

## INSTRUCTIONS

You are about to participate in an experiment on decision-making. Please turn off cell phones and similar devices now. Please do not talk or in any way try to communicate with other participants. We will start with a brief instruction period. If you have any questions during this period, raise your hand and your question will be answered so everyone can hear.

What you earn in the session depends partly on your decisions, and partly on chance. In this experiment you will make a series of predictions. At the end of the session, one of these predictions is selected at random to determine your payoff. Your payment is equal to \$10 plus the earnings in the randomly selected prediction task.

Below we describe the structure of a prediction task.

## Your Task

Your main task in this experiment is to make predictions. This task will be repeated several times.

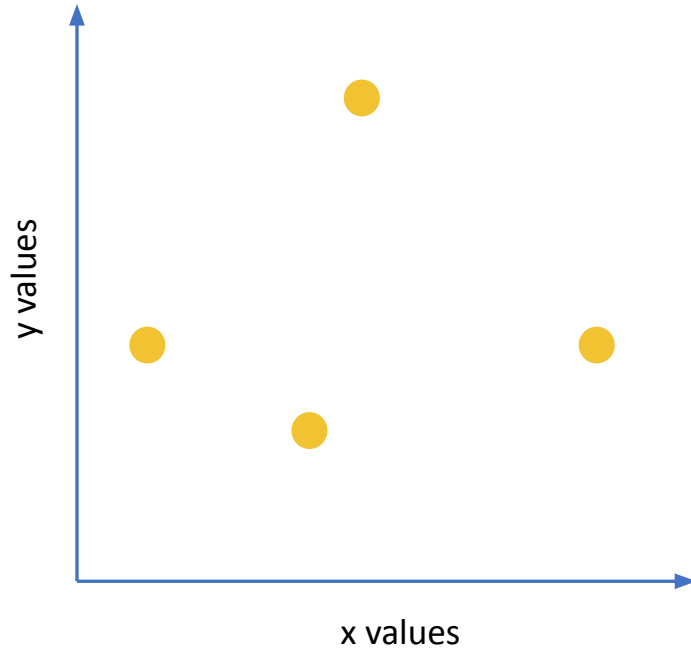
In each task, you will be making predictions about a publicly available data set. These data sets cover a range of topics (health, weather, traffic, etc.), so the data might involve measurements of time, weight, income, happiness, health, or other variables. We will reveal the data sources at the end of the experiment.

Each data point in a data set includes two values (which we will refer to as  $\mathbf{x}$  and  $\mathbf{y}$ ). Note that you will not be told what  $\mathbf{x}$  and  $\mathbf{y}$  stand for.

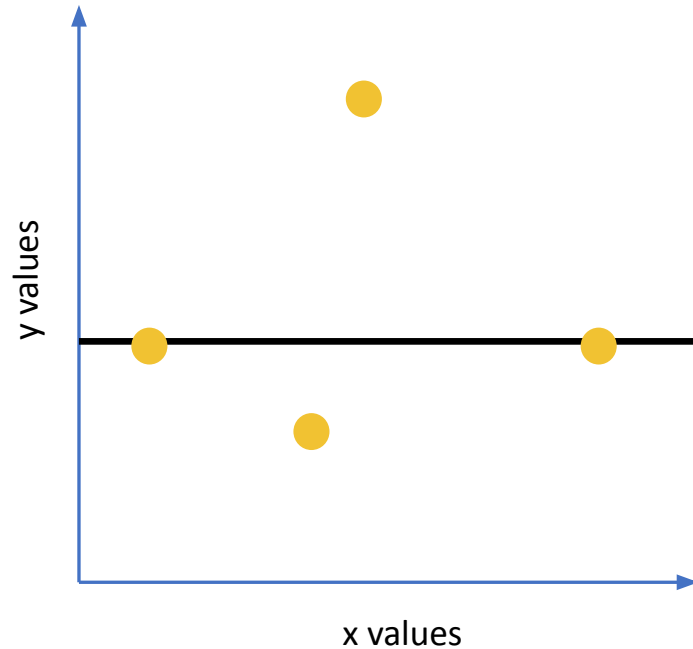
**We will show you a random subset of points from the data set. Based on these, you have to predict the values of  $\mathbf{y}$  for all possible values of  $\mathbf{x}$ .**

Here is a simple example.

