

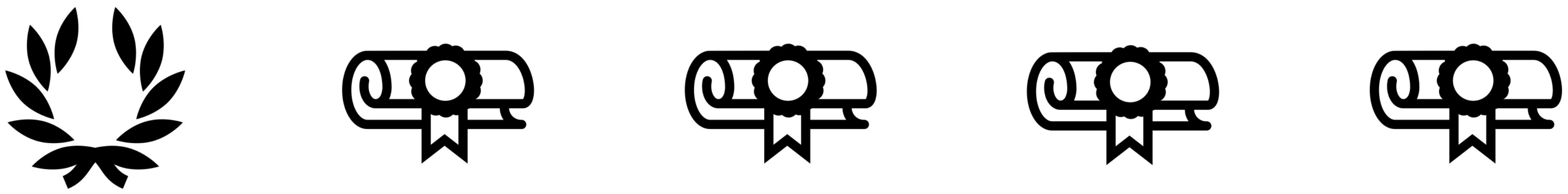
The **odds for** or **odds of** some *event* reflect the likelihood that the event will take place, while **odds against** reflect the likelihood that it will not.

An example …

Odds

An example

We may say the **odds** in favour of students to graduate with 1st-class honours is 1 to 4:



Visually, there are 5 students total.

1 of them will graduate with 1st-class honours.

4 of them will graduate without 1st-class honours.

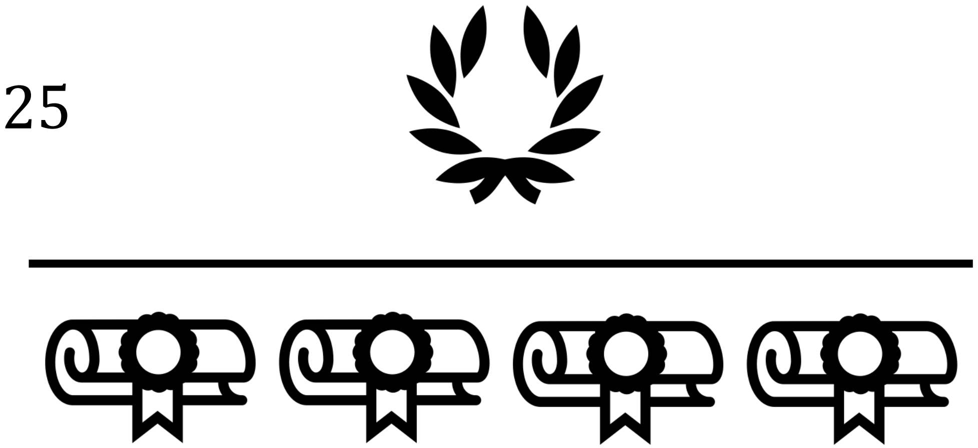
So, the odds are 1 to 4.

Odds

An example

We may say the **odds** in favour of students to graduate with 1st-class honours is 1 to 4:

Alternatively, we can write this as a **fraction** $\frac{1}{4} = 0.25$

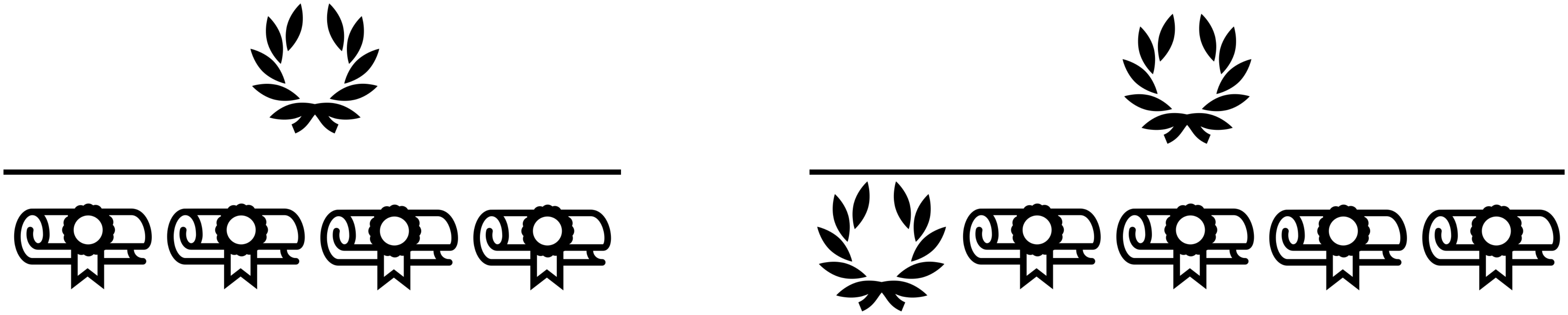


Visually, we have one students graduate with 1st-class honours, divided by the 4 who not.

NOTE: Odds are not probabilities.

Odds

NOTE: Odds are not probabilities.

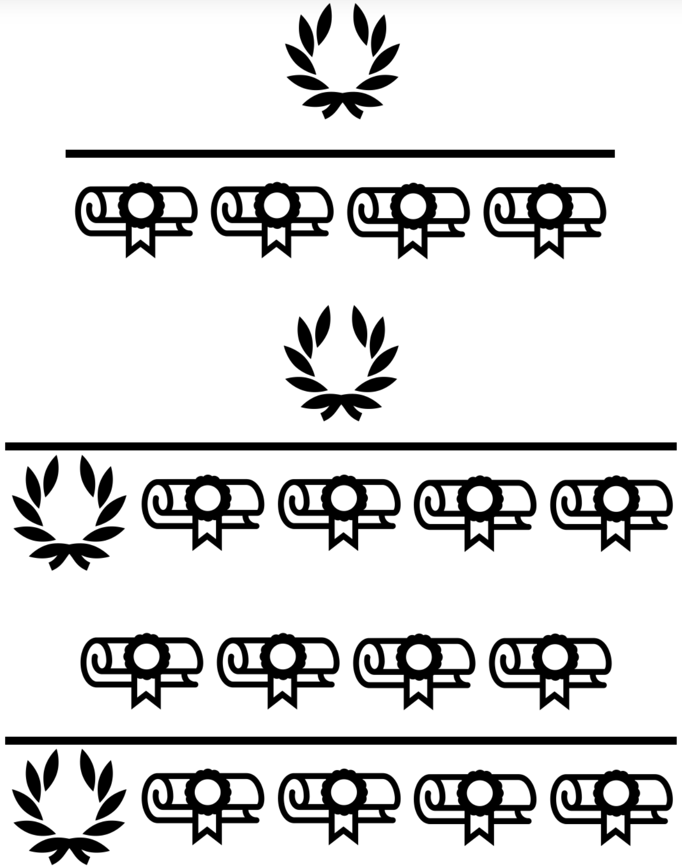


odds vs **probability**

1 to 4

1 to 5

Odds



odds(success) $= \frac{1}{4} = 0.25$

probability(success) $= \frac{1}{5} = 0.20$

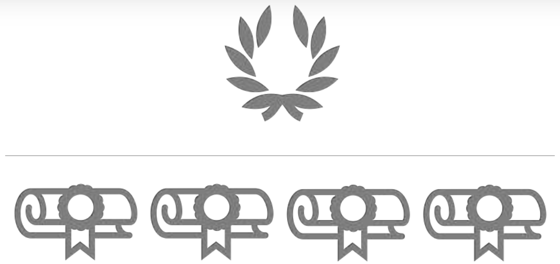
probability(unsuccess) $= \frac{4}{5} = 0.80$

either...

probability(unsuccess) = 1 – probability(unsuccess) $= 1 - \frac{1}{5} = \frac{4}{5} = 0.80$

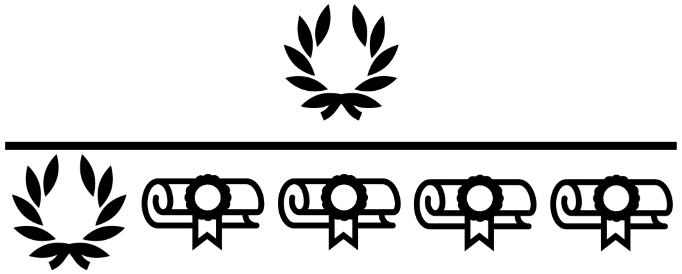
Odds

odds(success)



$$= \frac{1}{4} = 0.25$$

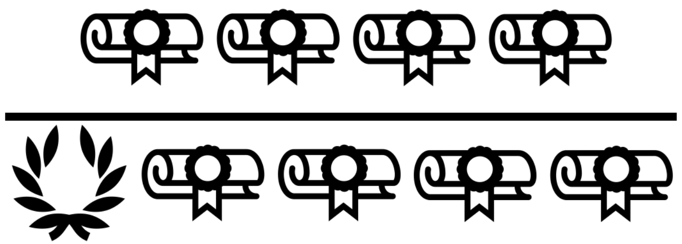
probability(success)



$$= \frac{1}{5} = 0.20 \qquad p$$

$$\updownarrow$$

probability(unsuccess)



$$= \frac{4}{5} = 0.80 \qquad q$$

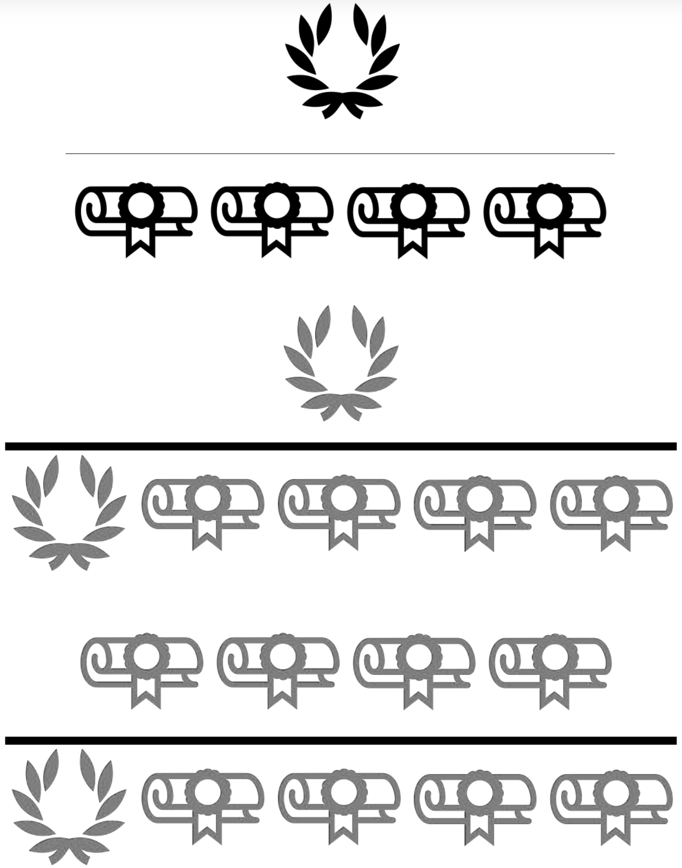
$$\frac{\text{probability(success)}}{\text{probability(unsuccess)}}$$

$$\frac{p}{q} = \frac{0.20}{0.80} = 0.25$$

or

$$\frac{p}{1-p} = \frac{1/\cancel{5}}{4/\cancel{5}} = \frac{1}{4} = 0.25$$

Odds



odds(success) $= \frac{1}{4} = 0.25$

probability(success) $= \frac{1}{5} = 0.20$

probability(unsuccess) $= \frac{4}{5} = 0.80$

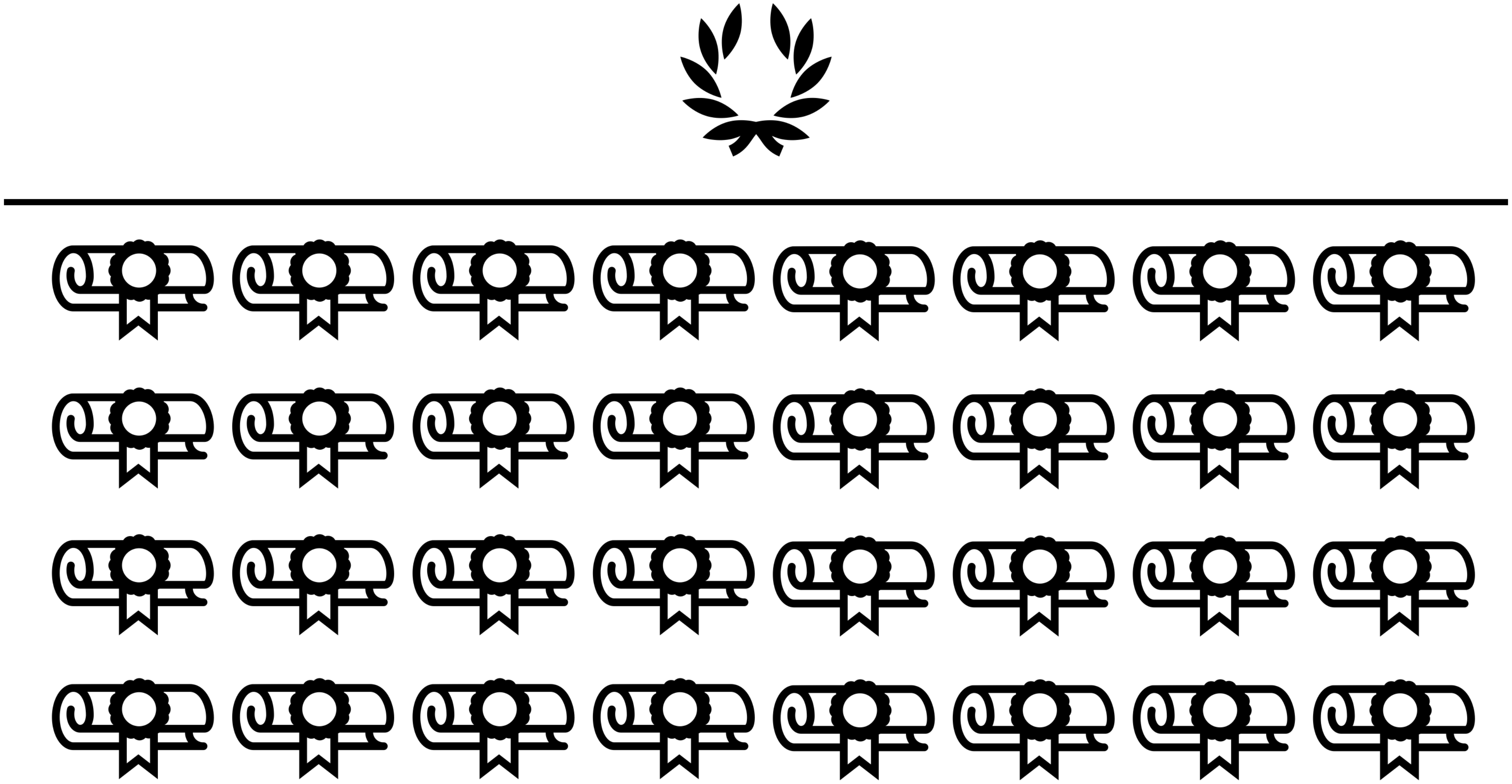
$$\frac{\text{probability(success)}}{\text{probability(unsuccess)}}$$

$\frac{p}{q} = \frac{0.20}{0.80} = 0.25$ or $\frac{p}{1-p} = \frac{1/\cancel{5}}{4/\cancel{5}} = \frac{1}{4} = 0.25$

Log of odds

log(odds)

odds(success), if students were the worst



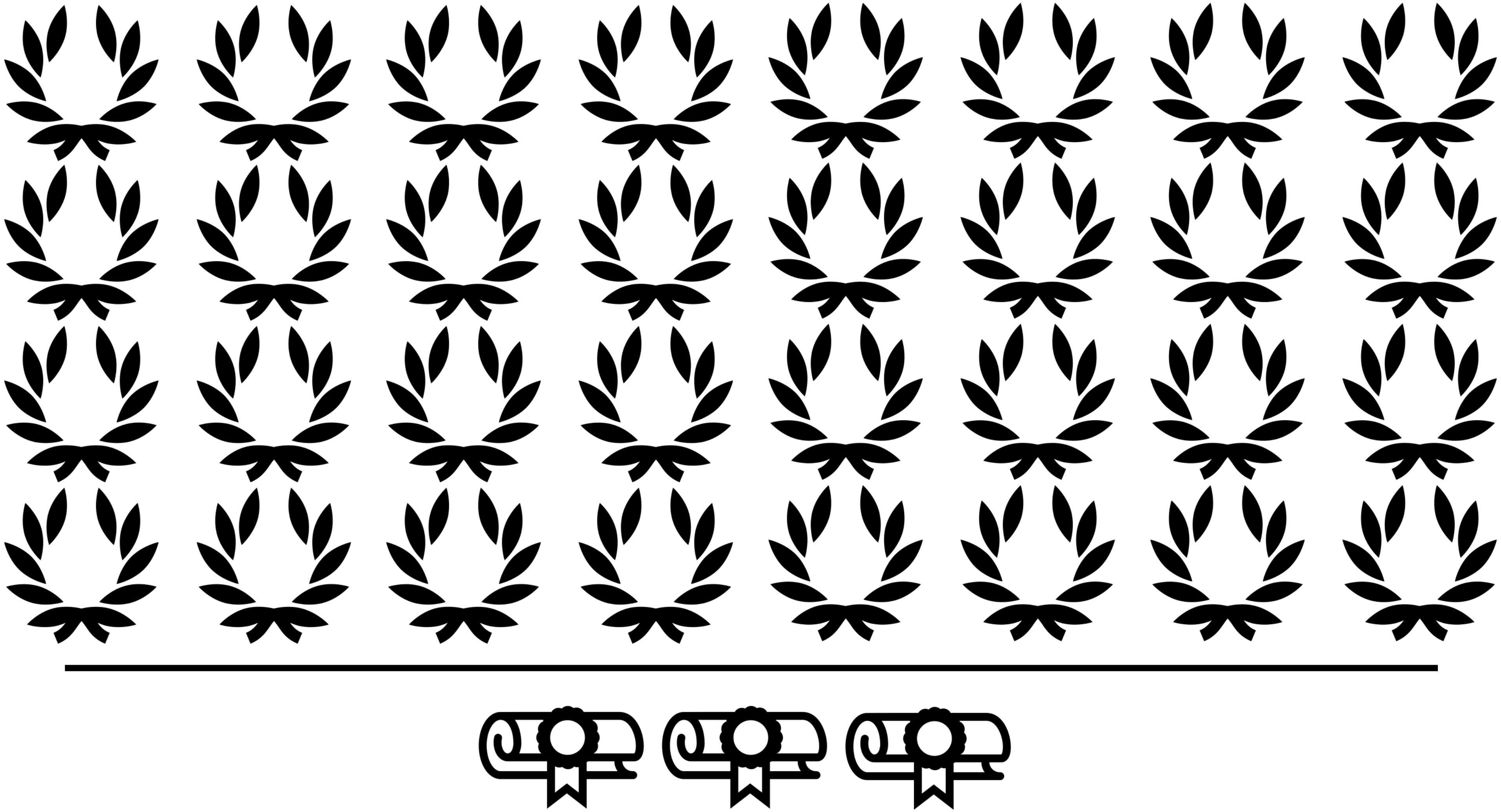
$$= \frac{1}{32} = 0.031$$

log(odds)

Odds against success is between 0 and 1

log(odds)

odds(success), if students were really good



$$= \frac{32}{3} = 10.7$$

log(odds)

