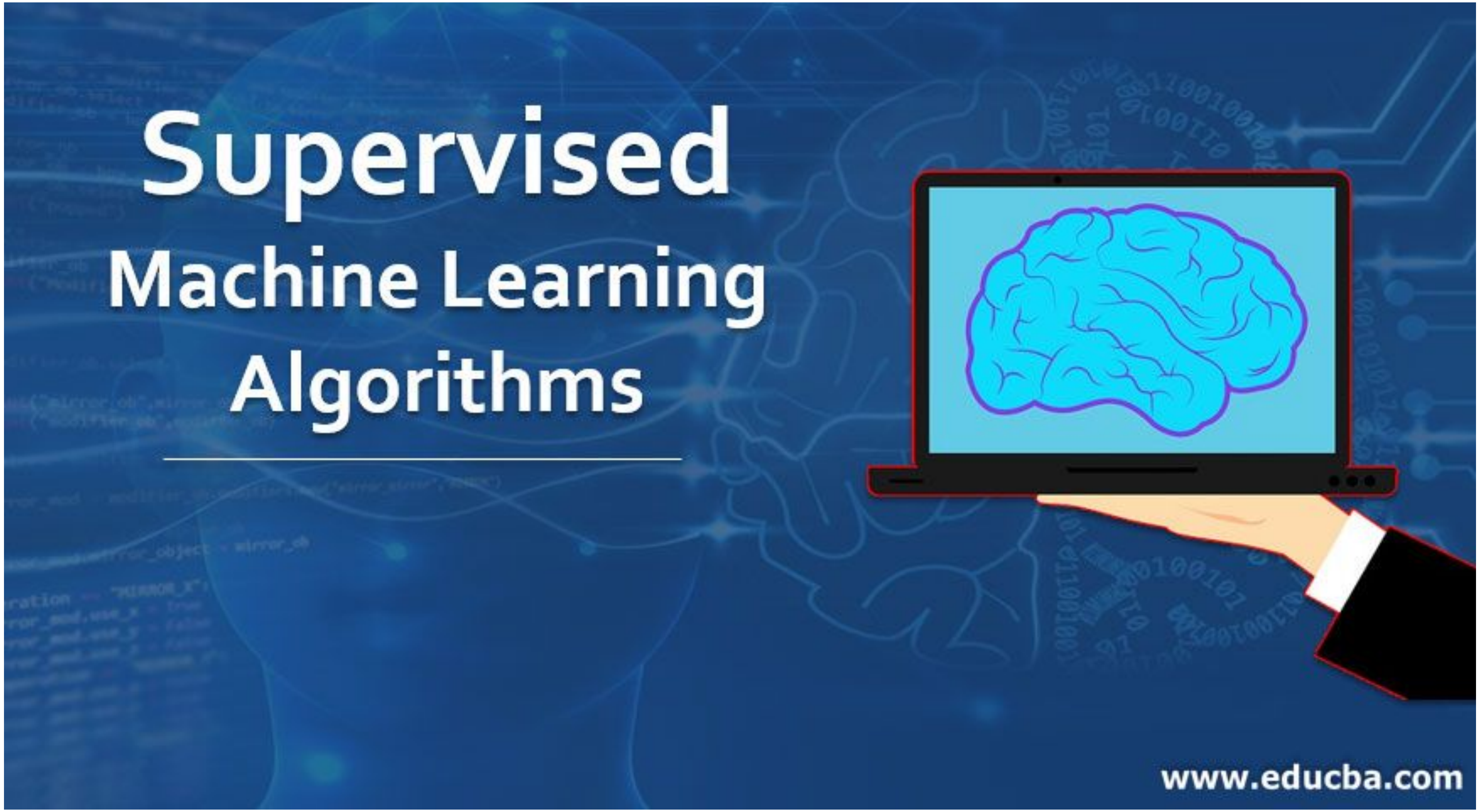

Supervised
Machine Learning
Algorithms
www.educba.com

Last lecture reminder

We learned about:

- L1 Regularization - Lasso Regression
- Lasso Regression vs Ridge Regression
- LassoCV - Python example
- Lasso Regression parameters - Lambda value and scoring metric
- Introduction to Cross-Validation
- Leave-One-Out-Cross-Validation (LOOCV)
- Cross-Validation - Python example (‘cross_val_score’ & ‘corss_validate)

Logistic Regression - Introduction

Logistic Regression → A supervised machine learning algorithm used for classification problems. It is typically used when the dependent variable (target) is categorical. A common application could be in binary problems like whether an email is spam or not, or if a tumor is malignant or not, etc.

