



Universidad  
de Alcalá

# Artificial Intelligence

## Foundations of Machine Learning II

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

# LINEAR MODELS

The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than the "Cartagena" part. The text is set against a light blue background with a white swoosh underneath, all contained within a white rectangular box.

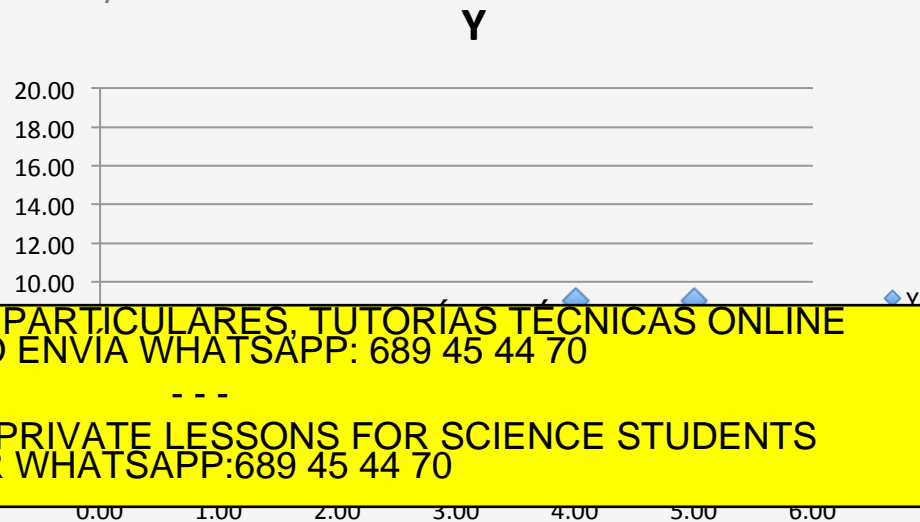
CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- As we have mentioned ML allows to create models in a supervised, semi-supervised, unsupervised or reinforced fashion.
- Nevertheless most of the models need to be pre-specified in some sense, that is, we can not simply tell the machine “to learn”, we need to formulate some particular hypothesis or, intuitively, to provide with some “hints”.
- For example, assume that, after observation, we record and plot the behavior of two variables X and Y, where Y is the variable we are interested in, let us assume that we have:

X	Y
3.00	7.00
2.00	3.00
4.00	8.00
5.00	9.00



**Cartagena99**

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---  
ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- We see that whenever X is bigger Y is bigger so we can postulate that there may be a proportional relationship between X and Y.

$$Y = \alpha X$$

- This is a *model*, that is, a mathematical entity that allows to explain some particular fact, in this case, the relationship between X and Y.
- Note that the model involves not only the variables (X and Y, in our case) but *parameters* ( $\alpha$  i our case) that give flexibility so that if either X or Y changes then the model still holds in some sense.
- Note also that the parameters may differ even if the relationship still holds. e.a.

Cartagena99

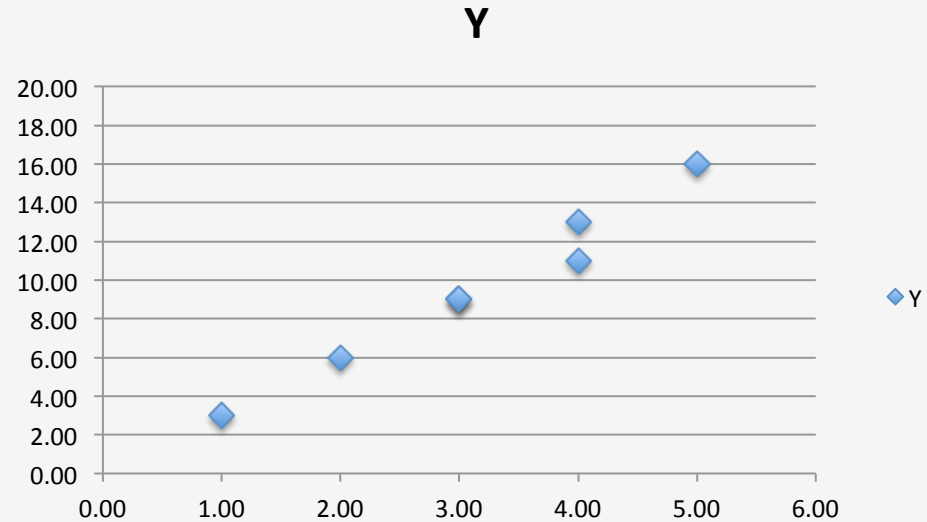
CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Assume now the following data and look at the corresponding plot:

X	Y
3.00	9.00
2.00	6.00
4.00	11.00
5.00	16.00
4.00	13.00
1.00	3.00
3.00	9.00



- Note that the relationship still holds ( $Y$  is proportional to  $X$ ) but it is not exactly the same,  $Y$  is more responsive to  $X$  of, in geometric terms, the slope of the points has increased.

**Cartagena99**

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Notice that even though the model is valid, the value of the parameter will not be, being higher in the second case.
- The process of finding the optimal parameters for a model is what we call *estimation* (in statistical terms) or *learning* (in the machine learning *argot*).
- *Algorithms* are computational procedures (most of the times *iterative* procedures) that allow to find the parameters of the model which are optimal under some criterion.
- For example, we may try to find the value of the parameter  $\beta$  so that the model “fits” the data as close as possible.

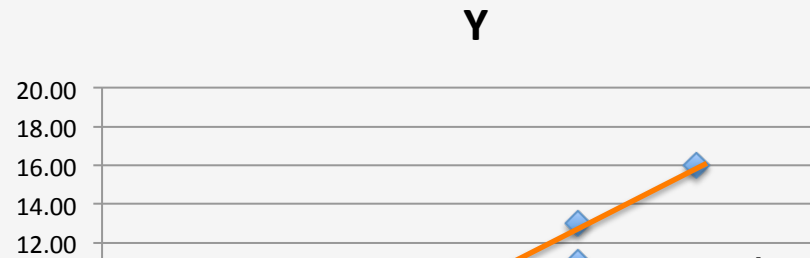
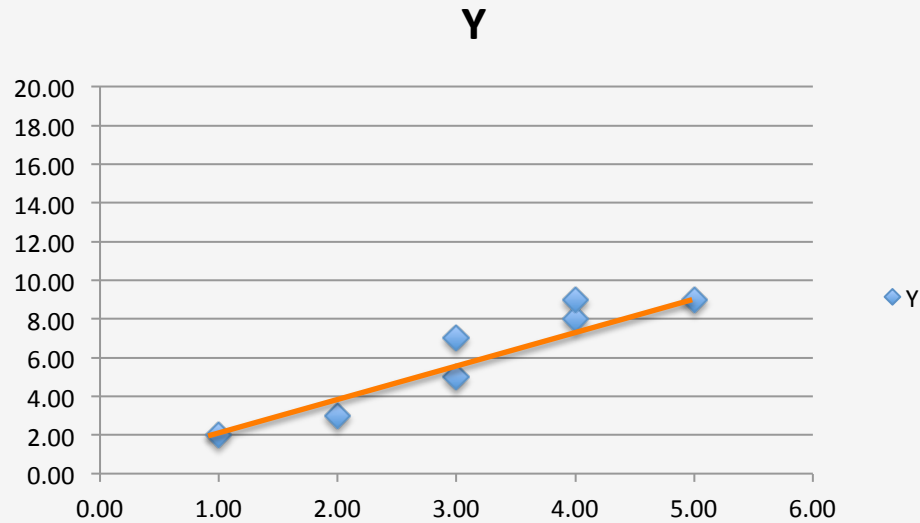
The logo for Cartagena99 features the text 'Cartagena99' in a stylized, dark green font. The '99' is significantly larger and more prominent than the rest of the text. The logo is set against a light blue and white background with a subtle wave-like pattern.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Intuitively (no formal method used) we may suggest a value of  $\beta=2$  in the first case and  $\beta=3$  in the second.