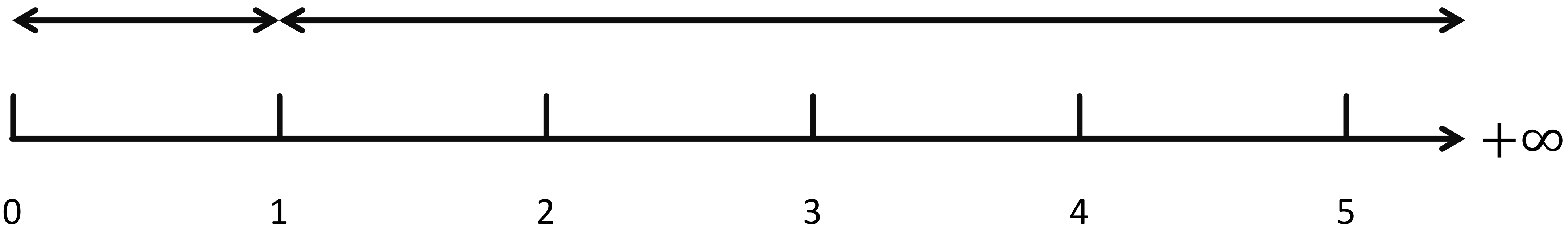
Odds in favour of success is between 1 and $+\infty$

log(odds)

Another way to look at this is with a number line

Odds(unsuccess) go from 0 to 1

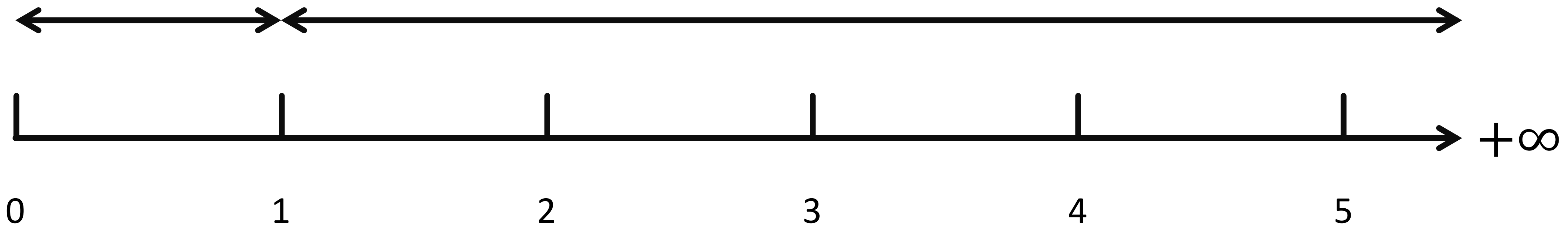
Odds(success) go from 1 to +∞



log(odds)

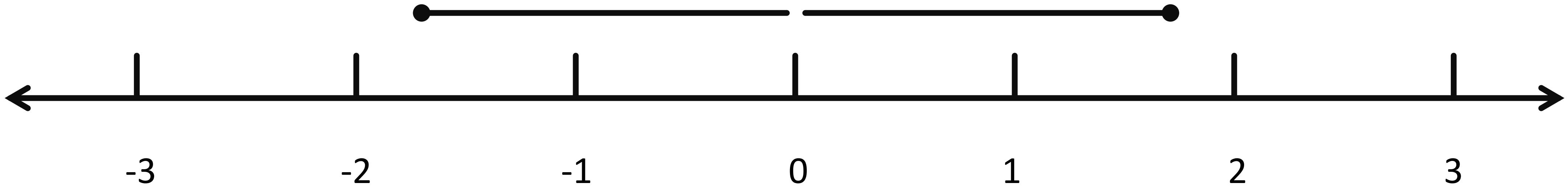
The **asymmetry** makes it difficult to compare odds(success) and odds(unsuccess)

The **magnitude** of these odds looks way smaller



log(odds)

Taking the log() of the odds (i.e. log(odds)) solves this problem by making everything symmetrical.



e.g. If odds(success) 1 to 6, then

log(odds)=log(1/6)=log(0.17)=-1.79

If odds(success) 6 to 1, then

log(odds)=log(6/1)=log(6)=1.79

Using the log function, the distance from the origin (or 0) is the same for 1 to 6 and 6 to 1 odds.

log(odds)

In Summary

The **odds** are the ratio of something happening to something not happening $\frac{p}{1-p}$

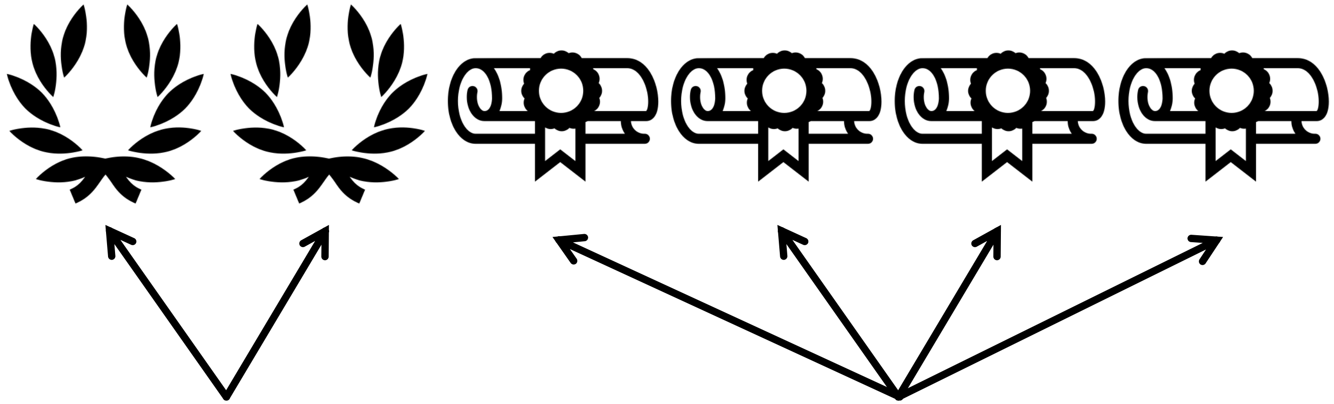
$$\log(\text{odds}) = \log\left(\frac{p}{1-p}\right)$$

logit function $\longrightarrow$ The basis for **logistic regression**

Odds Ratios

Odds Ratios

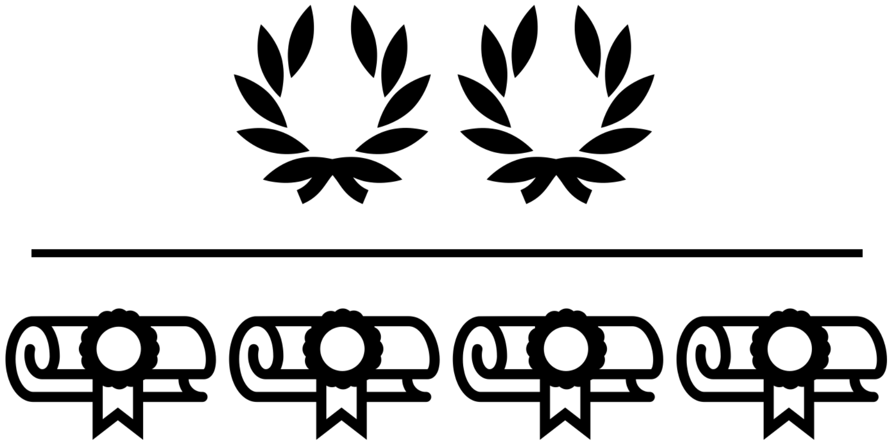
$$Odds = \frac{something\ happening}{something\ not\ happening}$$



Happening / success   Not happening /unsuccess

Odds Ratios

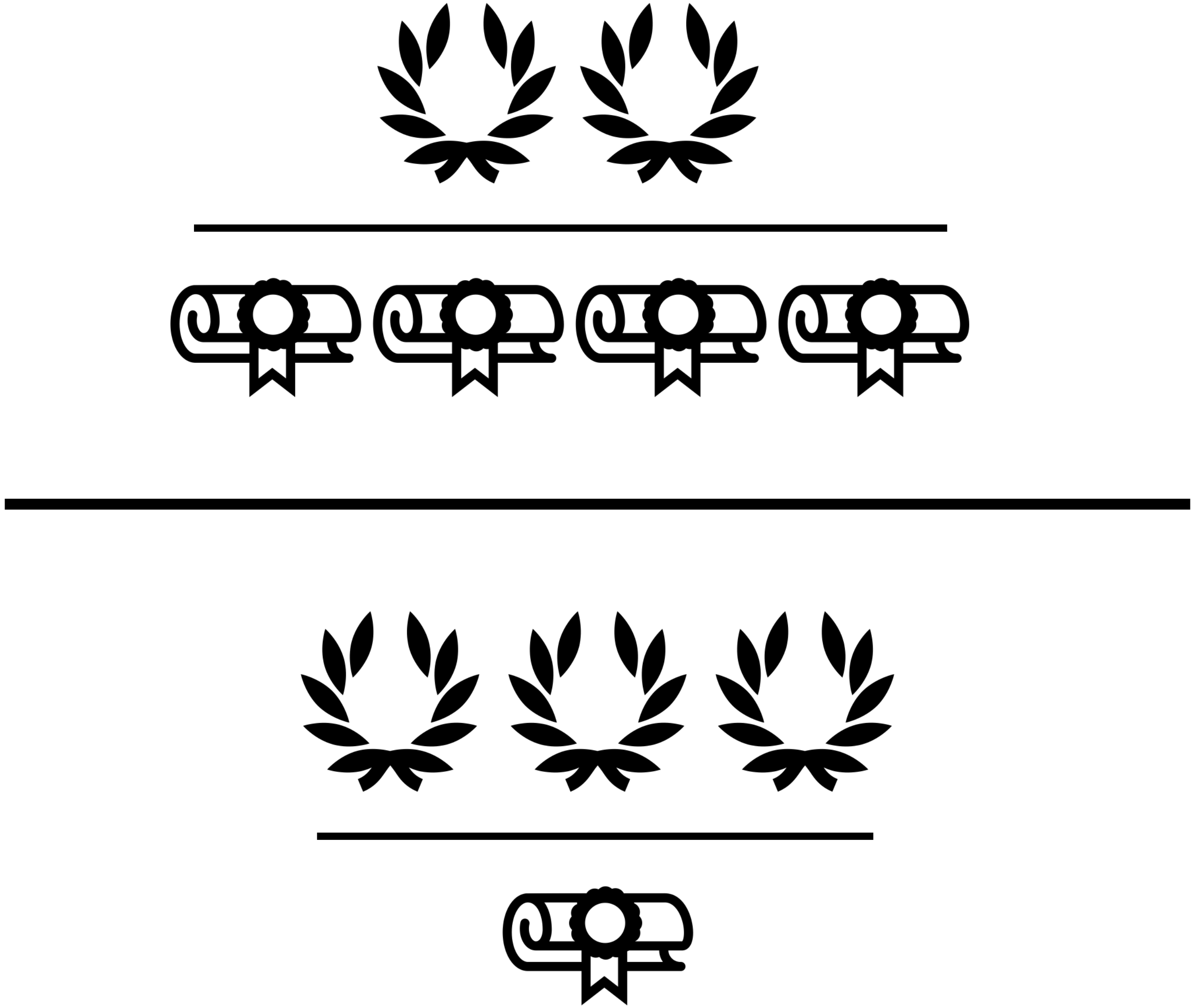
$$Odds = \frac{something\ happening}{something\ not\ happening}$$



$$= \frac{2}{4} \quad = 0.5$$

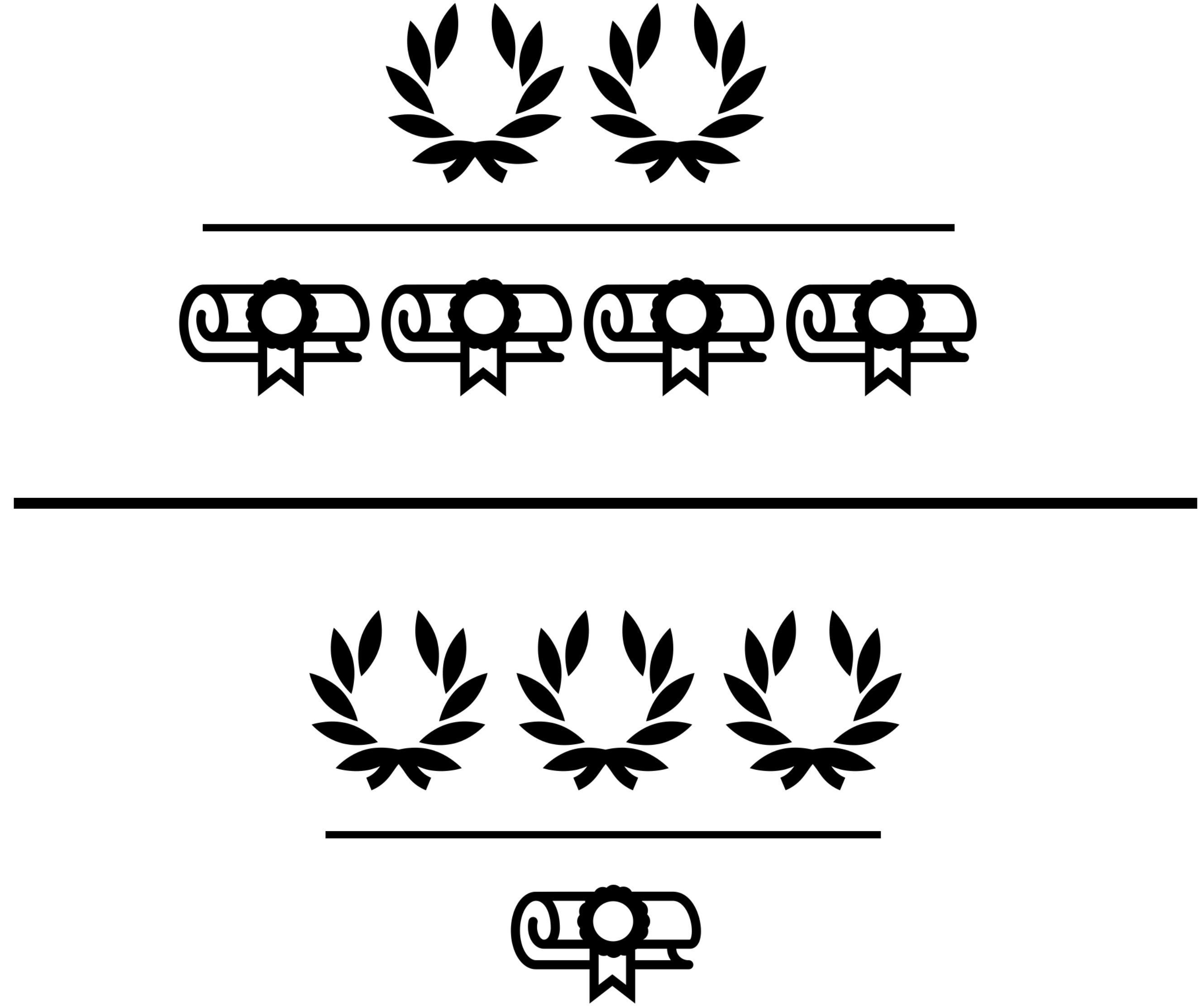
Odds Ratios

Odds ratio



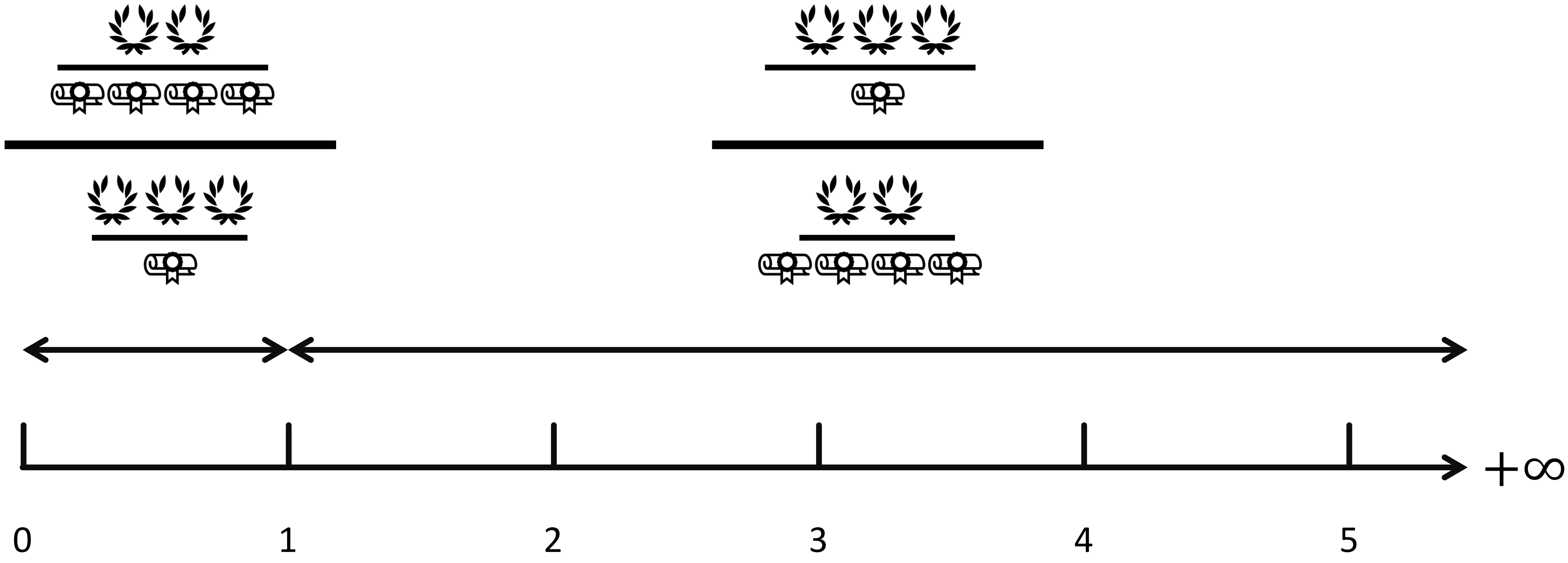
$$= \frac{2/4}{3/1}$$

Odds Ratios



$$= \frac{2/4}{3/1} \quad = 0.17$$

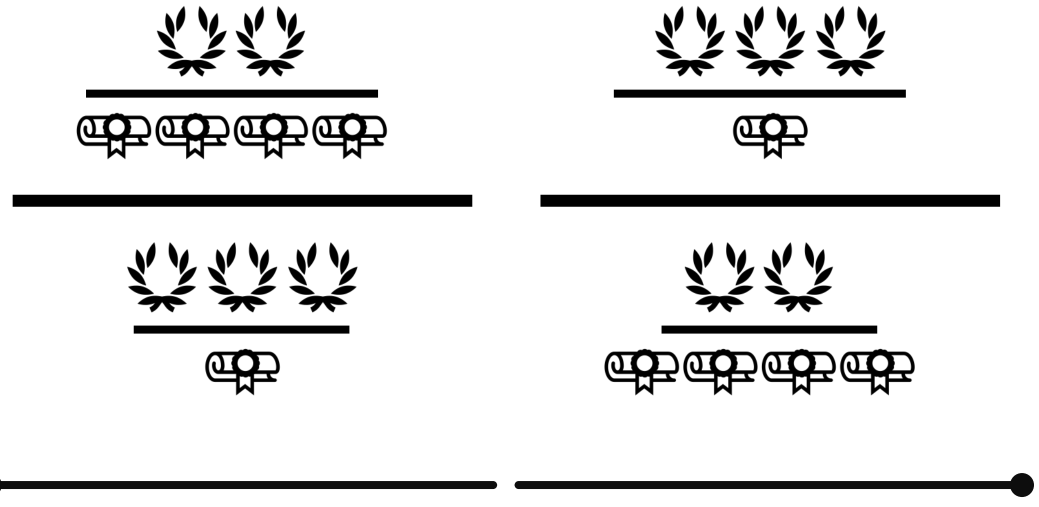
Odds Ratios



Odds Ratios

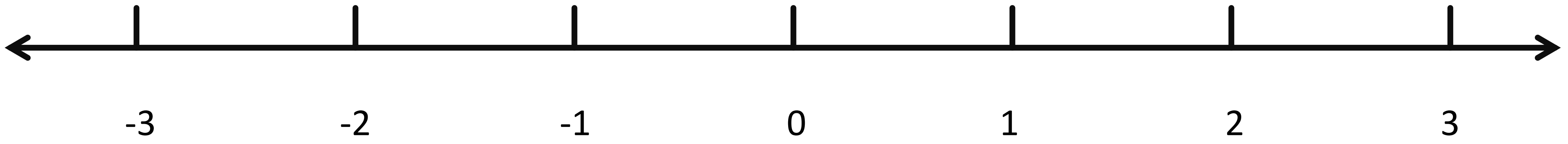
$$Odds\ ratio = \log\left(\frac{2/4}{3/1}\right)$$

$$= -1.79$$



$$Odds\ ratio = \log\left(\frac{3/1}{2/4}\right)$$

$$= 1.79$$



Odds Ratios

Has the mutated gene	Has Cancer ✓	Has Cancer ✗	
✓	23	117	140
✗	6	210	216
	29	327	

Can we use odds ratio to determine if there is a **relationship** between the mutated gene and cancer? If someone has the mutated gene, are odds higher that they will get cancer?