Classification algorithms → Classification algorithms are a type of supervised machine learning algorithms used to predict categorical outcomes. Basically, they categorize or "classify" input data into distinct categories or groups. Logistic regression is one algorithm belonging to this group.

Categorical data → Categorical data are variables that contain label values rather than numeric values. They represent characteristics such as a person's gender, marital status, hometown, or the types of sports they like. Categorical data can take on numerical values (such as "1" indicating male and "2" indicating female), but those numbers don't have mathematical meaning.

Logistic Regression - Introduction

Binary categorical data → A specific type of categorical data that has exactly two distinct values or categories. It's essentially a yes-or-no, presence-or-absence kind of data. Examples of binary categorical data include: gender (male / female), or yes / no questions like survived or not, pass or fail, positive or negative.

We can also transform any numeric data into categorical data incase we want.

For example → Suppose you have a dataset of people's ages ranging from 0 to 100. While these are numeric data, they can be converted to categorical data like this:

- **Young** → 0 - 17 years
- **Adult** → 18 - 64 years
- **Senior** → 65+ years

Note: Converting categorical data back to numeric data is not possible without the original numeric data itself.

Logistic Regression - Introduction

In categorical data models, our primary goal is to predict the likelihood or probability of a particular observation falling into a specific category.

For example → A medical diagnosis system that uses patient data to predict whether they have a particular disease or not. Here, the observations are patients, and there are two categories: "Has Disease" and "No Disease".

In this context, our categorical data model would analyze the patient's data (observations) like age, medical history, symptoms etc., and give us the probability of the patient falling into either of the two categories. It might say, for instance, that based on certain combinations of features, there's a 20% chance the patient "Has Disease" and an 80% chance the patient falls into the "No Disease" category.

In this way, categorical data models allow us to predict not just the most likely category for an observation, but the relative likelihood of each possible category.

Logistic Regression - Introduction

What it is mean likelihood and what is the difference between likelihood and probability?

So in order to answer those questions we need to understand some basic concepts in statistics

- **Probability** → Probability deals with calculating the chance of a given event's occurrence, which is expressed as a number between 0 and 1. An event with a probability of 1 can be considered a certainty.
- **Likelihood** → Likelihood refers to the plausibility of a statistical model parameter value given some observed data. It is not actually a probability, but it follows the same kind of scale. Higher likelihood values indicate that the data fit the statistical model well under the given parameters, while lower likelihood values indicate a poorer fit.
- **Odds** → Odds are another way of expressing the likelihood that a particular event will occur, but instead of expressing it as a ratio of the desired outcomes versus all outcomes (like probability), it is expressed as a ratio of the desired outcomes versus the undesired outcomes.

Logistic Regression - Introduction

Let’s take a simple example to explain each concept - Let’s use a fair dice with 6 possible options (1 - 6):

- **Probability** → The probability of an event refers to the number of ways that event can occur divided by the total number of possible outcomes. For instance, if we want to find the probability of rolling a 3 on a fair six-sided die, we know there is 1 way to roll a 3 and 6 total possible outcomes (1, 2, 3, 4, 5, or 6). Therefore, the probability of rolling a 3 is 1/6.

- **Likelihood** → Likelihood, in its simplest form, is a measure of the plausibility of a model given observed data. In the context of our die, say we have some observed data: we rolled the die 600 times and it landed on 3, 100 times. Based on this information, we could use likelihood to evaluate the accuracy of our proposed model (that a fair dice should land on 3 1/6th of the time). Our observed data is pretty close to what our model would predict for 600 rolls (600*1/6 = 100), so the likelihood of our model given the observed data is high.