**Assume the yellow dots are the four data points randomly shown to you from the data set.  
You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .  
In other words, this will represent your best guess of the value of  $y$  for any  $x$ .**



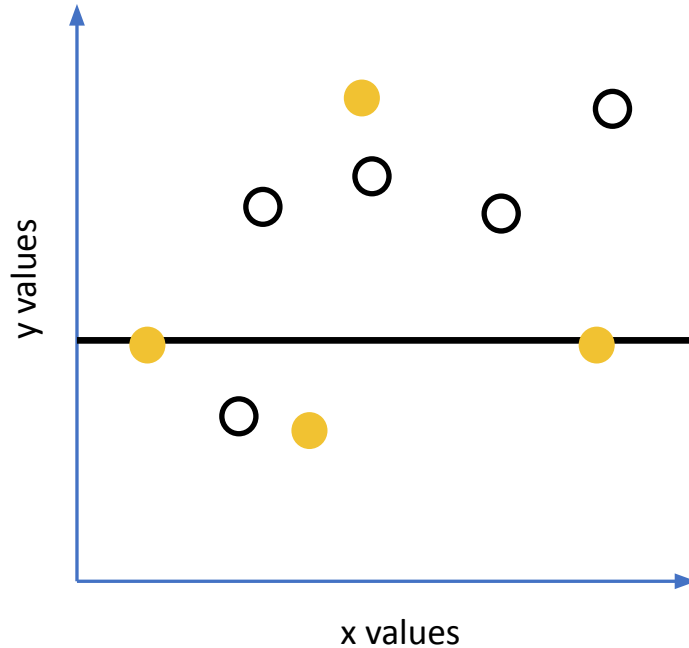
Assume the yellow dots are the four data points randomly shown to you from the data set.  
You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .



**Assume you draw this linear function.**

Assume the yellow dots are the four data points randomly shown to you from the data set.  
You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

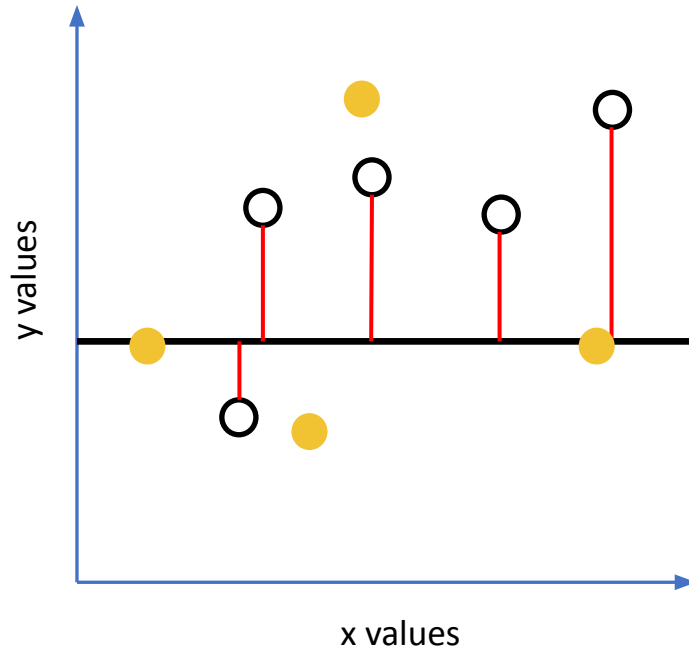
**We will then use the function you submitted to make predictions about the data points you did not see.**



**Now assume that the rest of the data points—the ones that were not shown to you—are represented by the hollow dots.**

Assume the yellow dots are the four data points randomly shown to you from the data set. You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

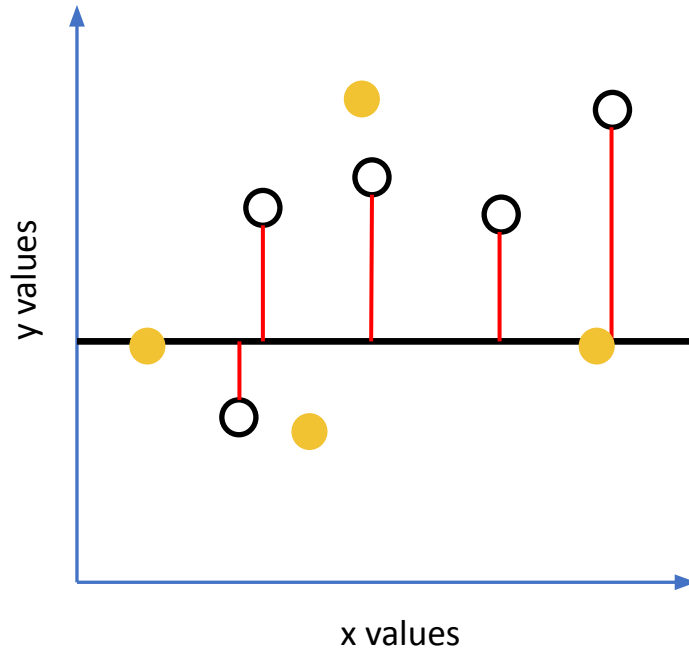
We will then use the function you submitted to make predictions about the data points you did not see.