Odds Ratios

Has the mutated gene	Has Cancer ✓	Has Cancer ✗	
✓	23	117	140
✗	6	210	216
	29	327	

$$\frac{\frac{23}{117}}{\frac{6}{210}} = \frac{0.2}{0.03} = 6.88$$ odds ratio

$$\log(6.88) = 1.93$$ log(odds ratio)

Larger values mean that the mutated gene is a good predictor of cancer. Smaller values mean that the mutated gene is not a good predictor of cancer.

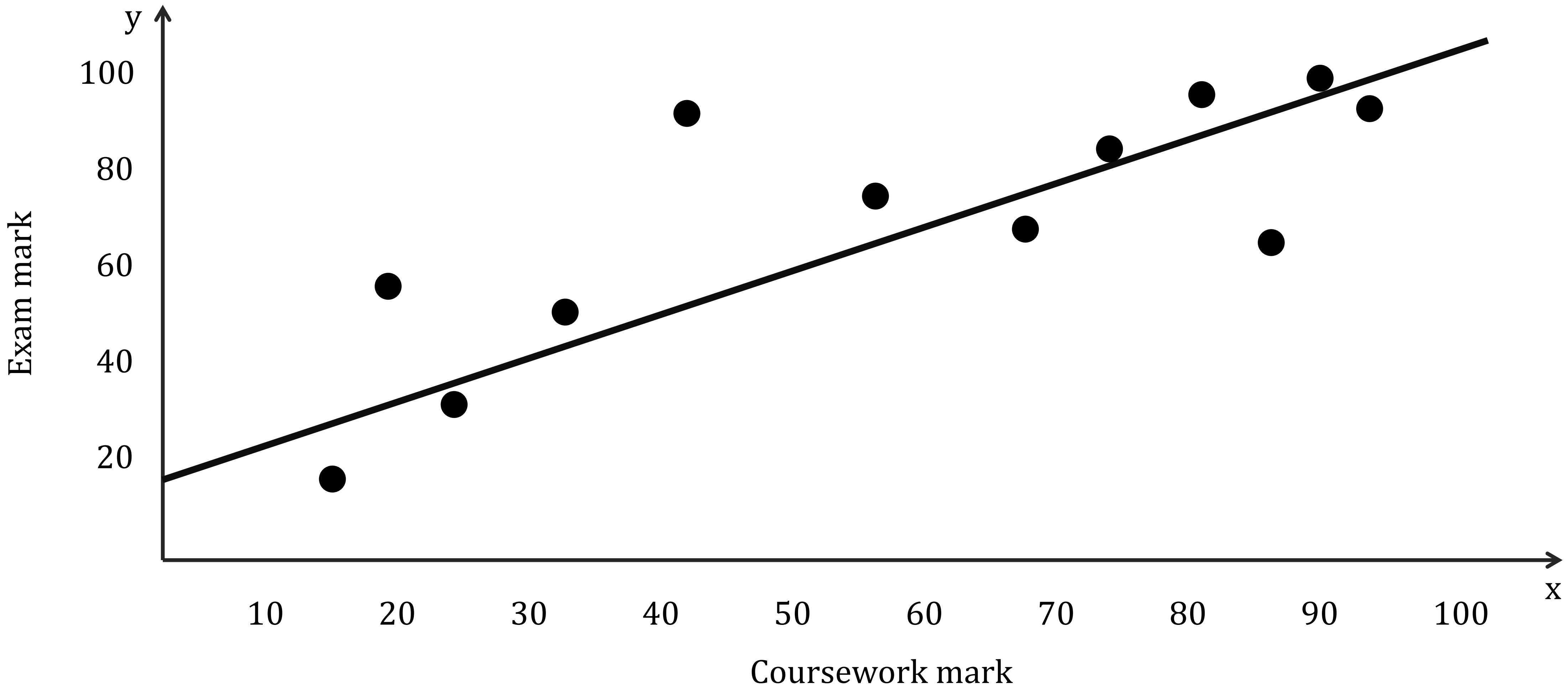
Today

- Odds
- Logistic Regression

Logistic Regression

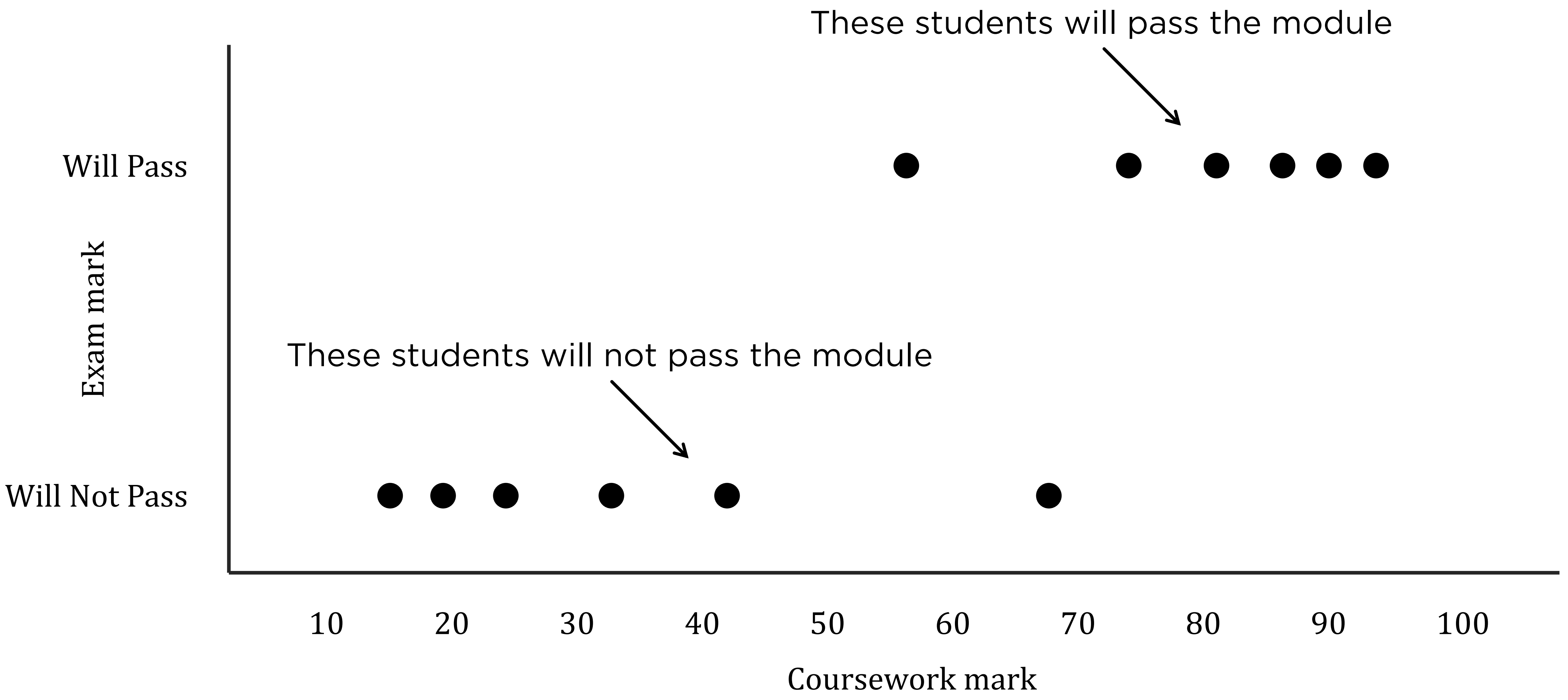
Logistic Regression

- is similar to **Linear Regression**, except…



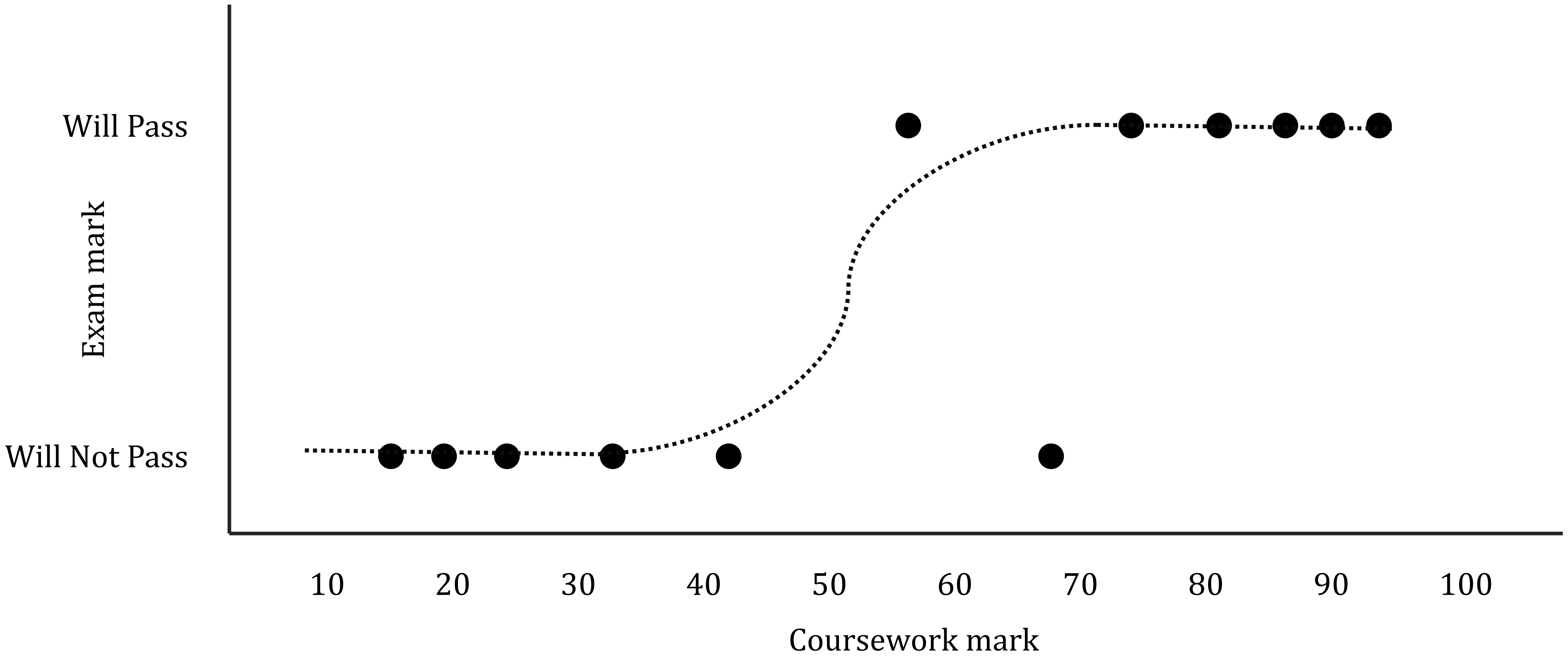
Logistic Regression

- predicts **True** / **False**



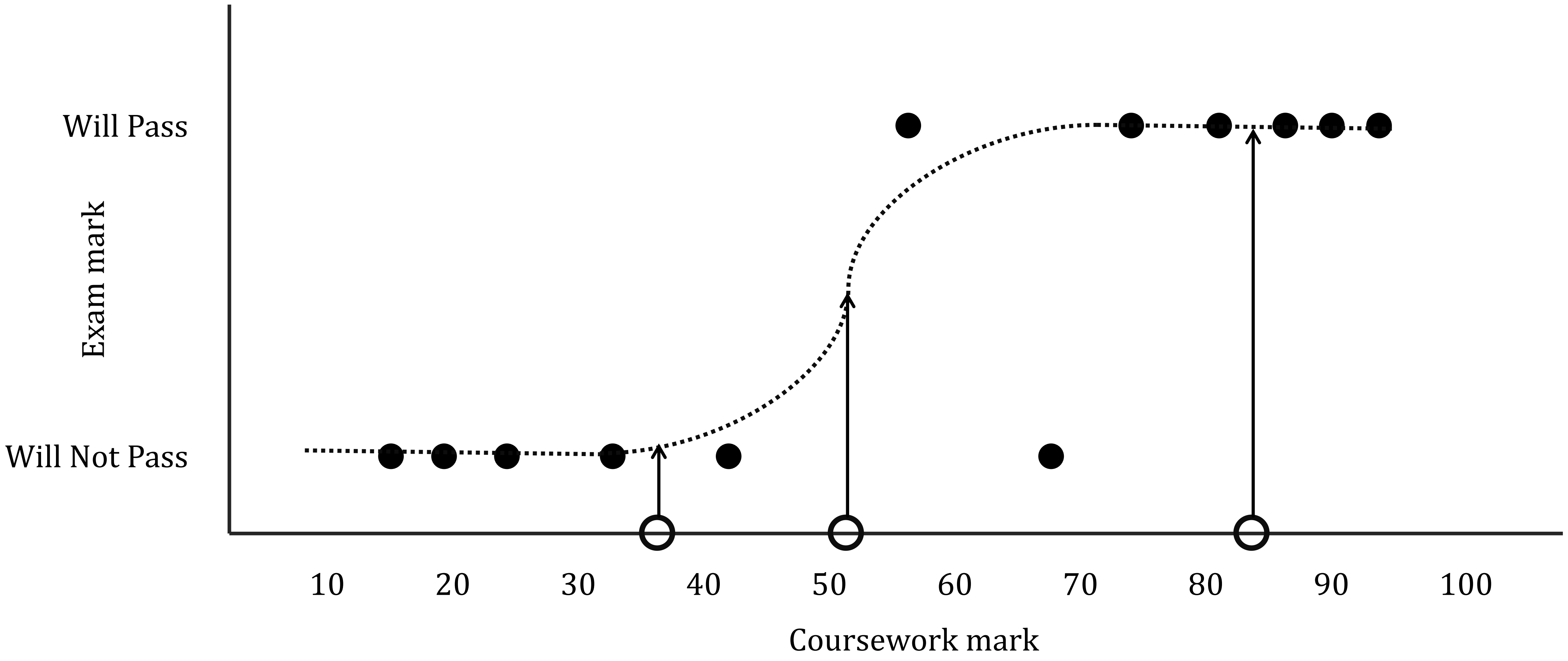
Logistic Regression

- fits an “S” shaped “**logistic function**”



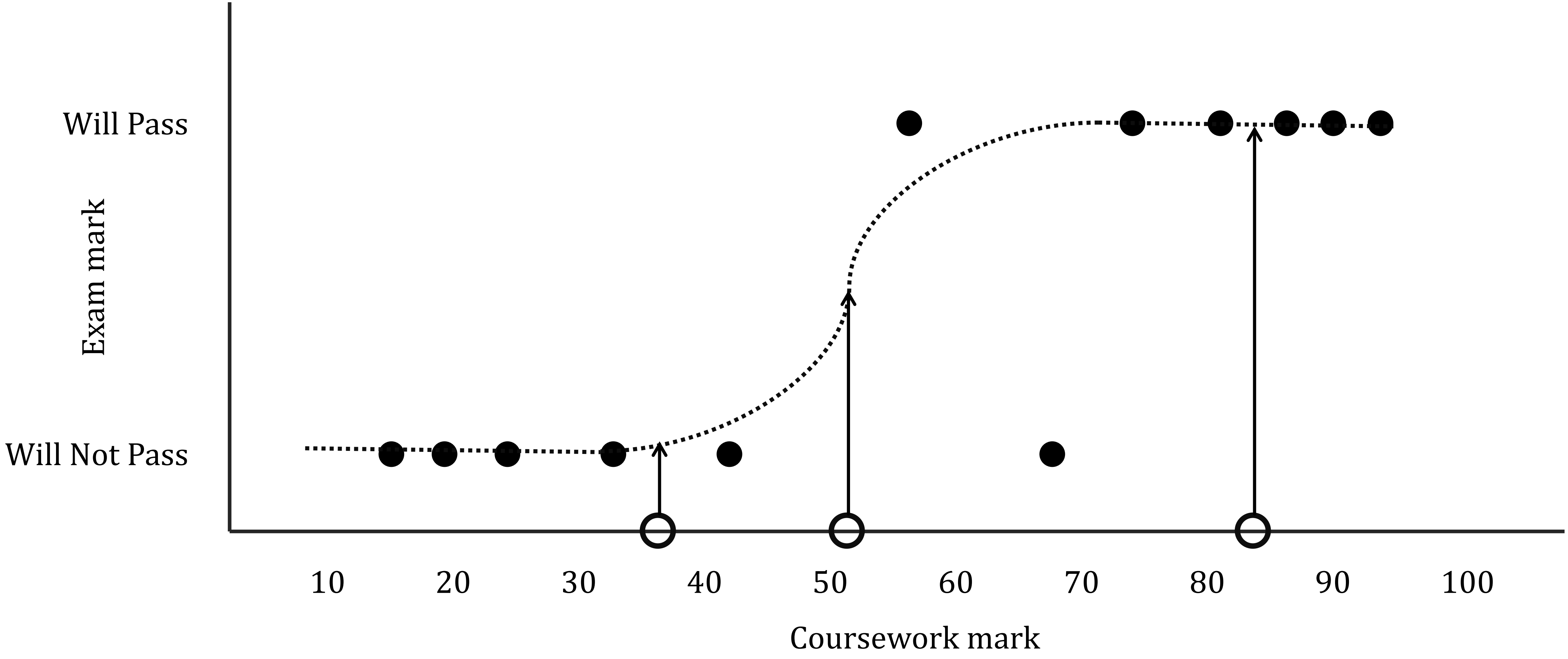
Logistic Regression

- its **curves** goes from 0 to 1



Logistic Regression

- is used for **prediction**



Logistic Regression

- like with Linear Regression, we can make simple models:

  Exam result is predicted by **coursework mark**

- or more complicated models:

  Exam result is predicted by **coursework mark**

  + lecture attendance

  } **continuous data**

  + is ~~sunny~~

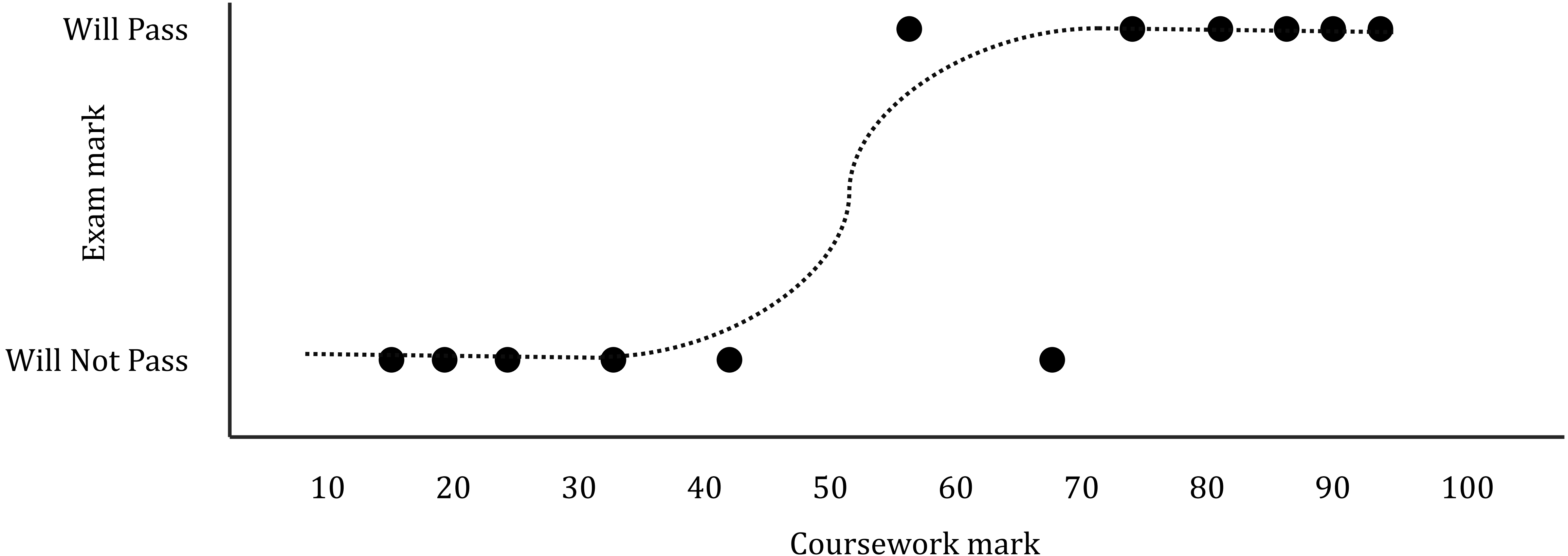
  + went to party

  } **discrete data**

~~vs~~ Exam result is predicted by coursework mark + lecture attendance

Logistic Regression

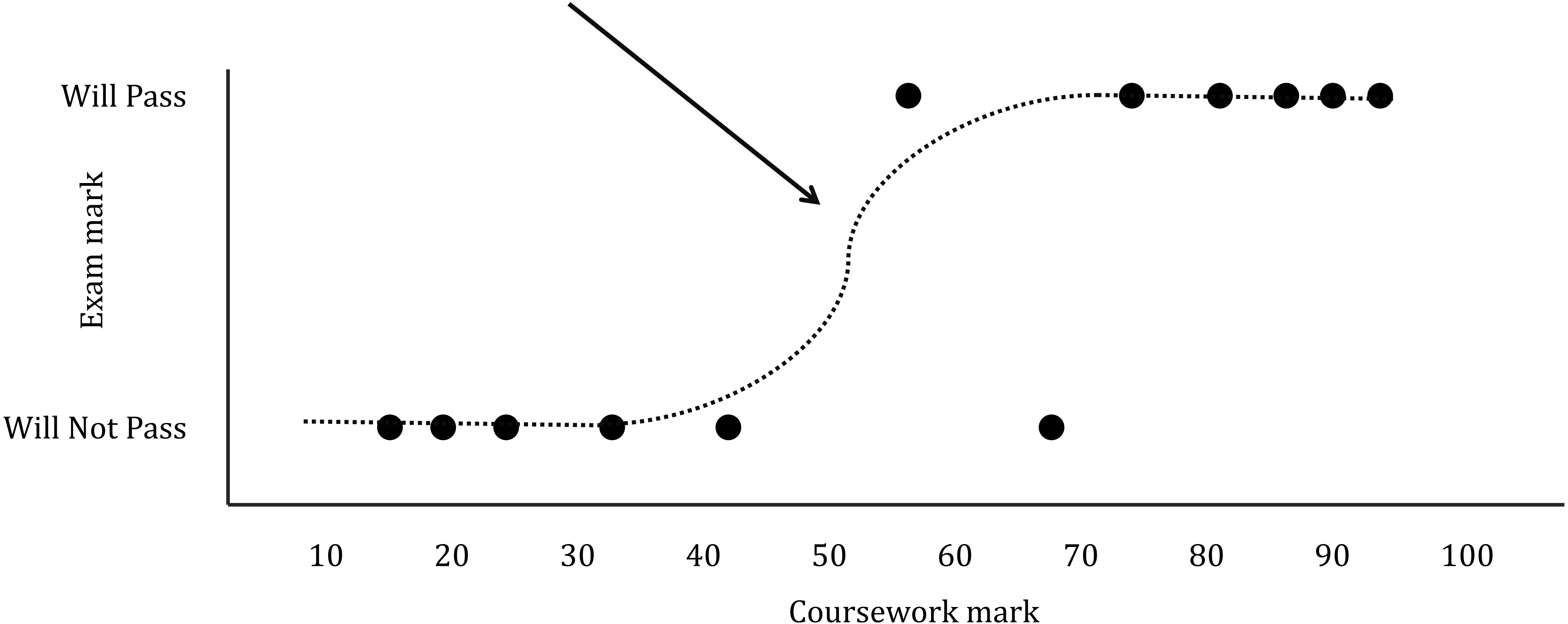
- Logistic Regression provides probabilities and classifies new samples using **continuous** and **discrete** measurements.
- A popular machine learning method.



Logistic Regression

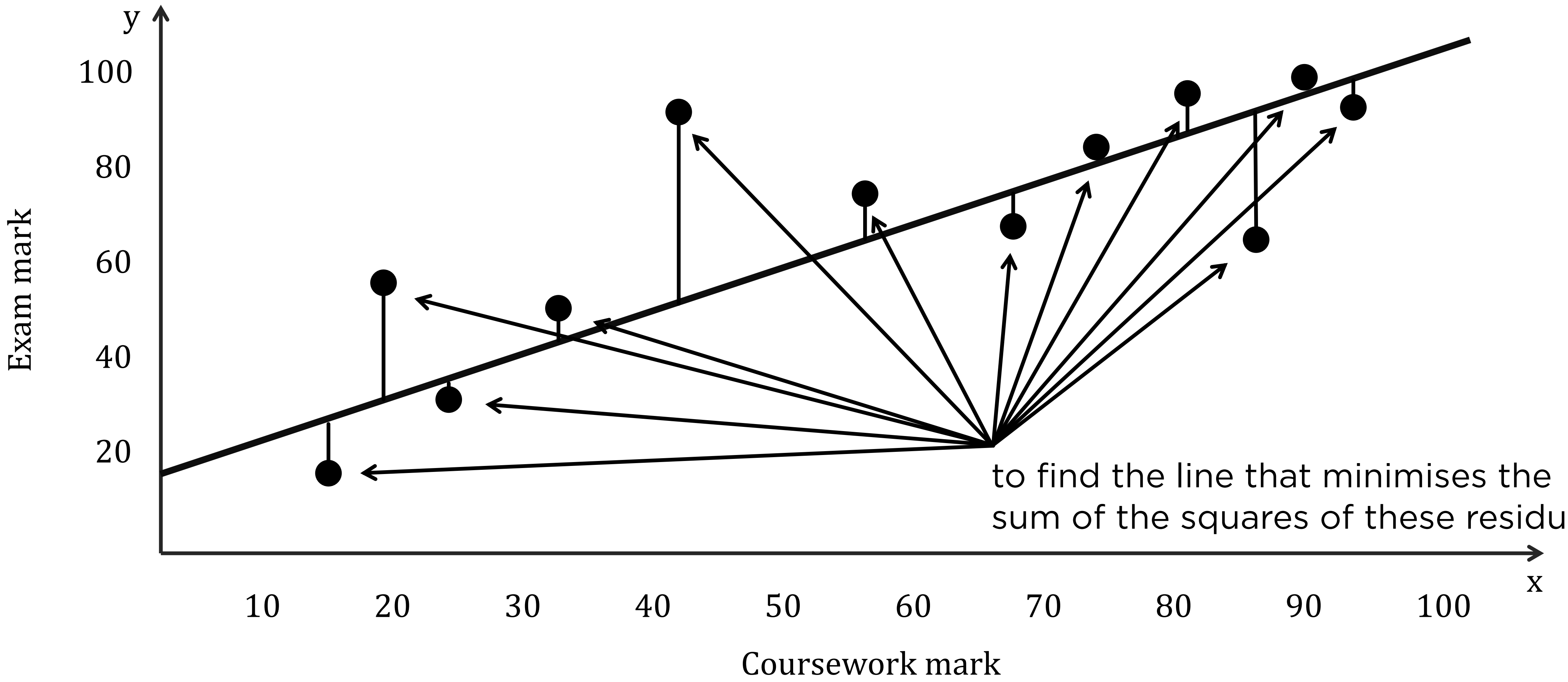
- Logistic Regression vs Linear Regression

  How the line is fit to the data



Logistic Regression

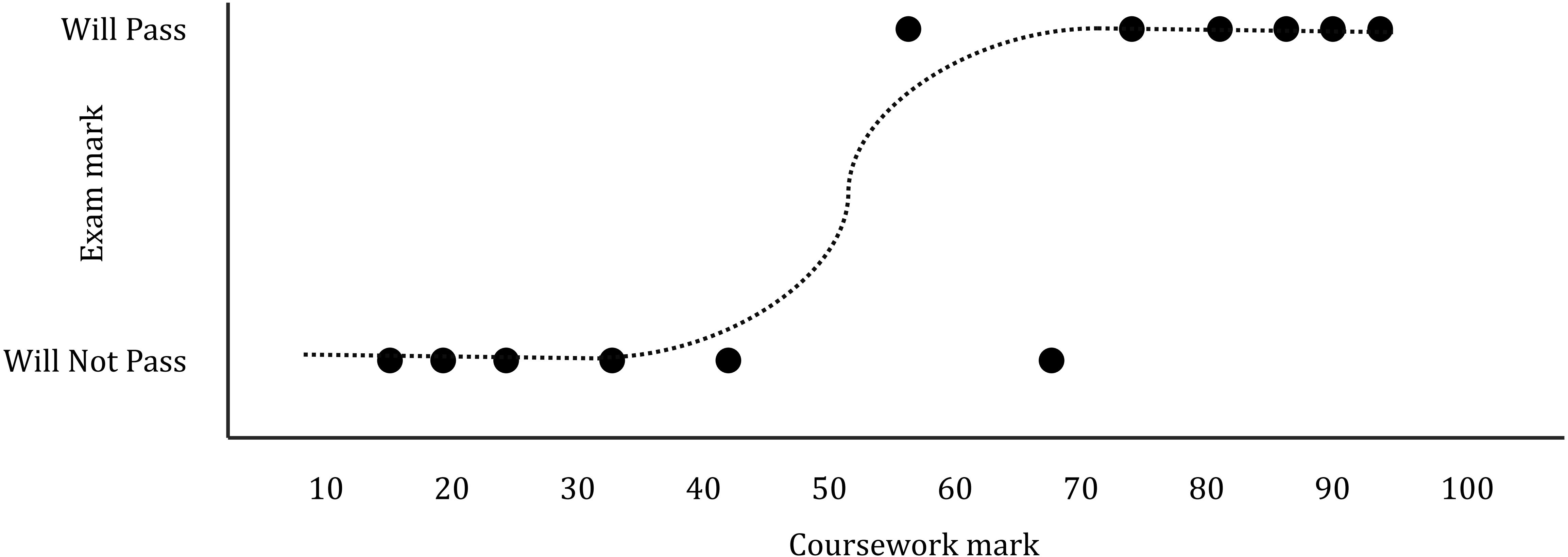
Linear Regression



to find the line that minimises the sum of the squares of these residuals.

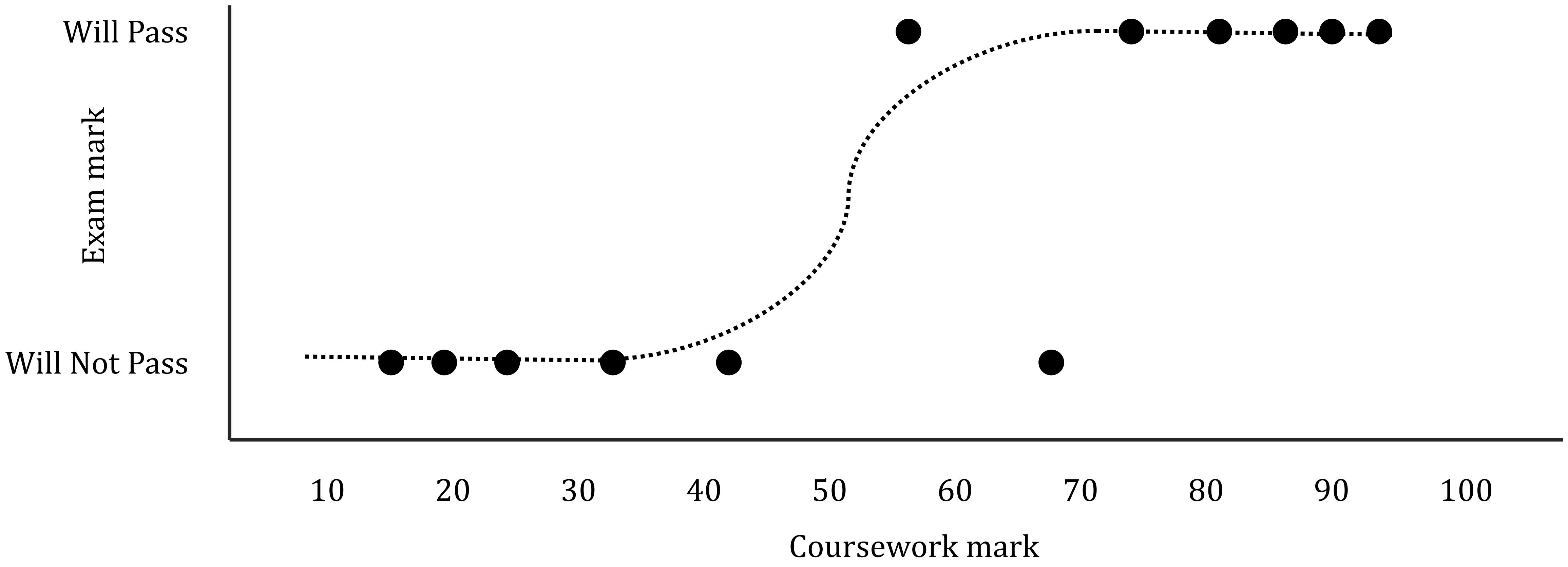
Logistic Regression

Does not have concept of a “Residual”, so it cannot use least squares and cannot calculate R^2.



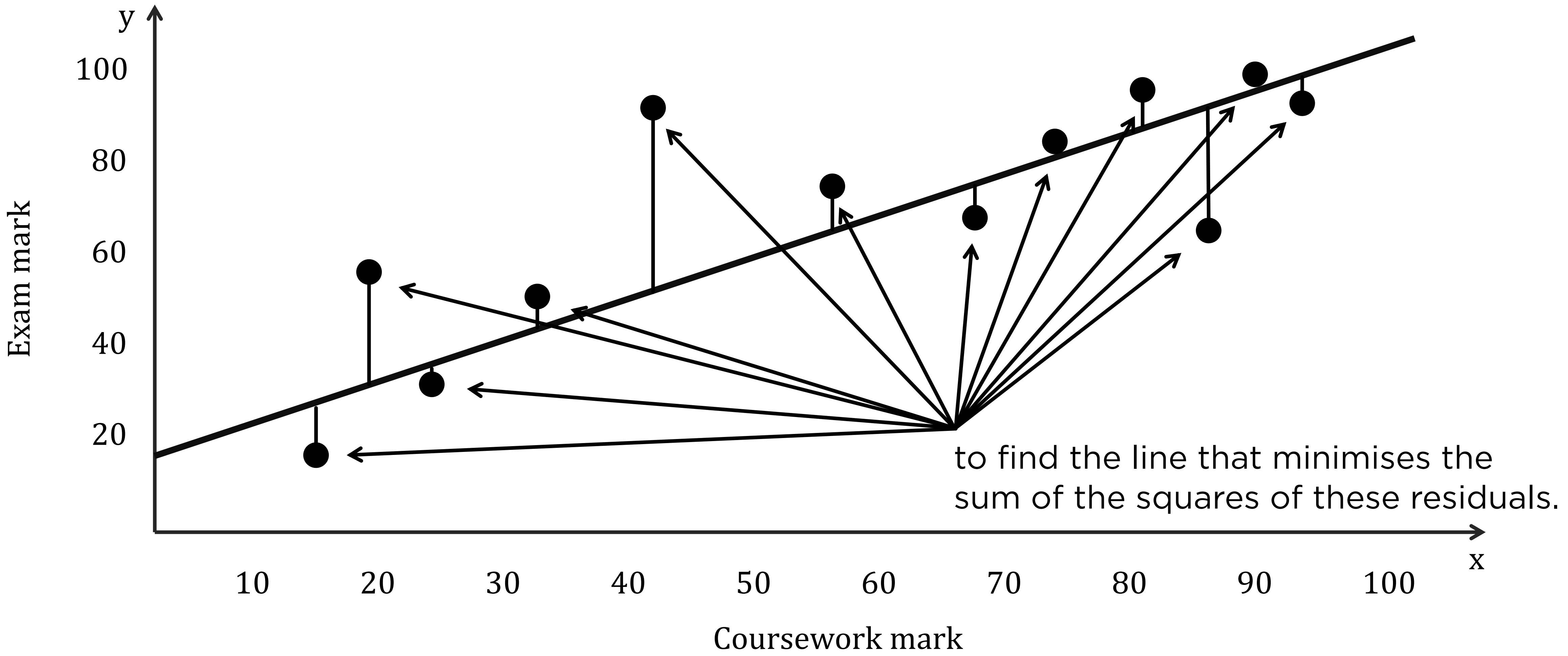
Logistic Regression

How is this squiggle optimised to fit the data the best? - **Maximum Likelihood**



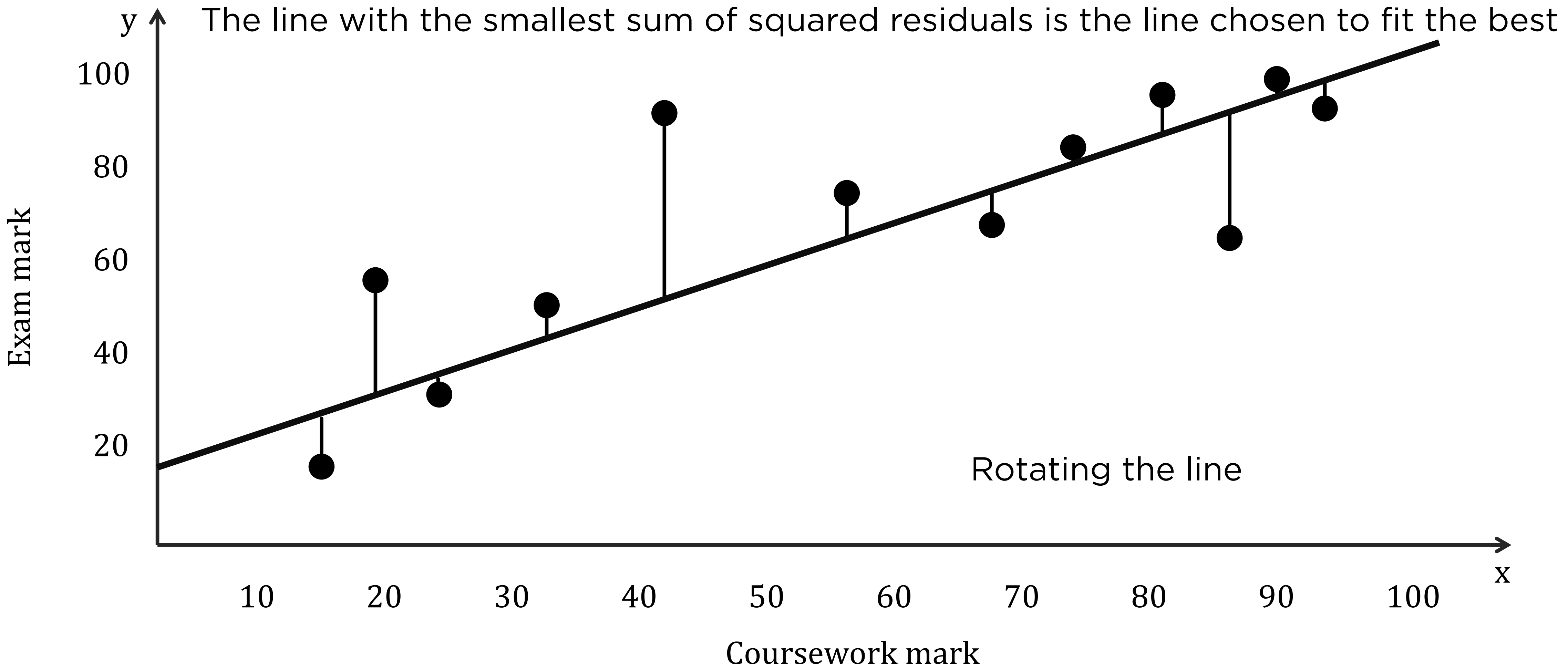
Logistic Regression

How are lines fit in linear regression?