Logistic Regression - Introduction

- **Odds** → The odds of an event occurring is defined as the number of ways the event can occur versus the number of ways it can't occur. If we are still looking to roll a 3 on a fair six-sided die: we know there is 1 way to roll a 3 and 5 ways not to roll a 3 (rolling a 1, 2, 4, 5, or 6). So, the odds of rolling a 3 on a fair six-sided die are 1:5.

We can use the following formulas to calculate the probability and the odds of a specific event:

P(A) = Number of ways event A can occur / Total number of possible outcomes

In our dice example → P(role 3) = 1 / 6 = ⅙ (probability)

Odds(A) = P(A) / 1 - P(A)

In our dice example → Odds(role 3) = 1 / 5 = ⅕ (odds)

Class Exercise - Probability & Odds

Instructions:

Use the probability and odds formulas to answer the following questions:

- Suppose you toss a fair coin:

  What is the probability that it lands on heads?

  What is the probability that it lands on tails?

- Suppose you roll a fair six-sided die:

  What is the probability and odds that it lands on even numbers (2, 4, or 6)?

  What is the probability and odds that it lands on a number greater than 4?

- Suppose you draw a single card from a well-shuffled standard deck of 52 playing cards:

  What is the probability and odds that you draw a heart?

  What is the probability and odds that you draw a face card (king, queen, or jack)?

Class Exercise Solution - Probability & Odds

Logistic Regression - Theory

Logistic function → Logistic regression is based on the logistic function. The logistic function, also known as the sigmoid function, is an S-shaped curve that maps any real-valued number to a value between 0 and 1, but never exactly at the extremes. This is particularly useful in models where the output can be interpreted as a probability of occurrence of an event.

The formula for the logistic function is:

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

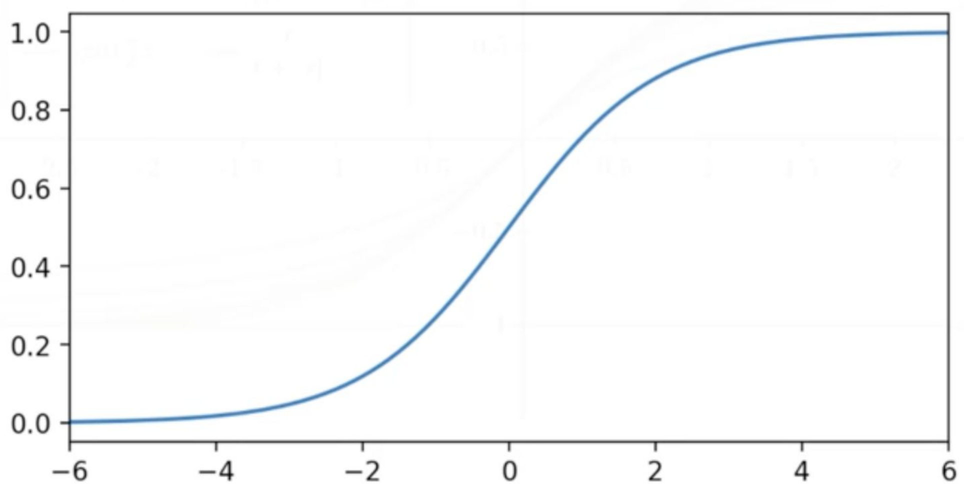
Formula explanation:

e symbol → The symbol e represents Euler's number, which is a mathematical constant that is the base of the natural logarithm. The value of e is approximately equal to 2.71828.

Negative degree → any number raised to the power of a negative integer is equal to one divided by that number. For example → 2^-3 = ½ ^ 3 = ½ * ½ * ½ = ⅛.

Logistic Regression - Theory

The logistic function graph looks like this:



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

We can see that the logistic function graph has limits in the y-axis between 0 and 1 so no matter how small or how large the x value will be it will always converge to 0 or 1 y value.

This represent perfectly a binary categorical data, no matter what the value of the feature x the result y will be 0 or 1.

Logistic Regression - Theory

Let's understand why the logistic function behave like this and bounded between 0 and 1 no matter what the value of the input (x) is. So the logistic function formula is the following:

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

Inputting a very small X (like -1000) into the logistic function f(x) = 1 / (1 + e^-(-1000)) yields almost 0, as e to the power of positive 1000 is an extremely large number. Hence, 1 divided by a large number is near 0.