**For each of the hidden data points (hollow dots), we will measure how far it is from your line—using the vertical distance. These distances are shown in red.**

Assume the yellow dots are the four data points randomly shown to you from the data set. You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

We will then use the function you submitted to make predictions about the data points you did not see.

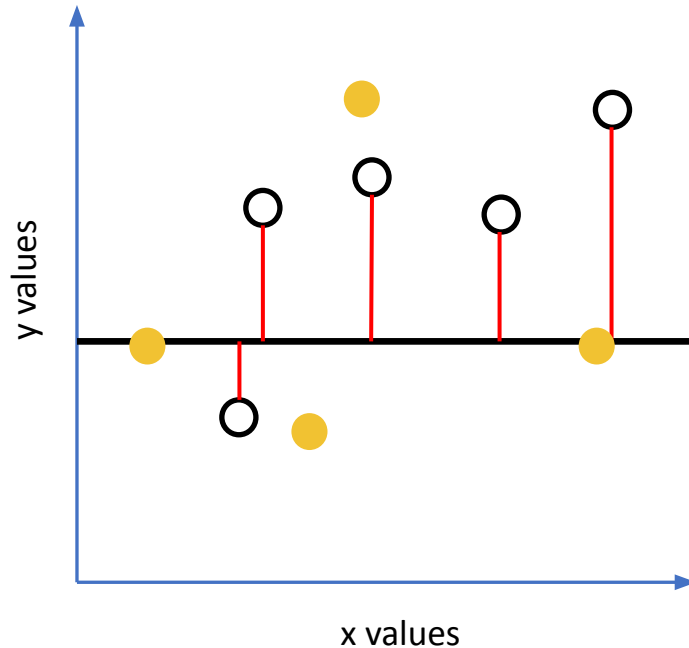


For each of the hidden data points (hollow dots), we'll measure how far it is from your line—using the vertical distance. These distances are shown in red.

**Your payoff depends on how accurate your predictions are: the smaller these distances are, the higher your payoff will be.**

Assume the yellow dots are the four data points randomly shown to you from the data set. You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

We will then use the function you submitted to make predictions about the data points you did not see.



For each of the hidden data points (hollow dots), we'll measure how far it is from your line—using the vertical distance. These distances are shown in red.

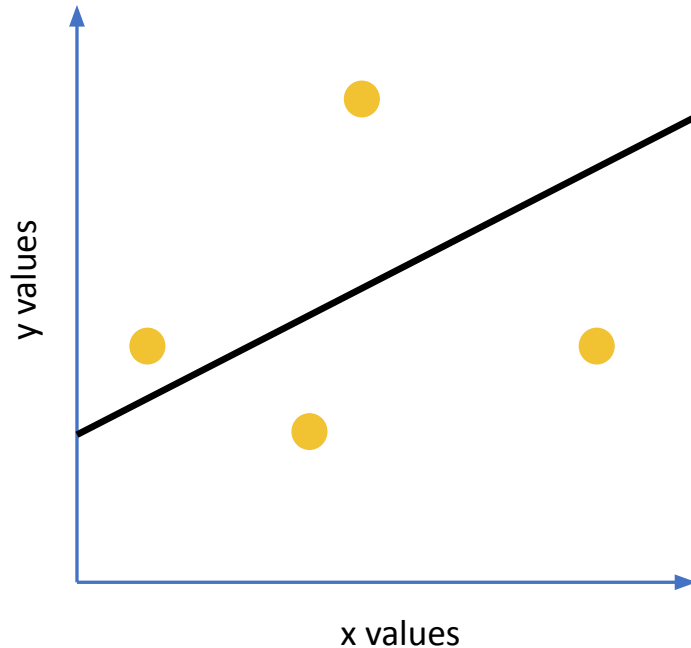
Your payoff depends on how accurate your predictions are: the smaller these distances are, the higher your payoff will be.

**The exact formula to determine your payoffs will be explained later, but before that we will explore the task a bit further.**



Assume the yellow dots are the four data points randomly shown to you from the data set. You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

We will then use the function you submitted to make predictions about the data points you did not see.