**Cartagena99**

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
 CALL OR WHATSAPP: 689 45 44 70

0.00 1.00 2.00 3.00 4.00 5.00 6.00

- The model is still the same but the *calibration* is different.
- Even generative models that modify their structure adding complexity to capture the relation between independent and dependent variables.
- For the moment we can assume that the parameters are found by trial and error.
- Notice that, after finding the optimal value (optimality is defined later) we can employ the model to forecast unseen examples, e.g.

$$\hat{Y}' = \hat{\beta}X'$$



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
 CALL OR WHATSAPP:689 45 44 70



- Models may have some inertial component (in models such as artificial neural networks it is called the *bias* and in Statistics the *intercept*) so that there is a response even if the input is zero:

$$Y = \alpha + \beta X$$

- And similarly we can forecast for new observations:

$$\hat{Y}' = \hat{\alpha} + \hat{\beta}X'$$

- The model that we have just described is called a *linear* model for obvious reasons, if we plot Y against X the result is a line (which passes through the origin if  $\alpha=0$ ) and we say that there is a *linear relationship* between Y and X.
- More specifically, this model is called a *univariate* linear model because there is only one independent variable.

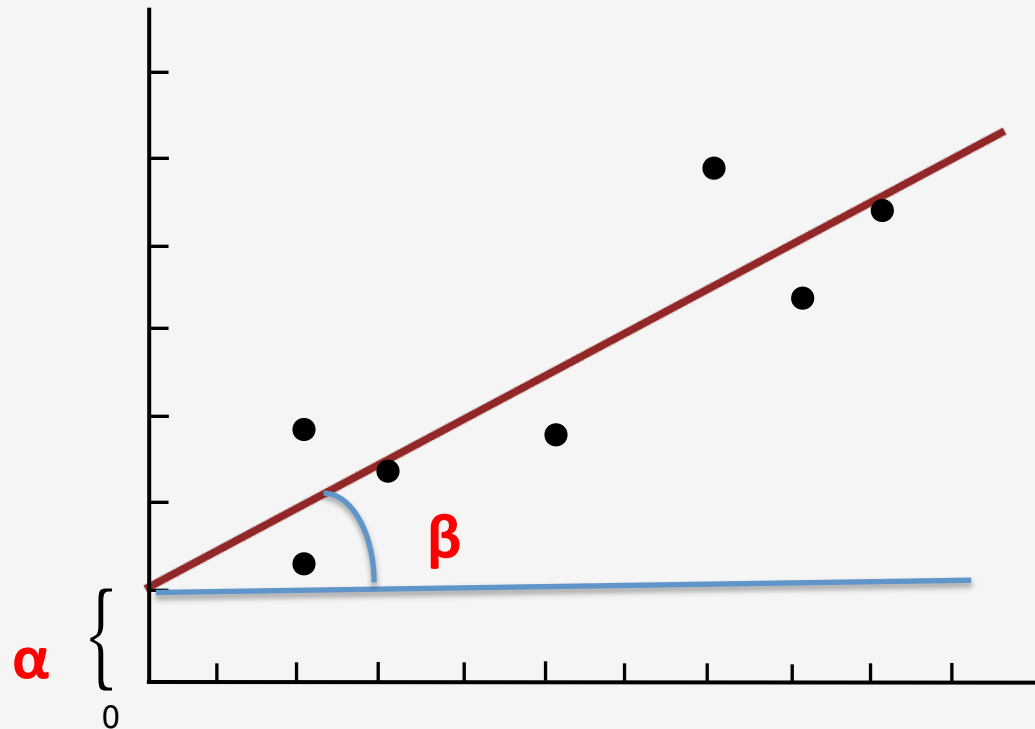
The logo for Cartagena99, featuring the text 'Cartagena99' in a stylized font with a blue and orange gradient background.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Geometrically:



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- In most real applications, the relationship between  $Y$  and  $X$  will not be perfect but affected by unpredictable components that are called *noise*, in such cases we have:

$$Y = \alpha + \beta X + \varepsilon$$

- Note that, by definition,  $\varepsilon$  is unpredictable so it will be useless trying to make any guess about values of  $\varepsilon$ , all that can be “learnt” of the relationship between  $Y$  and  $X$  is summarized in the parameters of the model,  $\alpha$  and  $\beta$  in our case.
- Finally, note that we may have not just one independent variable but a set of them, in such case we can proceed similarly and consider a *multivariate linear model*:

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, bold font. The 'C' is large and green, while the rest of the text is in a dark green color. Below the text is a horizontal bar with a gradient from orange to yellow.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

# LINEAR MODEL ESTIMATION

The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than the "Cartagena" part. The text is set against a light blue background with a white swoosh underneath, all contained within a white rectangular box.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- As mentioned, the parameters that index any model need to be found, we have also mentioned that they should be the “best” ones under some particular criterion.
- One of the commonest criteria is *mean squared error*, there are technical details why this criterion is optimal in many applications but for the moment we will rely on an intuitive interpretation.
- Note that for any value of the parameters we can calculate the difference between the actual data and our prediction using those parameters:

$$\varepsilon = g(Y, \hat{Y}) = (Y - \hat{Y}) = (Y - (\hat{\alpha} + \hat{\beta}X))$$

- Intuitively, it is obvious that a “good” model would provide forecasts that are as close as possible to the true observations,



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Of course one would like to do this for ALL the observations in the dataset, assume we have a *sample*

$$\{X_i, Y_i\}$$

- Naturally one would like:

$$\min_{\alpha, \beta} \sum \varepsilon_i = (Y_i - (\alpha + \beta X_i))$$

- Since positive errors could compensate negative errors it seems more reasonable