Similarly, a very large X (like 1000) in f(x) = 1 / (1 + e^-1000), with e^-1000 being extremely small, results denominator value to be 1, because 1 + extremely small value will be equal to 1.
So 1 divided by 1 will equal to 1.

Logistic Regression - Example

So we can see now why logistic regression is optimistic to represent binary categorical situation (when we have only 2 categorial options for the y axis).

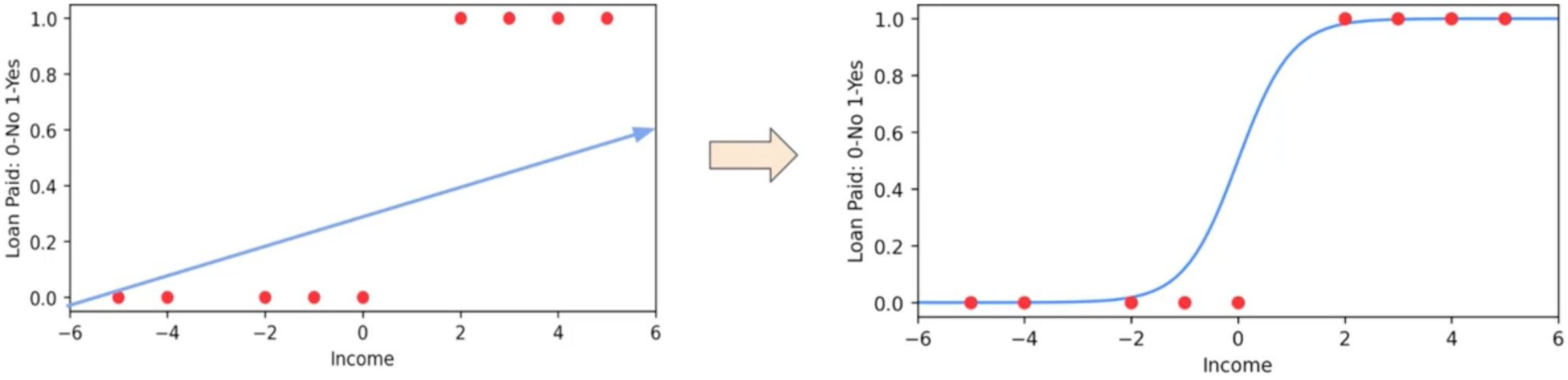
Now let’s say we have the following data set that represent income value and if loan was paid.

We want to perform supervised learning in order to predict for future income the ability to pay the loan.

Income	Loan Paid
-5	0
-4	0
-2	0
-1	0
0	0
2	1
3	1
4	1
5	1

Logistic Regression - Example

We can see that this is binary categorical data because we have only 2 options that we can predict (0 or 1). If we will try to use linear regression we won't be able to predict correctly because the linear regression line is not fit for binary categorical regression.
But the logistic function line (which bound between 0 and 1) will fit perfectly for us.



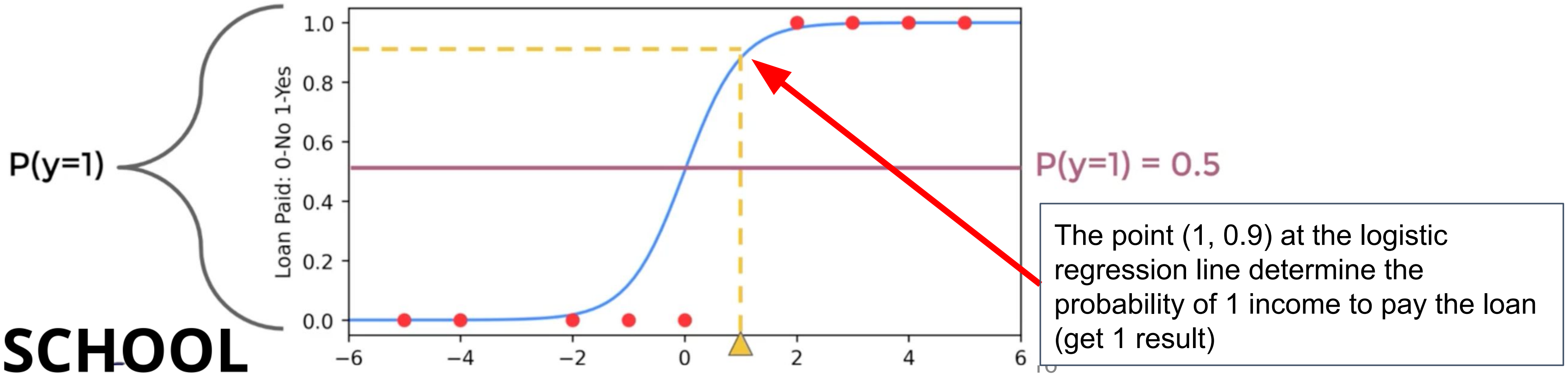
What we want to do is perform conversion between linear regression to logistic regression and predict according to the logistic regression line.