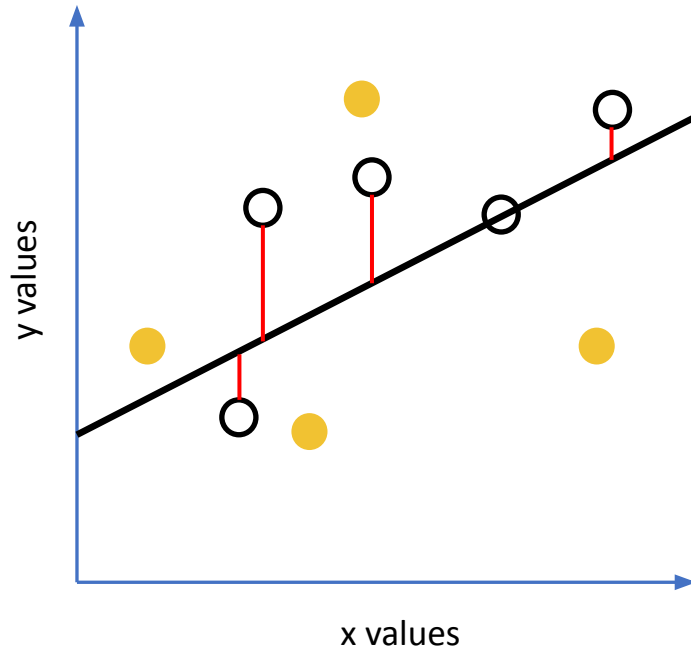
For each of the hidden data points (hollow dots), we'll measure how far it is from your line—using the vertical distance. These distances are shown in red.

Your payoff depends on how accurate your predictions are: the smaller these distances are, the higher your payoff will be.

**For example, drawing this alternative function would have resulted in a higher payoff.**

Assume the yellow dots are the four data points randomly shown to you from the data set. You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

We will then use the function you submitted to make predictions about the data points you did not see.



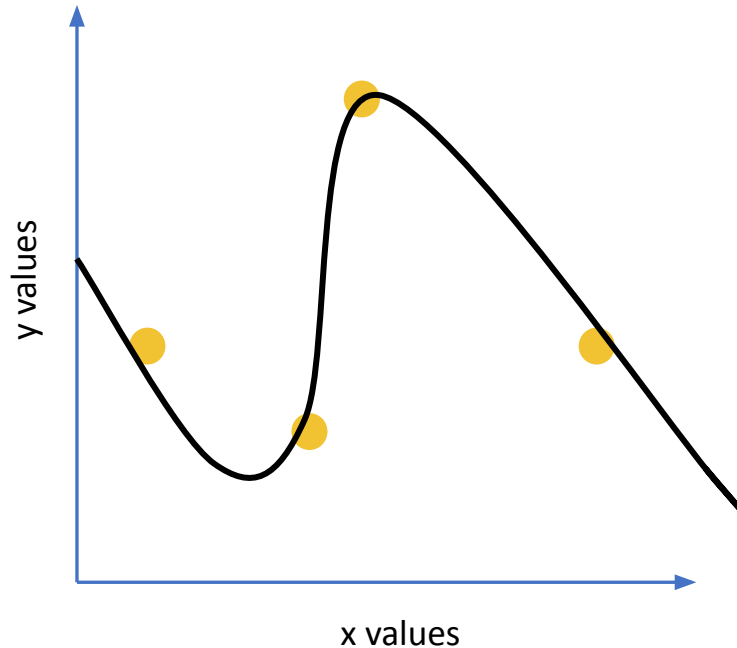
For each of the hidden data points (hollow dots), we'll measure how far it is from your line—using the vertical distance. These distances are shown in red.

Your payoff depends on how accurate your predictions are: the smaller these distances are, the higher your payoff will be.

**For example, drawing this alternative function would have resulted in a higher payoff.**

Assume the yellow dots are the four data points randomly shown to you from the data set. You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

We will then use the function you submitted to make predictions about the data points you did not see.



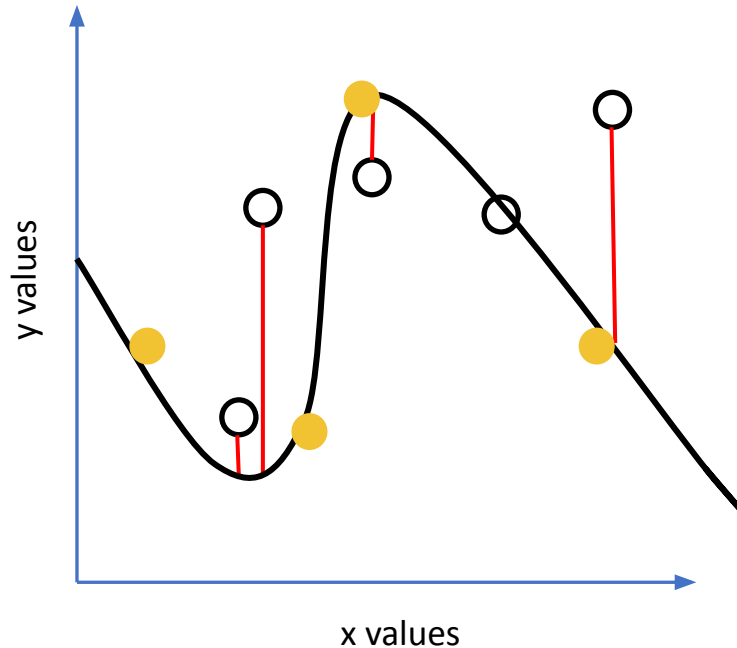
For each of the hidden data points (hollow dots), we'll measure how far it is from your line—using the vertical distance. These distances are shown in red.