$$\min_{\alpha, \beta} \sum \varepsilon_i = (Y_i - (\alpha + \beta X_i))^2$$

- Note that the above formula explains the name: *minimum*

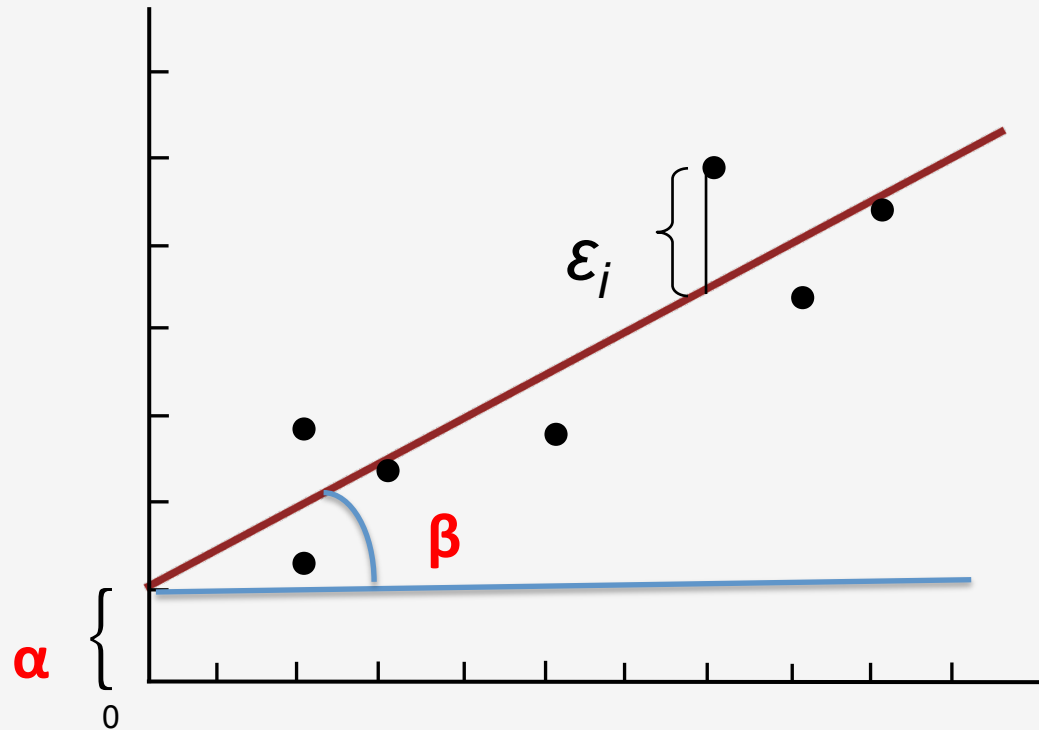
**Cartagena99**

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP: 689 45 44 70

- Geometrically:



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Since, as mentioned, random guess is generally unfeasible, the problem now becomes on using an efficient algorithm to find the parameters.
- There are several ways to do that but the simplest method is to employ the *normal equation*, in the multivariate case assume we have

$$X = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{pmatrix} \quad Y = \begin{pmatrix} Y_{11} \\ Y_{12} \\ \dots \\ Y_{1m} \end{pmatrix}$$

**Cartagena99**

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
 LLAMA O ENVIA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
 CALL OR WHATSAPP: 689 45 44 70



- The optimal parameters, under the *mse* criterion, can be found solving the equation:

$$\beta = (X^t X)^{-1} X^t Y$$

- Note, that if we have an intercept the above formula holds by considering

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1n} \\ 1 & X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{m1} & X_{m2} & \dots & X_{mn} \end{pmatrix}$$

- In this case:



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- In order to reduce notation complexity it is common to simplify as

$$\theta = (X^t X)^{-1} X^t Y$$

- Where

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_n X_n$$

- Note that these parameters found are optimal under the mse criterion, but they do not need to be if we change the *performance* function, e.g.

$$\varepsilon = g(Y, \hat{Y}) = |Y - \hat{Y}|$$

The logo for Cartagena99 features the text "Cartagena99" in a stylized, bold font. The "C" is large and green, while "artagena99" is in a smaller, dark green font. The text is set against a light blue background with a white swoosh underneath.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

# NONLINEAR MODELS

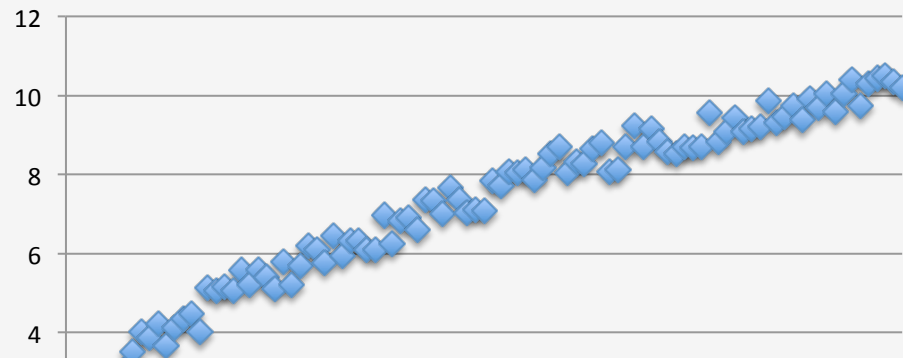
The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than the rest of the text. The logo is set against a light blue and orange background that resembles a stylized wave or a banner.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- In many situations the relationship between dependent and independent variables can not be captured using the preceeding models.
- For example, let us suppose that we want to estimate the number of individuals of some population, as time passes the population increases fast but after some point it increases at a lower pace because individuals compete for scarce resources.