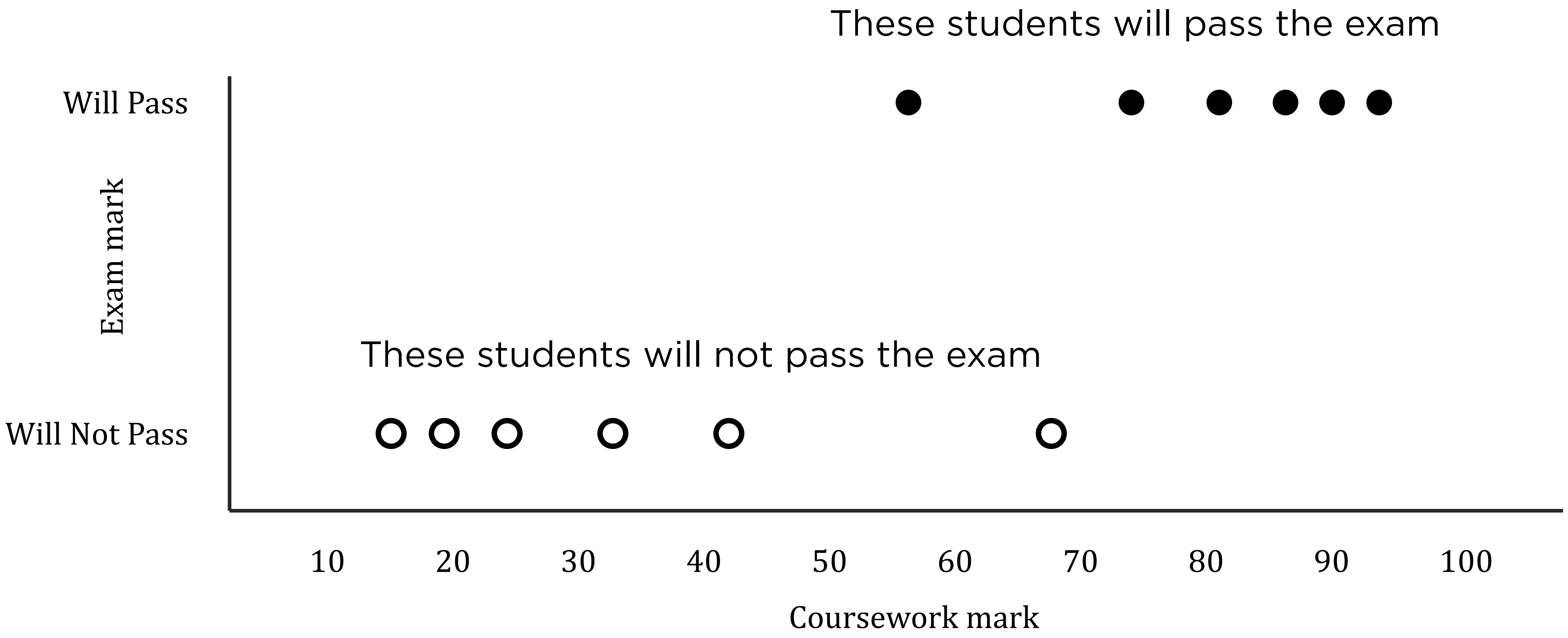


Logistic Regression

How are lines fit in linear regression?

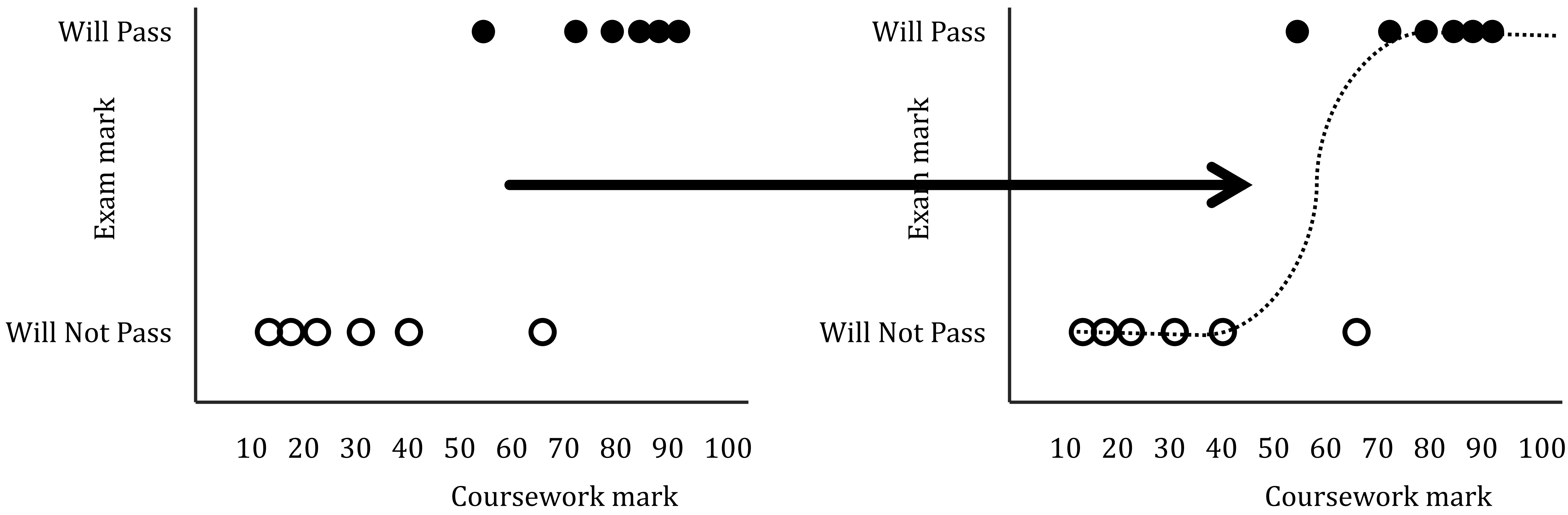


Logistic Regression



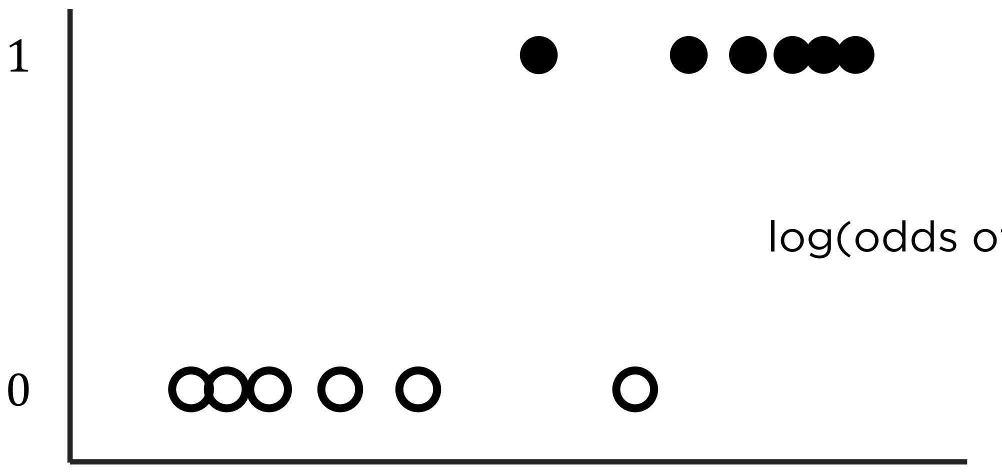
Logistic Regression

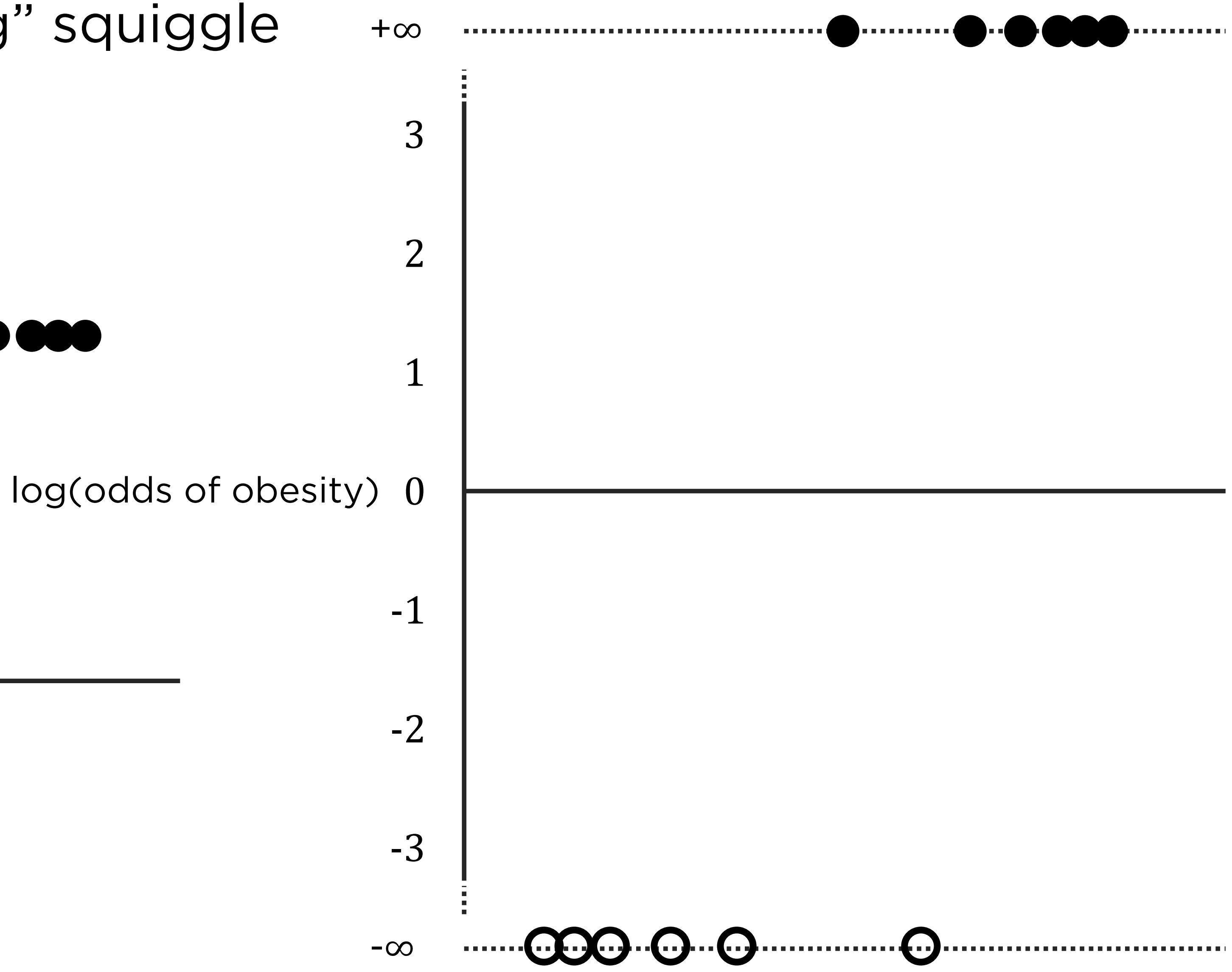
To draw the "best fitting" squiggle



Logistic Regression

To draw the "best fitting" squiggle





Logistic Regression

To draw the ”best fitting” squiggle

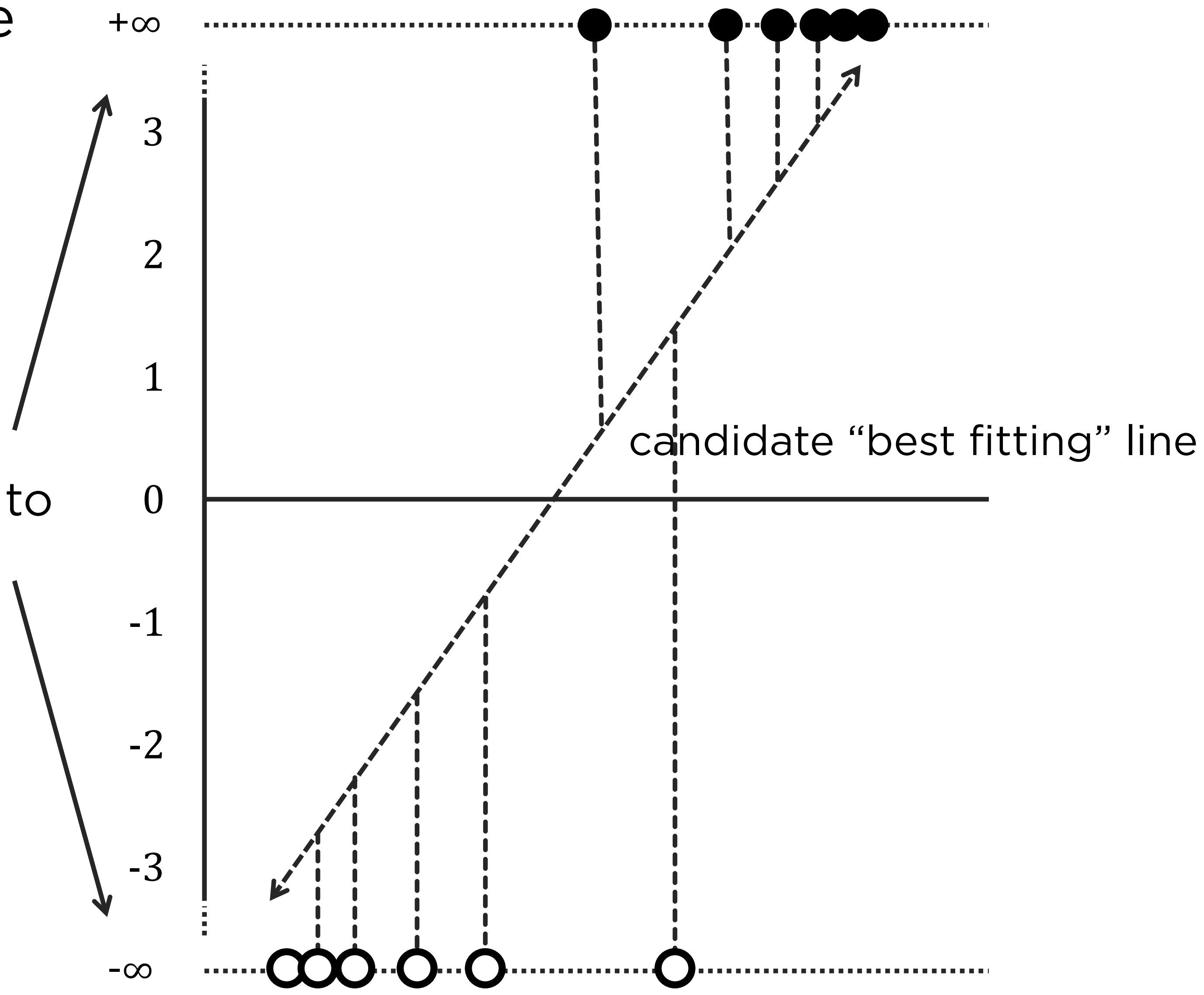
The only problem…

The transformation pushes the raw data to

The residuals are equal to +∞ and -∞

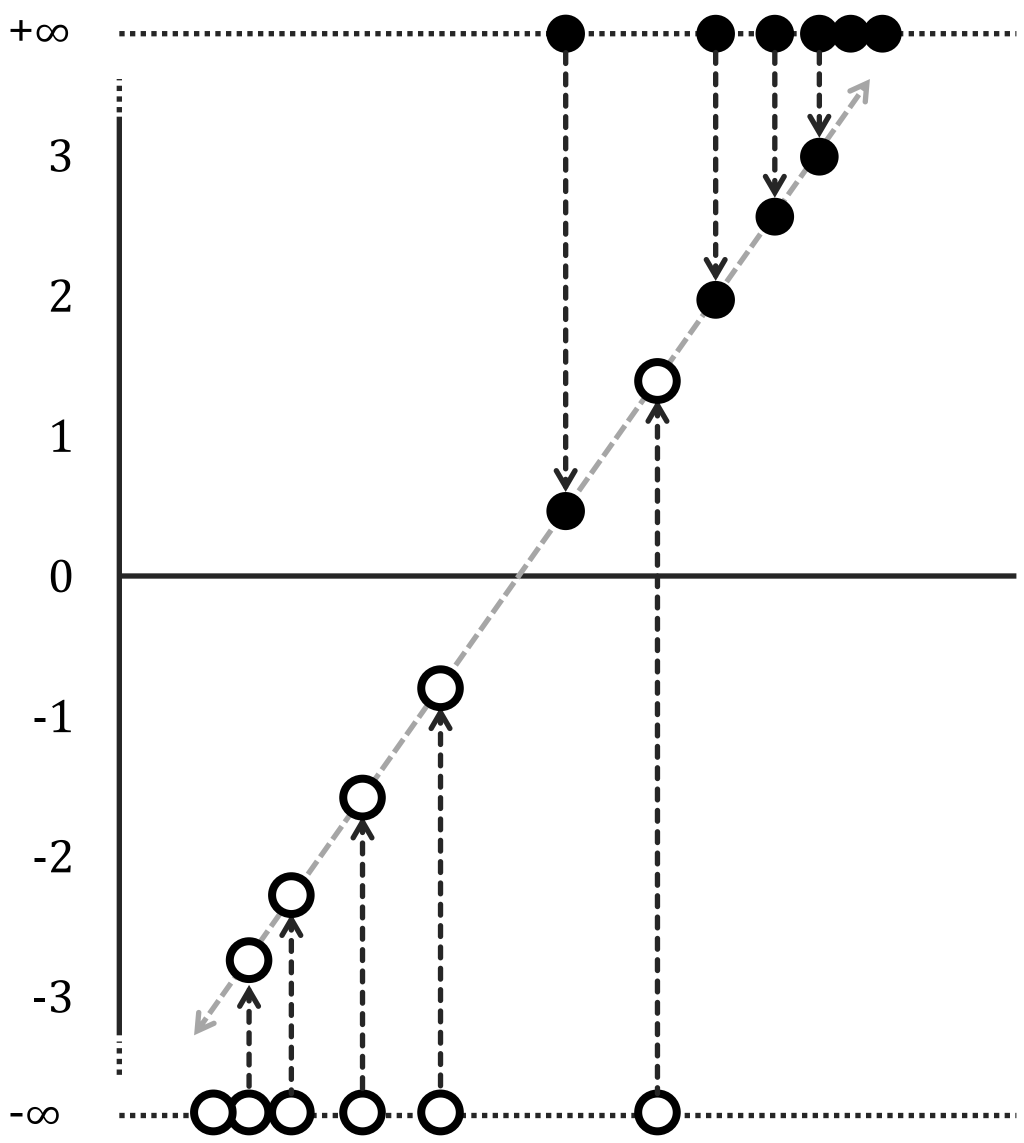
So, we cannot use least-squares
to find the best fitting line ☹