Your payoff depends on how accurate your predictions are: the smaller these distances are, the higher your payoff will be.

**But drawing this other function would have resulted in lower payoffs.**

Assume the yellow dots are the four data points randomly shown to you from the data set. You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

We will then use the function you submitted to make predictions about the data points you did not see.



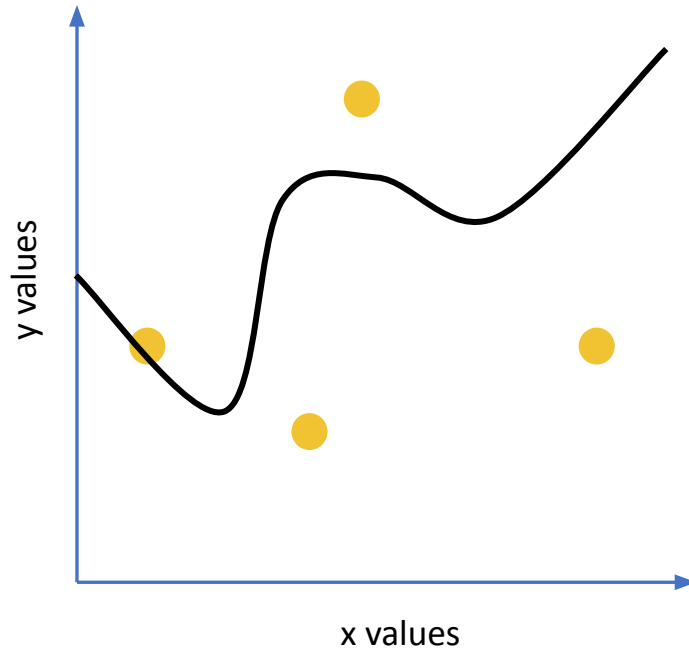
For each of the hidden data points (hollow dots), we'll measure how far it is from your line—using the vertical distance. These distances are shown in red.

Your payoff depends on how accurate your predictions are: the smaller these distances are, the higher your payoff will be.

**But drawing this line other function would have resulted in lower payoffs.**

Assume the yellow dots are the four data points randomly shown to you from the data set. You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

We will then use the function you submitted to make predictions about the data points you did not see.



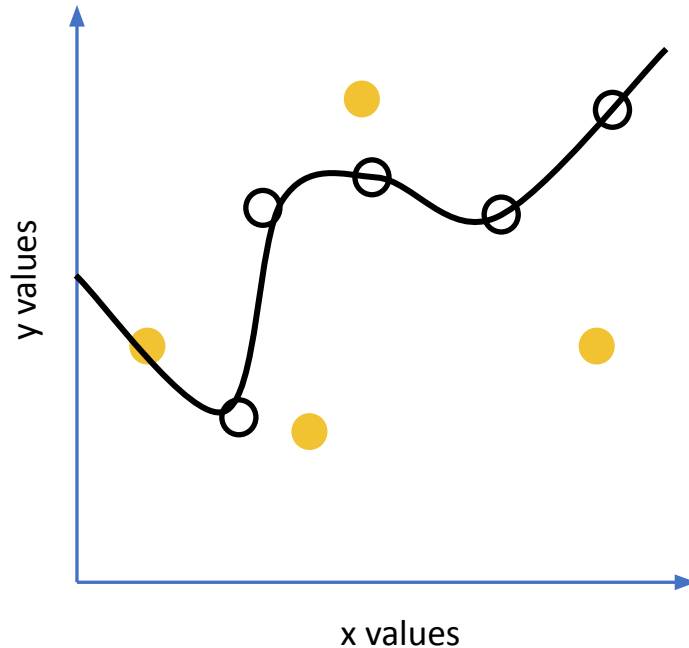
For each of the hidden data points (hollow dots), we'll measure how far it is from your line—using the vertical distance. These distances are shown in red.

Your payoff depends on how accurate your predictions are: the smaller these distances are, the higher your payoff will be.

**This function would have achieved the highest accuracy level.**

Assume the yellow dots are the four data points randomly shown to you from the data set. You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

We will then use the function you submitted to make predictions about the data points you did not see.