Logistic Regression - Example

As we mention, in categorical data prediction we predict probability for getting a specific category so the logistic regression line allowing us to get the probability of getting each category.

For example → Let’s say we got a new person with income of 1.

We can see that according to the logistic regression line the probability of this person to pay the loan is 0.9 (meaning 90%), and accordingly the probability of him not pay the loan is 0.1 (meaning 10%).



ECOM SCHOOL
המכללה למקצועות הדיגיטל וההייטק

Logistic Regression - Theory

Now let’s understand how we convert linear regression model into logistic regression model.

What we want to do is take our linear regression equation:

$$\hat{y} = \beta_0 x_0 + \cdots + \beta_n x_n$$

$$\hat{y} = \sum_{i=0}^{n} \beta_i x_i$$

And convert it to the logistic equation:

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

Logistic Regression - Theory

We can do it by replacing the (x) in the logistic equation to be equal the ŷ so that way we will get the logistic regression equation:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \longrightarrow \begin{aligned} \hat{y} &= \sigma(\beta_0 x_0 + \cdots + \beta_n x_n) \\ \hat{y} &= \sigma\left(\sum_{i=0}^{n} \beta_i x_i\right) \end{aligned}$$

The logistic regression equation will be:

$$\hat{y} = \frac{1}{1 + e^{-\sum_{i=0}^{n} \beta_i x_i}}$$

ŷ → Represent the probability of belonging to a specific class.

β coefficient → Represent the change in this probability by increasing or decreasing the feature value.

Logistic Regression - Theory

We can also isolate the coefficient in one side and ŷ in the other side of the equation.

So after some mathematical actions we will get the final equation:

$$\frac{\hat{y}}{1-\hat{y}} = e^{\sum_{i=0}^{n} \beta_i x_i}$$

$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \sum_{i=0}^{n} \beta_i x_i$$

Logistic Regression - Theory

The left side of the equation is basically the log("odds") because we saw that calculating the odds of event is Odds(A) = P(A) / 1 - P(A).

So we can say:

$$\ln\left(\frac{p}{1-p}\right) = \ln(odds) \longrightarrow \frac{p}{1-p} = e^{\ln(odds)} \longrightarrow p = \frac{e^{\ln(odds)}}{1+e^{\ln(odds)}}$$

In conclusion:

ln(odds) = β0 + β1*X1 + β2*X2 + ... + βn*Xn

ŷ = 1 / (1 + e^(-log odds))

We can see that if we calculated the "log-odds" of a specific event (using β values) we can simply extract it's probability and we know that the probability is basically ŷ.

This represent the connection between ŷ (the prediction) and β (the coefficients)

Logistic Regression - Theory

Once we understand the connection between the β coefficient value and the ŷ (probability) we can say the following:

- **Positive β** → Indicate that as the predictor variable increases, the log odds of the outcome happening increases. In terms of the probability, this means that the probability of the positive outcome increases as the predictor variable increases.

- **Negative β** → Indicate that as the predictor variable increases, the log odds of the outcome happening decreases. This means that the probability of the positive outcome declines as the predictor variable increases.

- **Zero β** → Indicate that the predictor variable has no effect on the outcome variable. This means any changes in the predictor variable will not influence the probability of the outcome happening.