But we can use **maximum likelihood** ☺



Logistic Regression

To draw the "best fitting" squiggle

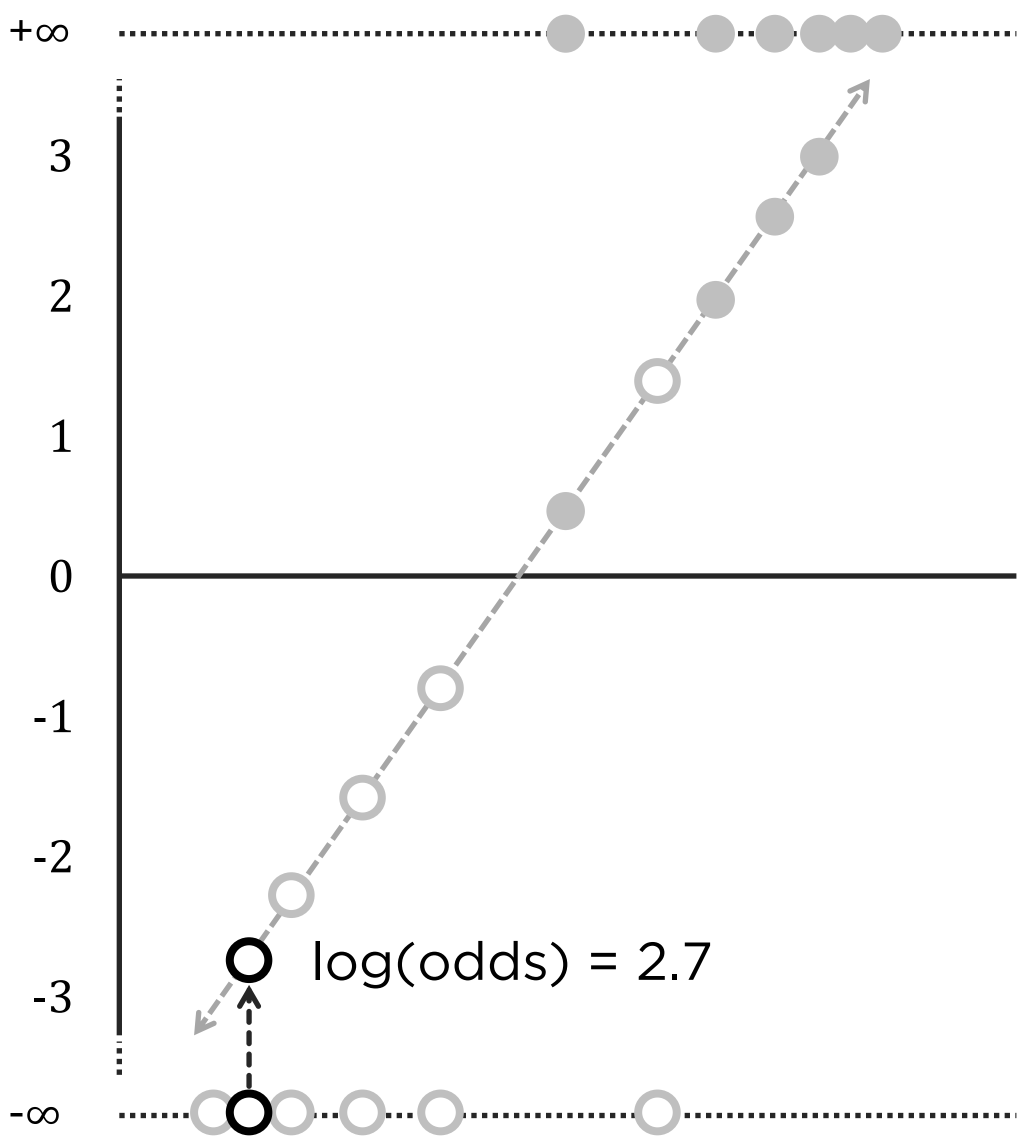


First, project the original data points onto the candidate line. This gives each sample a candidate log(odds) value.

Logistic Regression

To draw the "best fitting" squiggle

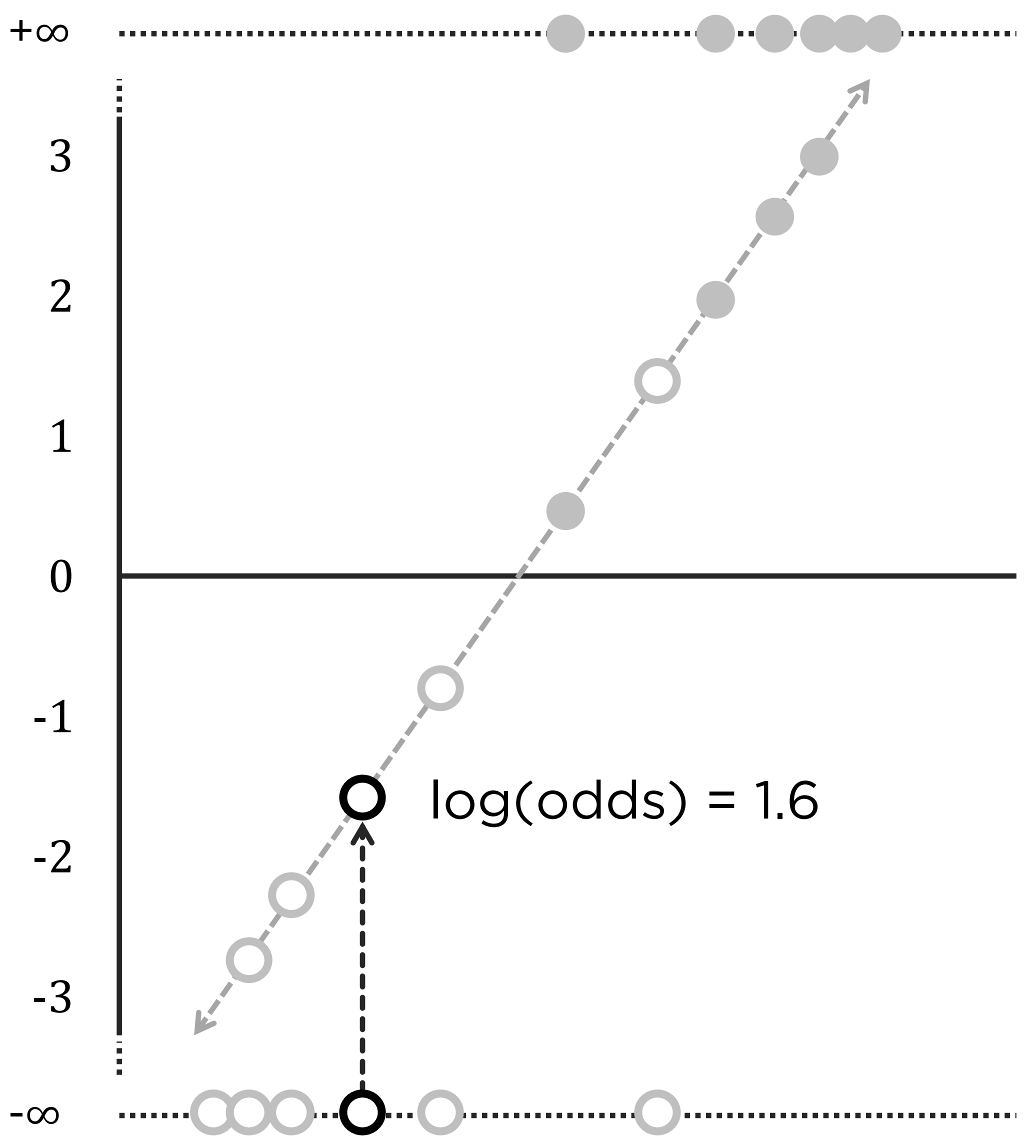
First, project the original data points onto the candidate line. This gives each sample a candidate log(odds) value.



Logistic Regression

To draw the "best fitting" squiggle

First, project the original data points onto the candidate line. This gives each sample a candidate log(odds) value.

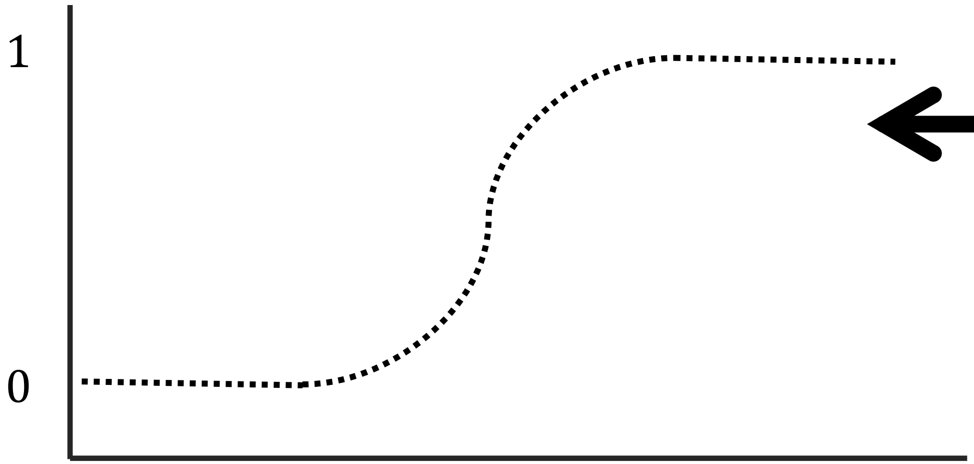


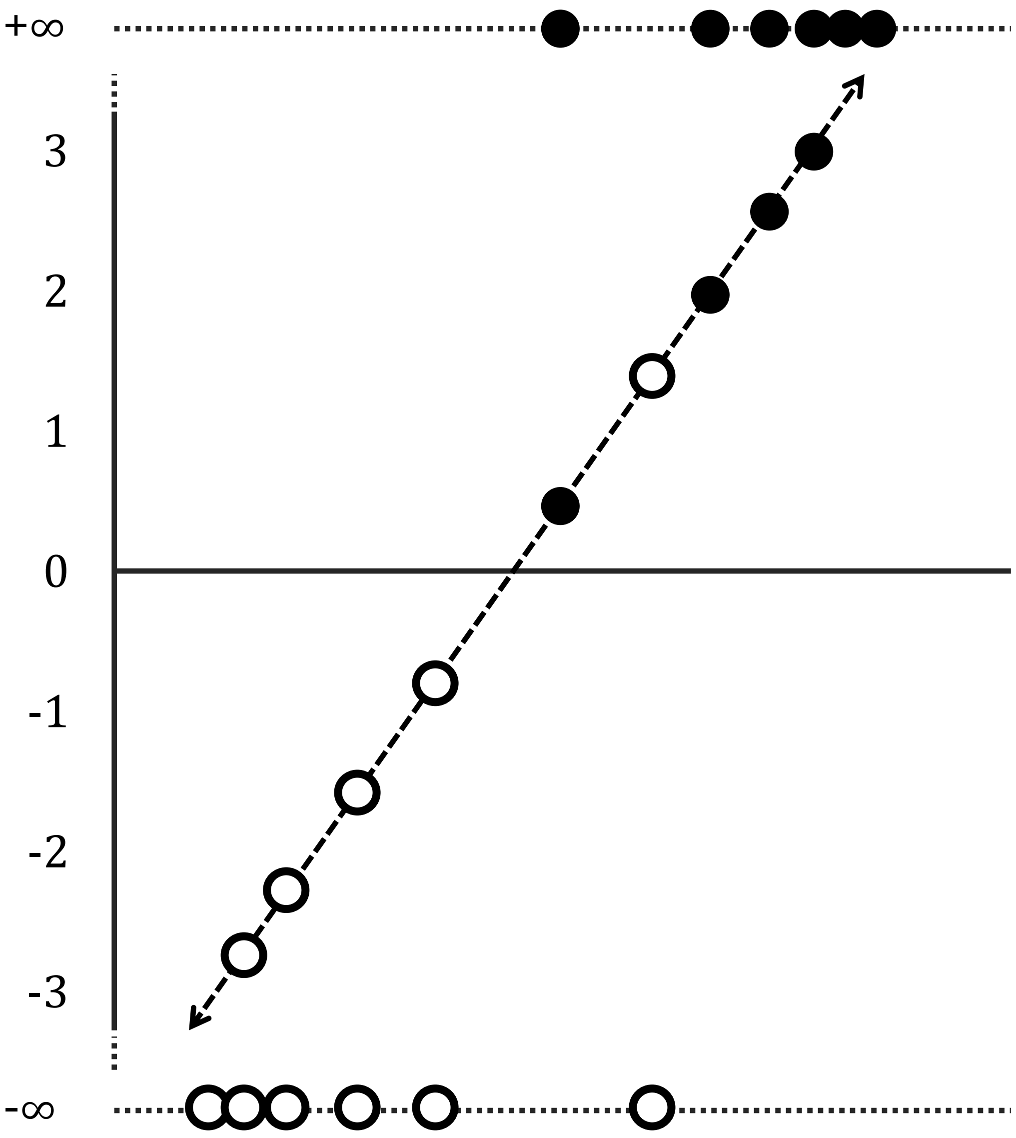
Logistic Regression

To draw the "best fitting" squiggle

Then, transform the candidate log(odds) to candidate probabilities using:

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



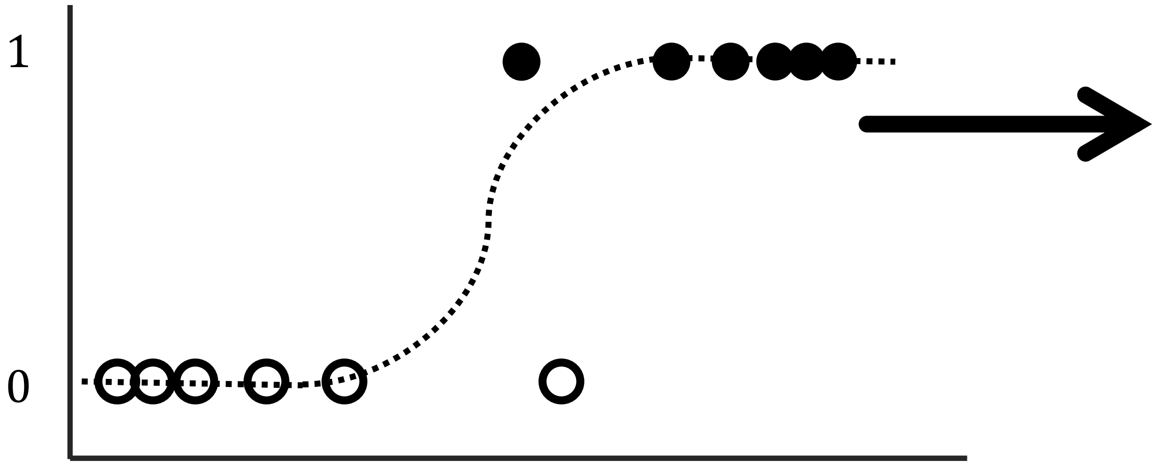


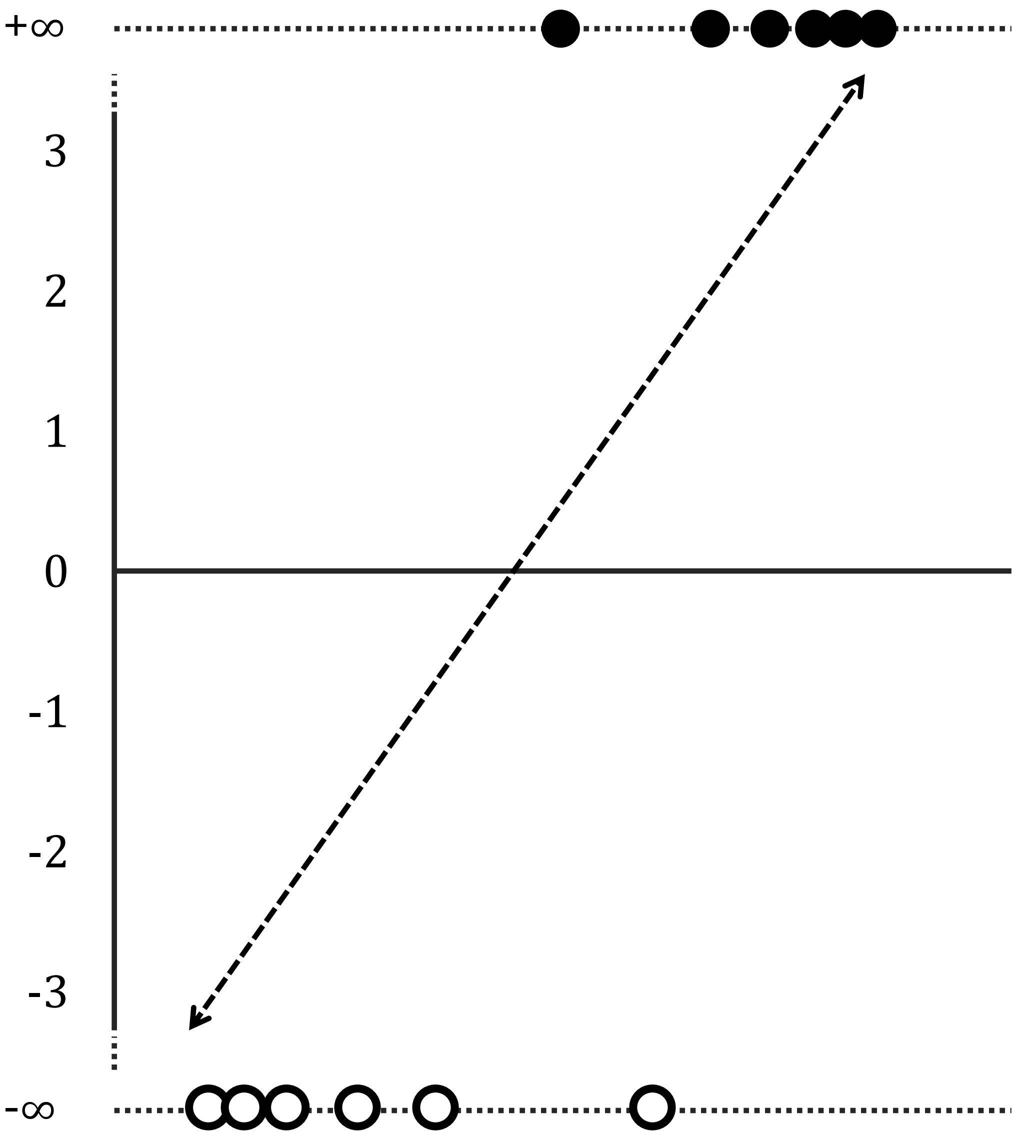
Logistic Regression

To draw the "best fitting" squiggle

Then, transform the candidate log(odds) to candidate probabilities using:

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



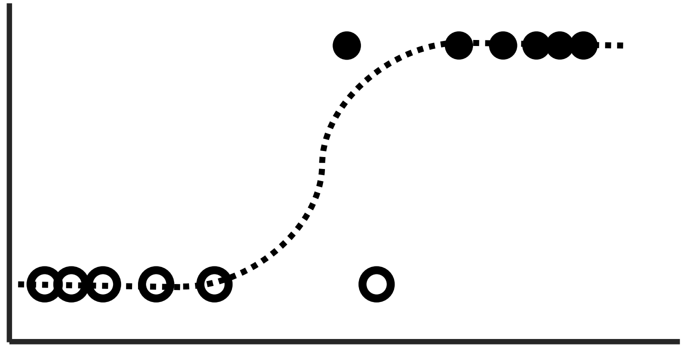


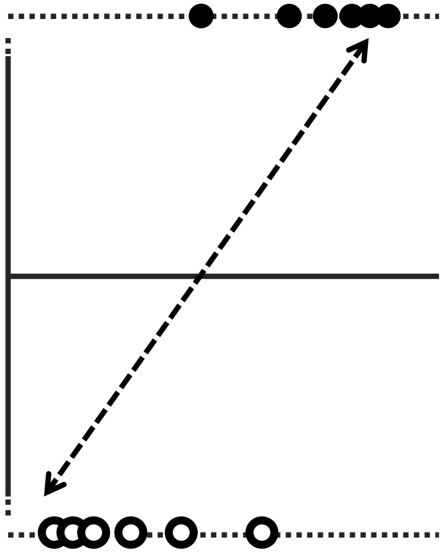
How to convert the equation…

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

Input: probability

Output: log(odds)



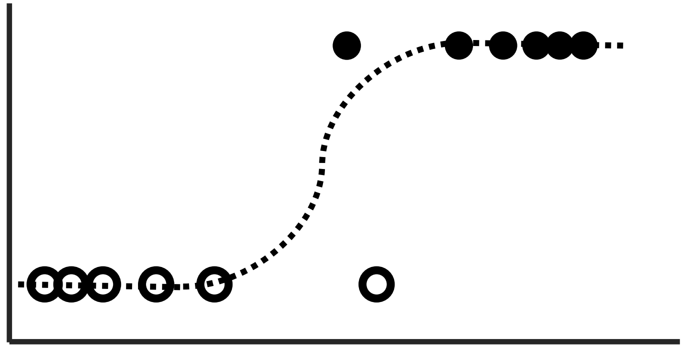


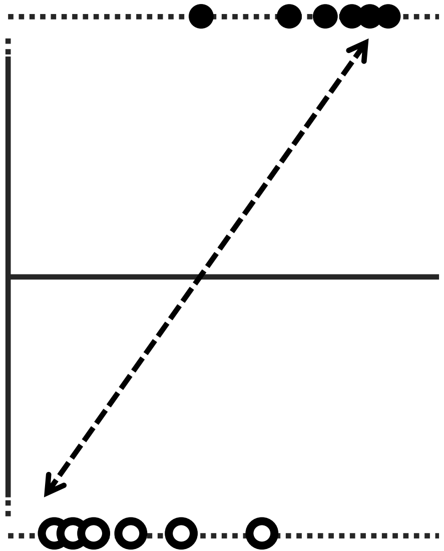
to

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

Input: log(odds)

Output: probability

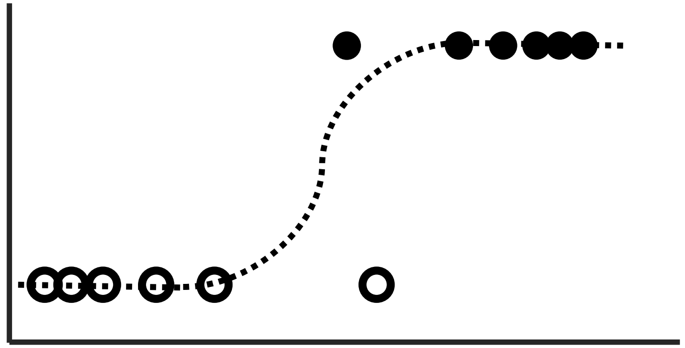


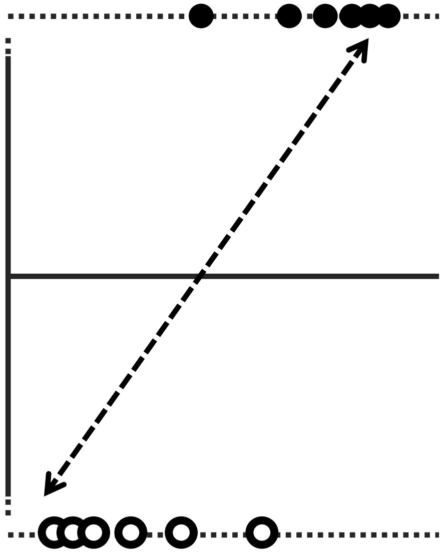


How to convert the equation…

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

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



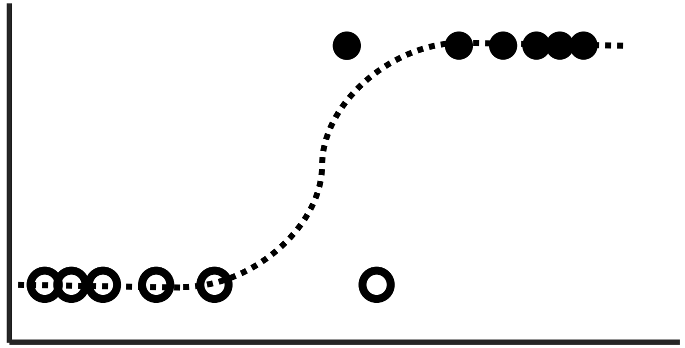


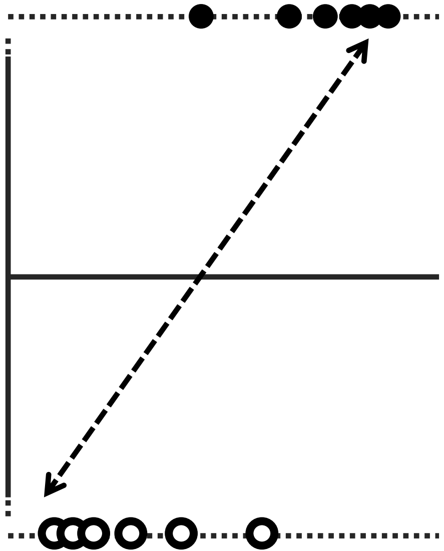
$$p = (1-p)e^{\log(odds)} = e^{\log(odds)} - pe^{\log(odds)}$$

$$p + pe^{\log(odds)} = e^{\log(odds)}$$

$$p(1 + e^{\log(odds)}) = e^{\log(odds)}$$

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



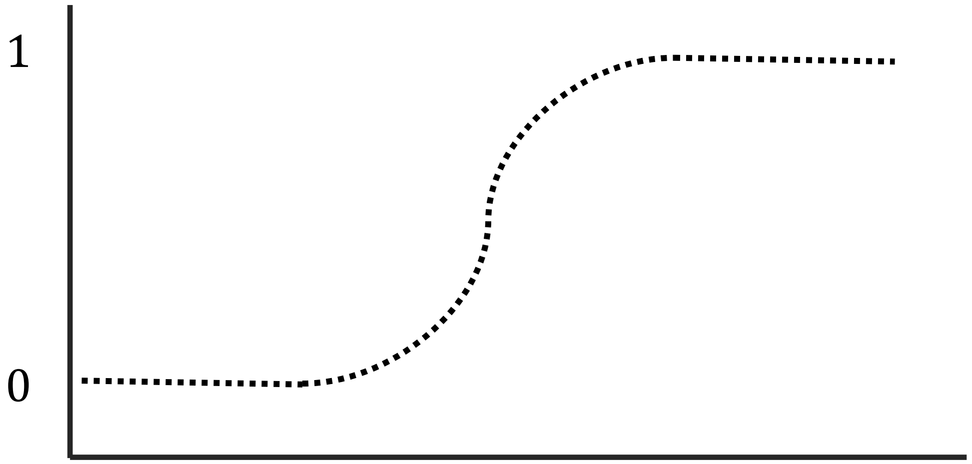


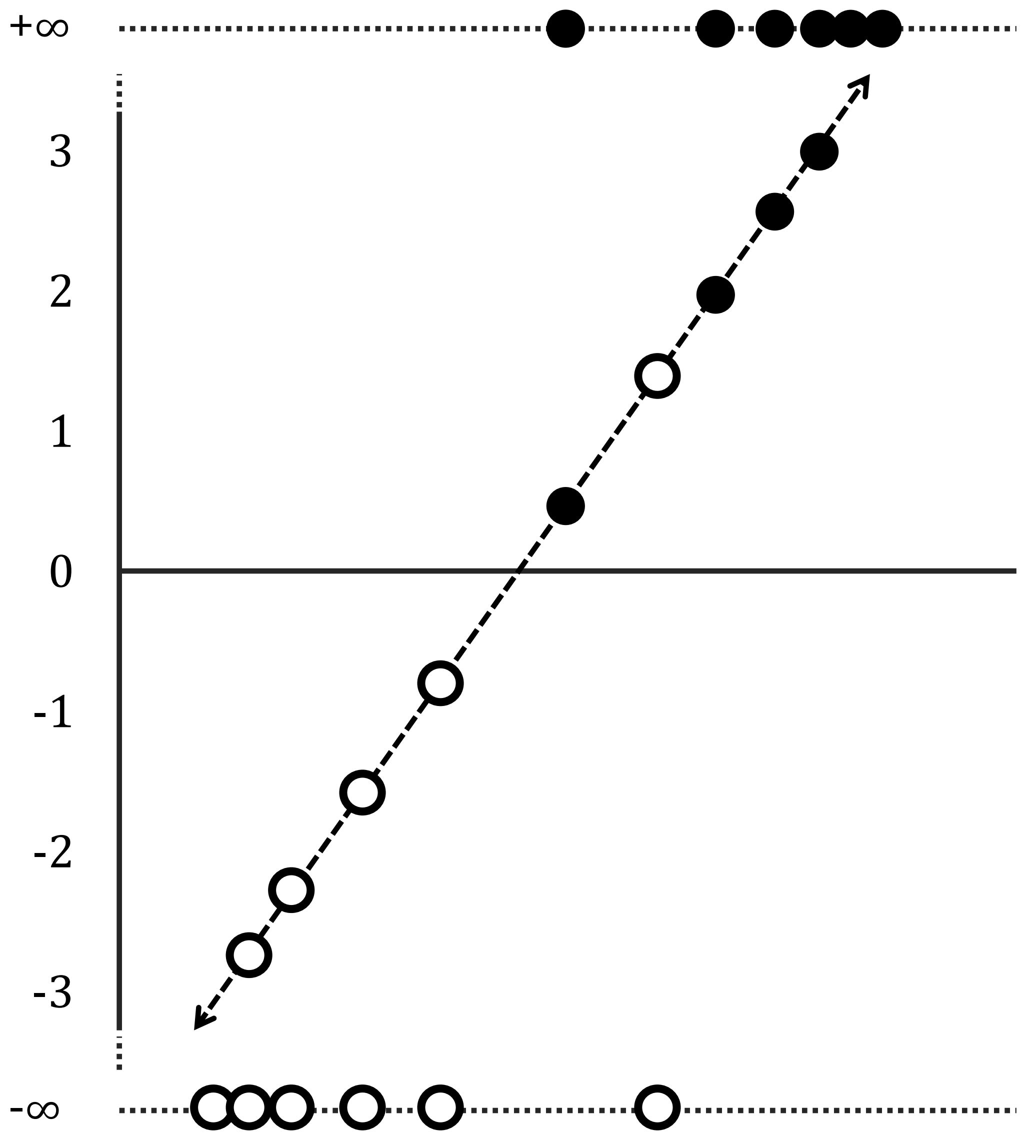
Logistic Regression

To draw the "best fitting" squiggle

Then, transform the candidate log(odds) to candidate probabilities using:

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





Logistic Regression

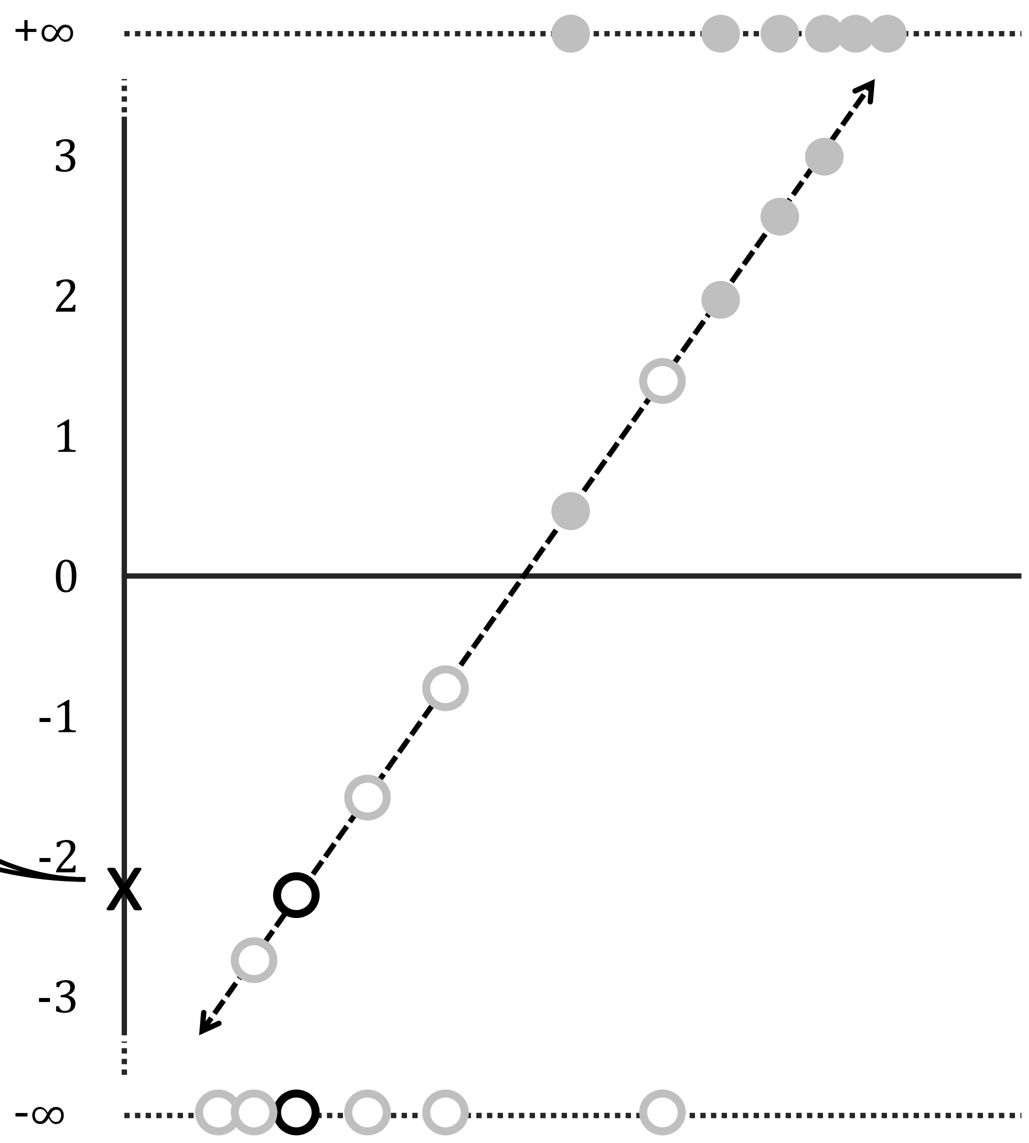
To draw the "best fitting" squiggle

Then, transform the candidate log(odds) to candidate probabilities using:

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

$$p = \frac{e^{-2.1}}{1 + e^{-2.1}}$$

$$p = 0.1$$

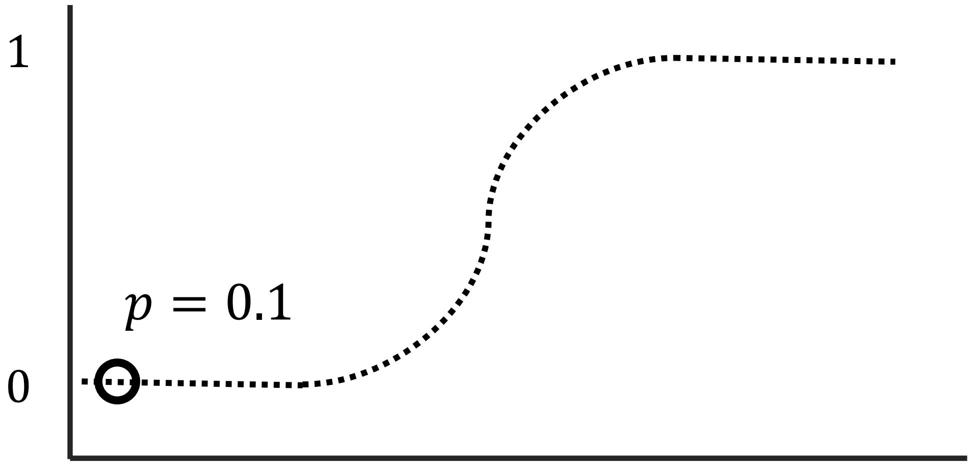


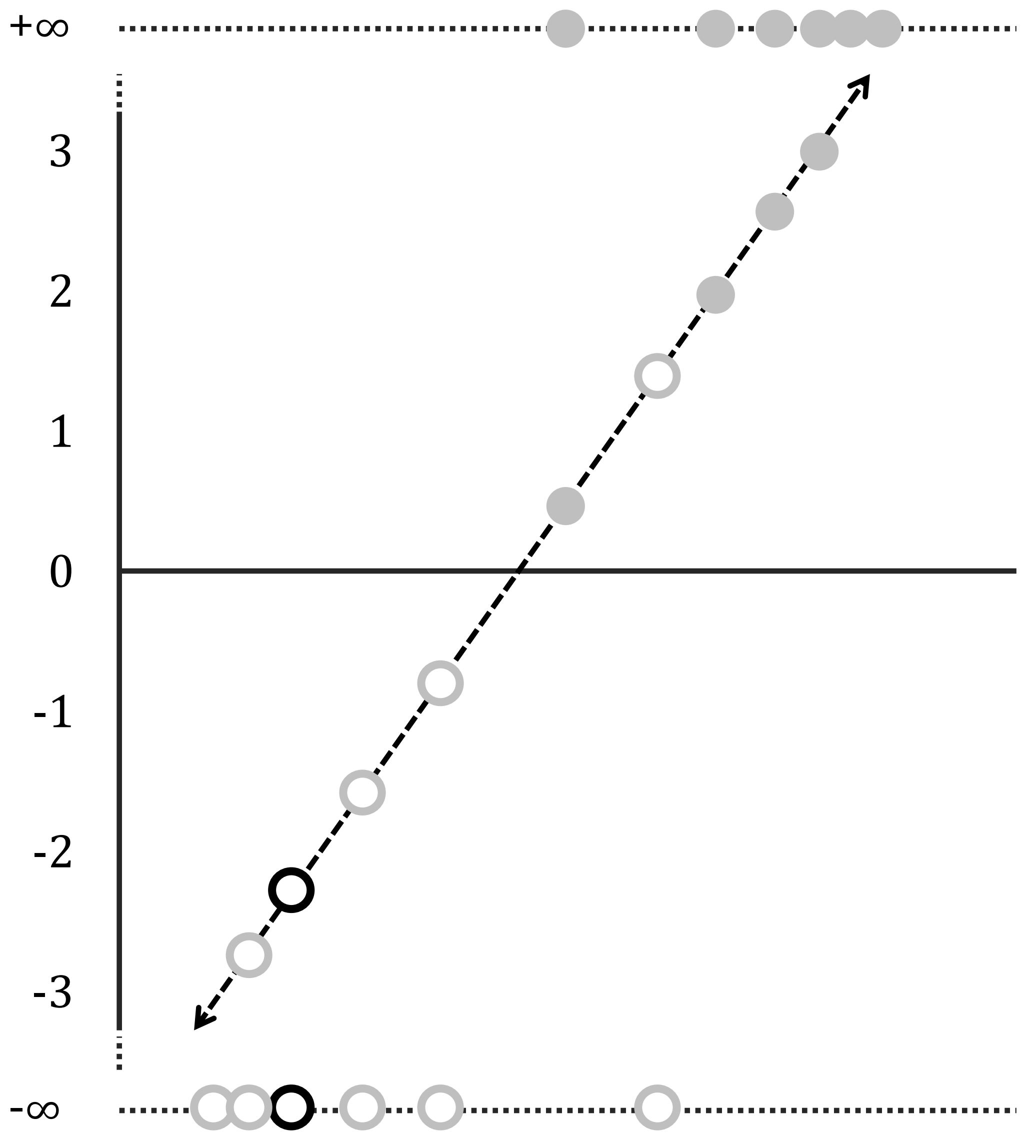
Logistic Regression

To draw the "best fitting" squiggle

Then, transform the candidate log(odds) to candidate probabilities using:

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



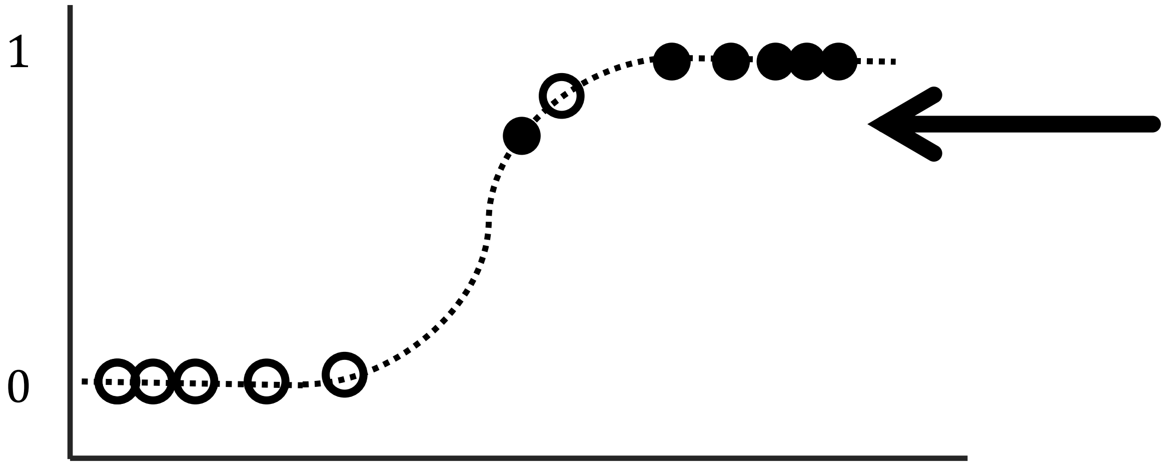


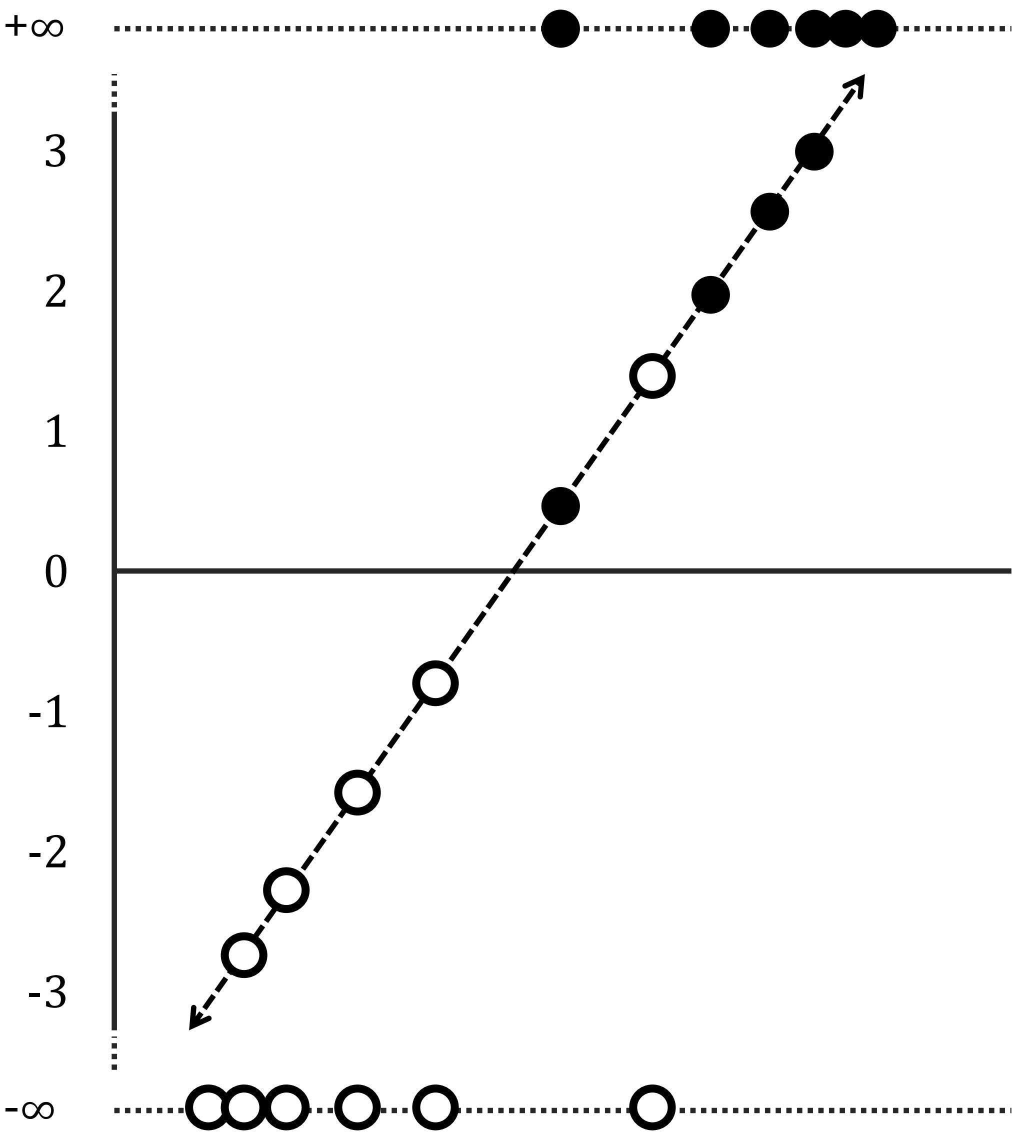
Logistic Regression

To draw the "best fitting" squiggle

Then, transform the candidate log(odds) to candidate probabilities using:

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

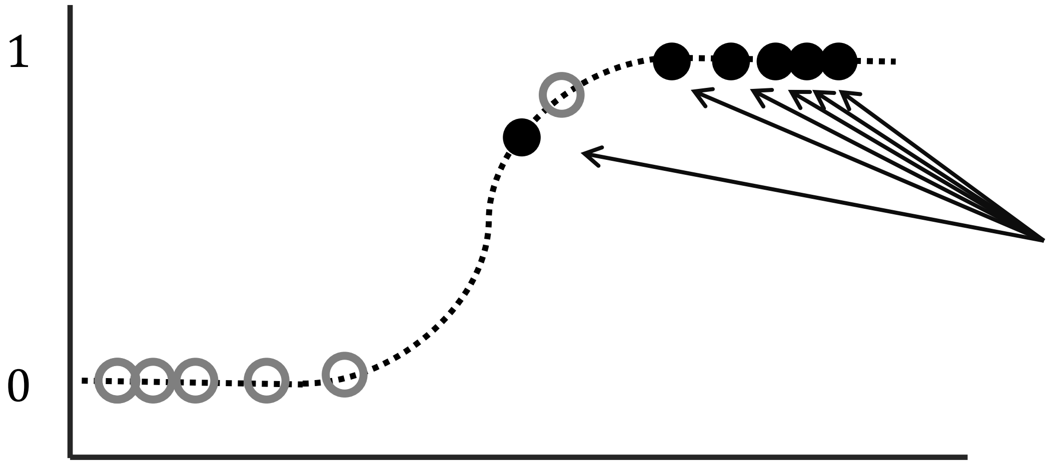




Logistic Regression

To draw the ”best fitting” squiggle

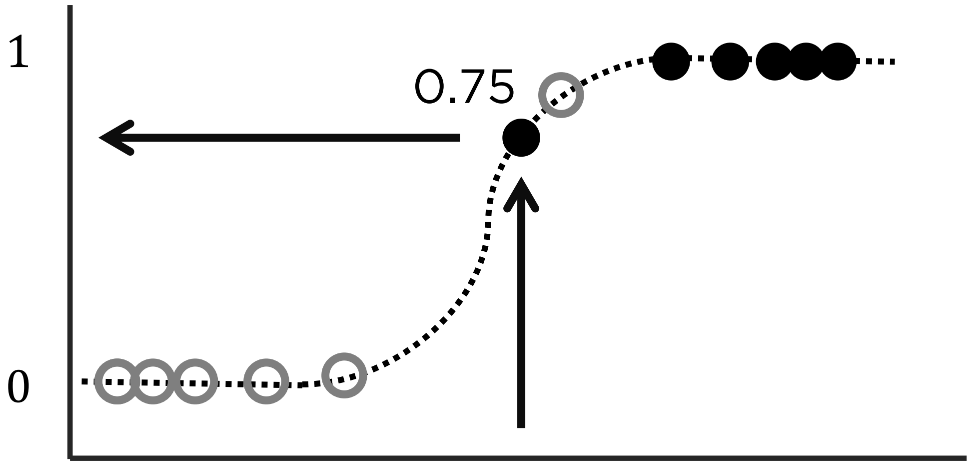
Now, calculate the likelihood.



Logistic Regression

To draw the ”best fitting” squiggle

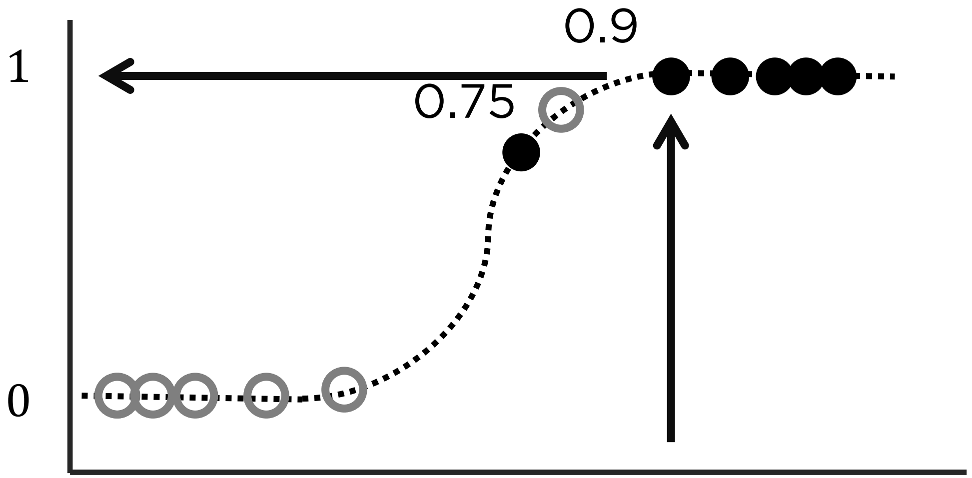
Now, calculate the likelihood.



Logistic Regression

To draw the "best fitting" squiggle

Now, calculate the likelihood.



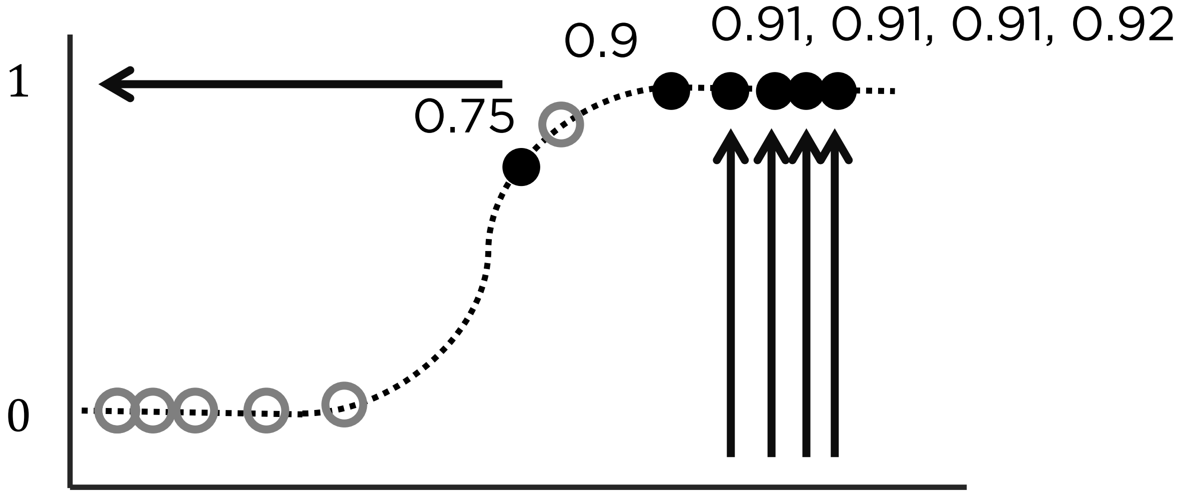
Logistic Regression

To draw the ”best fitting” squiggle

Now, calculate the likelihood.

Likelihood that these students will pass the exam

= 0.75 x 0.9 x 0.91 x 0.91 x0.91 x 0.92



Logistic Regression

To draw the ”best fitting” squiggle

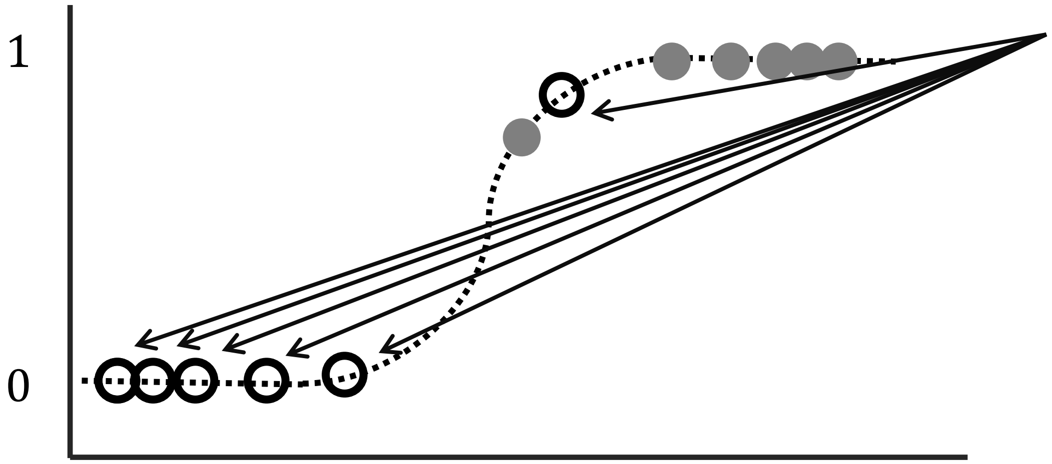
Now, calculate the likelihood.

Likelihood that these students will pass the exam

= 0.75 x 0.9 x 0.91 x 0.91 x0.91 x 0.92

Likelihood that these students will not pass the exam

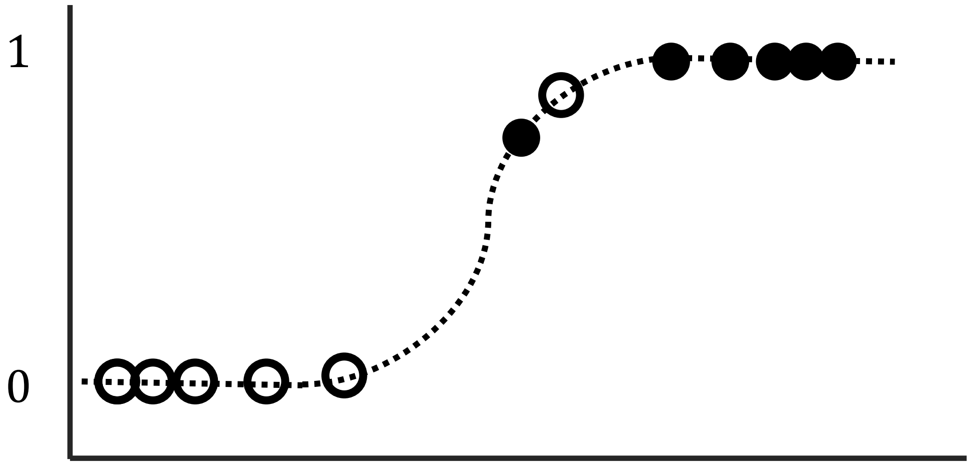
= (1-0.01)x(1-0.01)x(1-0.01)x(1-0.02)x(1-0.03)x(1-0.8)



Logistic Regression

To draw the "best fitting" squiggle

Now, calculate the likelihood.



Likelihood of data given the squiggle

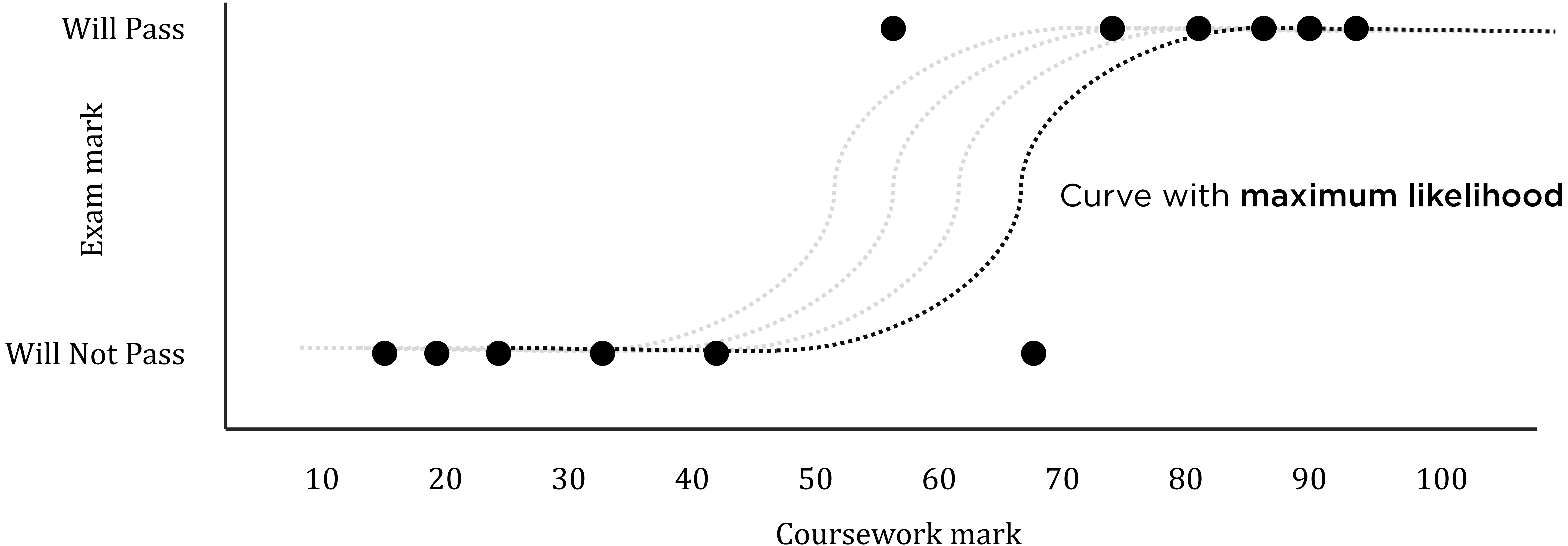
= 0.75 x 0.9 x 0.91 x 0.91 x0.91 x 0.92 x

(1-0.01)x(1-0.01)x(1-0.01)x(1-0.02)x(1-0.03)x(1-0.8)

= 0.086

Logistic Regression

Likelihood



Questions about Assignment