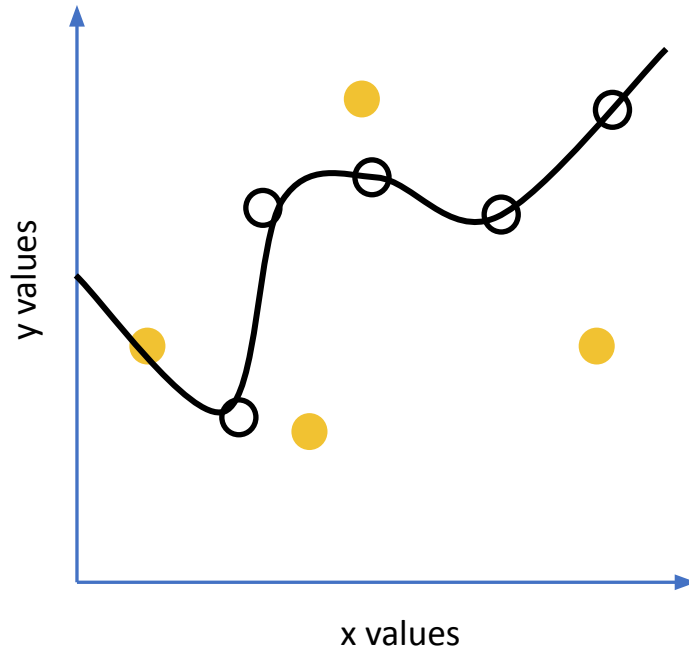
For each of the hidden data points (hollow dots), we'll measure how far it is from your line—using the vertical distance. These distances are shown in red.

Your payoff depends on how accurate your predictions are: the smaller these distances are, the higher your payoff will be.

**This function would have achieved the highest accuracy level.**

Assume the yellow dots are the four data points randomly shown to you from the data set.  
You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

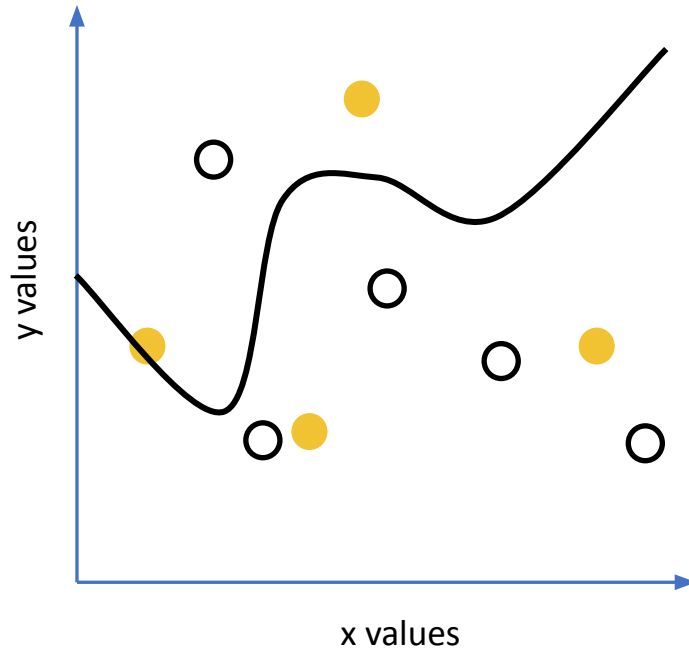
We will then use the function you submitted to make predictions about the data points you did not see.



**Although, in the example we just showed you, this function matches the points exactly (hollow dots), it could have led to poor predictions if the hidden data points (hollow dots) had looked very different.**

Assume the yellow dots are the four data points randomly shown to you from the data set.  
You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

We will then use the function you submitted to make predictions about the data points you did not see.



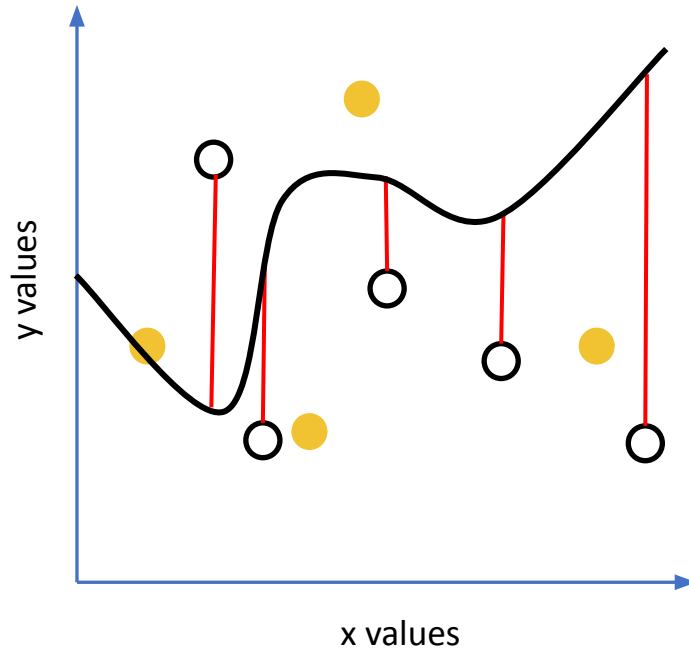
Although, in the example we just showed you, this function matches the points exactly (hollow dots), it could have led to poor predictions if the hidden data points (hollow dots) had looked very different.

**For instance if instead these were the other points in the data set.**



Assume the yellow dots are the four data points randomly shown to you from the data set.  
You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

We will then use the function you submitted to make predictions about the data points you did not see.

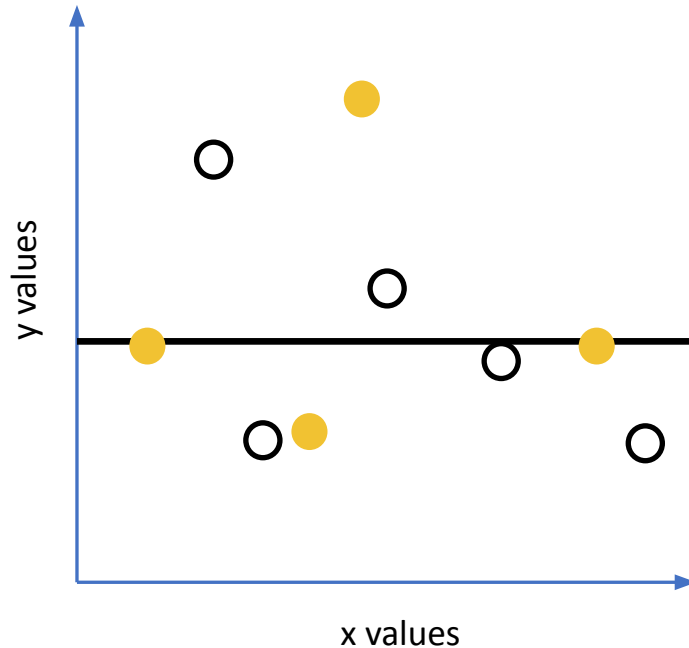


Although, in the example we just showed you, this function matches the points exactly (hollow dots), it could have led to poor predictions if the hidden data points (hollow dots) had looked very different.

**For instance if instead these were the other points in the data set.**

Assume the yellow dots are the four data points randomly shown to you from the data set.  
You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

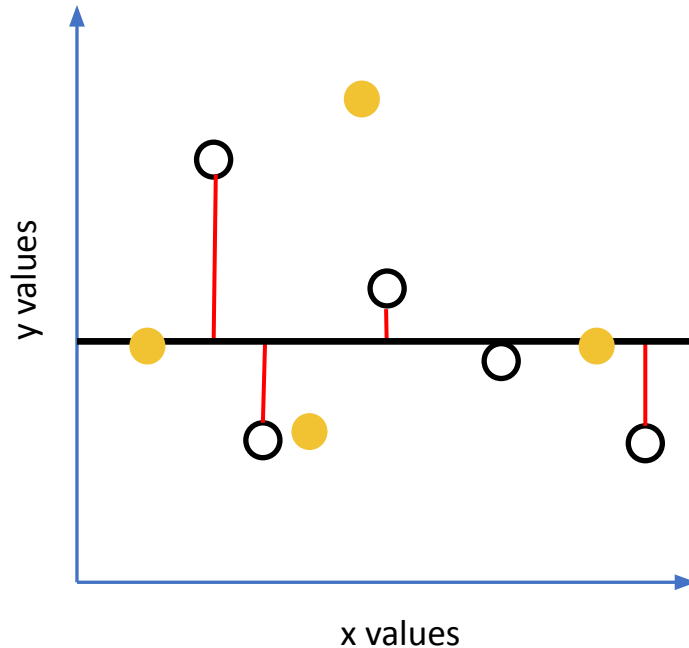
We will then use the function you submitted to make predictions about the data points you did not see.



**For these other hollow dots the first function would lead to much more accurate predictions.**

Assume the yellow dots are the four data points randomly shown to you from the data set.  
You will be asked to draw a function. This will represent the value of  $y$  you would predict for any  $x$ .

We will then use the function you submitted to make predictions about the data points you did not see.



For these other hollow dots the first function would lead to much more accurate predictions.

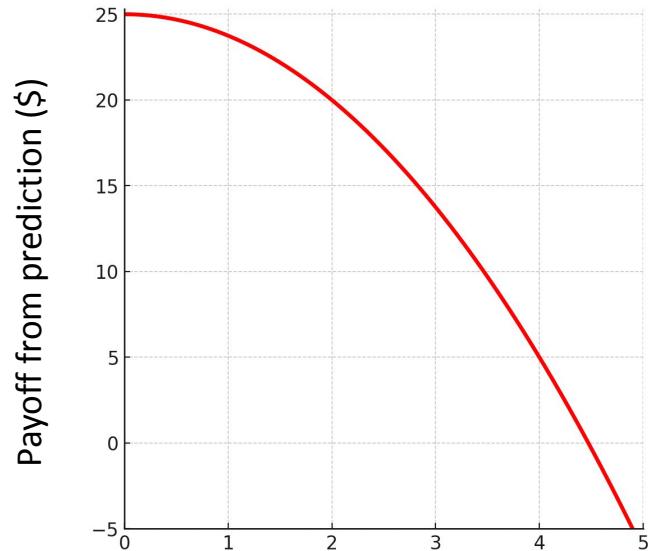
**To describe exactly how payoffs work, let  $D$  be the vertical distance between your prediction and the actual value of each hollow dot. (Length of the red line.)**

Here is the exact formula used to determine your payoffs for each prediction.

$$\text{Payoff per prediction} = \$25 \times (1 - D^2/20)$$

Recall that **D** is the vertical distance between your prediction and the true value.

The figure below describes how your payoff changes with **D**. To compute your final payoff, we average the payoff you obtained for each prediction.



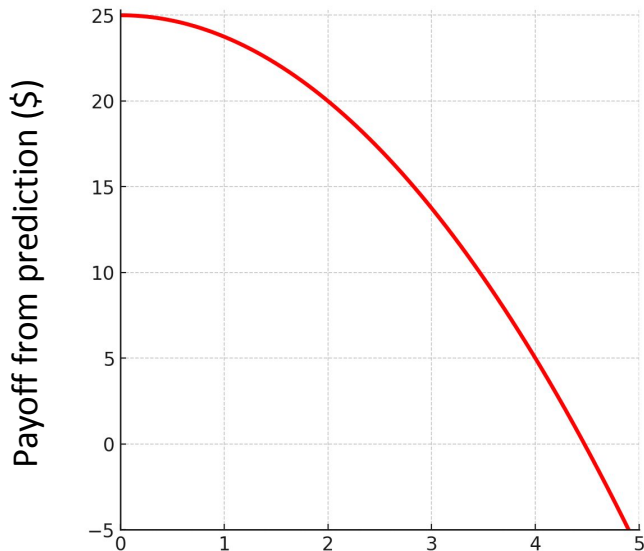
Distance between your prediction and actual value

Here is the exact formula used to determine your payoffs for each prediction.

$$\text{Payoff per prediction} = \$25 \times (1 - D^2/20)$$

Recall that **D** is the vertical distance between your prediction and the true value.

The figure below describes how your payoff changes with **D**. To compute your final payoff, we average the payoff you obtained for each prediction.



Distance between your prediction and actual value

Example:

Assume you predict 2.

If the true value is 1, your payoff is  $\$25 \times (1 - (2 - 1)^2/20) = \$23.75$

If the true value is 5, your payoff is  $\$25 \times (1 - (2 - 5)^2/20) = \$13.75$

If both points were in the data set,  
your average payoff would be  $(\$23.75 + \$13.75)/2 = \$18.75$ .

In other words, both positive and negative errors are costly, but larger errors are much more costly.

## Summary

- The experiment consists of 54 tasks. This will be followed by a short second part that we'll describe later.
- In each task, you will be presented with a different data set. Each data set consists of 350 points randomly sampled from the original data set.
- These data sets are real and cover a range of topics, but you will not know what the two variables **x** and **y** represent.
  - In fact, the data will be rescaled such that the **x** and **y** variables never go below 0 or above 10.
- You will only see a subset of the data points from that data set. (Either 5, 10 or 50 points [randomly selected points].)
- Based on these observed points, you will draw a function that will be used to predict **y** based on **x** in that data set.
- Specifically, we will measure the vertical distance between your function and each of the hidden points.
- Your payoff depends on the accuracy of these predictions: the smaller the distance between your prediction and the true value, the higher your payoff.
- Note that you are only rewarded for the accuracy on the data points you did **not** observe. How accurate your function is in predicting the points you were shown does not affect your payoff.
- Hence, we took a publicly available data set, sampled 350 data points from it and normalized those points so that their values are all between 0 and 10. In each task we randomly select 5, 10, or 50 of those points to show you. Your payment is determined using your prediction function and the remaining 345, 340, or 300 points; depending on the case.