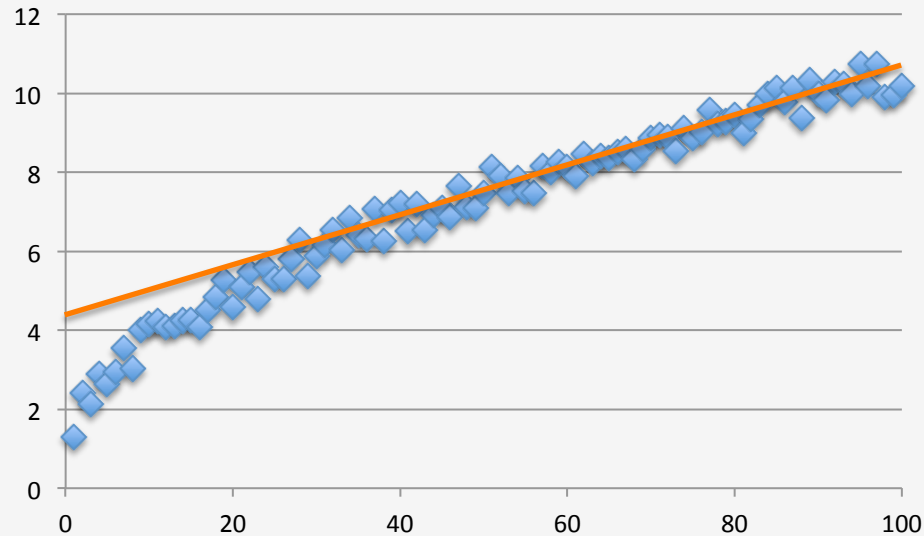
**Cartagena99**

**CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70**

---

**ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70**

- Note that a linear model would fail to capture the relationship between time and population size.



$$Y \neq \alpha + \beta t$$

- For example we could suggest

**Cartagena99**

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- After a nonlinear model has been proposed, the procedure to fit it to the data is very similar as for the linear case, again we have to find the optimal value of the parameters that index the model

$$Y = \hat{\alpha} + \hat{\beta}t + \hat{\gamma}t^2$$

- The complication comes from the fact that, in these cases, the normal equation can not be used and an iterative procedure must be used.
- There are many algorithms to estimate such parameters being the *Gauss-Newton method* one of the most commonly employed.

The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than "Cartagena". The text is set against a light blue background with a white swoosh underneath.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Notice, also, that some non-linear models can be transformed into linear ones using some transformation.
- For example, consider the *multiplicative model* (in contrast with the preceding which is an *additive model*):

$$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} \dots X_n^{\beta_n}$$

- Taking logarithms

$$\log(Y) = \log(\beta_0) + \beta_1 \log(X_1) + \beta_2 \log(X_2) + \dots + \beta_n \log(X_n)$$

- And renaming  $\log(Y) = Y'$ ,  $\log(\beta_0) = \beta_0'$ ,  $\log(X_i) = X_i'$ , we obtain:

$$Y' = \beta_0' + \beta_1 X_1' + \beta_2 X_2' + \dots + \beta_n X_n'$$



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
 CALL OR WHATSAPP: 689 45 44 70

- After the best values of the parameters are found, one can transform them back to find the model of interest, i.e.

$$\hat{\beta}_0 = e^{\hat{\beta}_0'}$$

- In practice, these transformations rarely do exist (except in very particular settings) and one has to consider the original functional form and estimate its parameters directly.

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, dark green font. The '99' is significantly larger and more prominent than the 'Cartagena' part. The text is set against a light blue background with a white swoosh underneath, all contained within a white rectangular box.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70



- Out of the infinite number of nonlinear models one can find some that are particularly useful and that are extensively used in machine learning.
- One of them is the *univariate logistic model*:

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

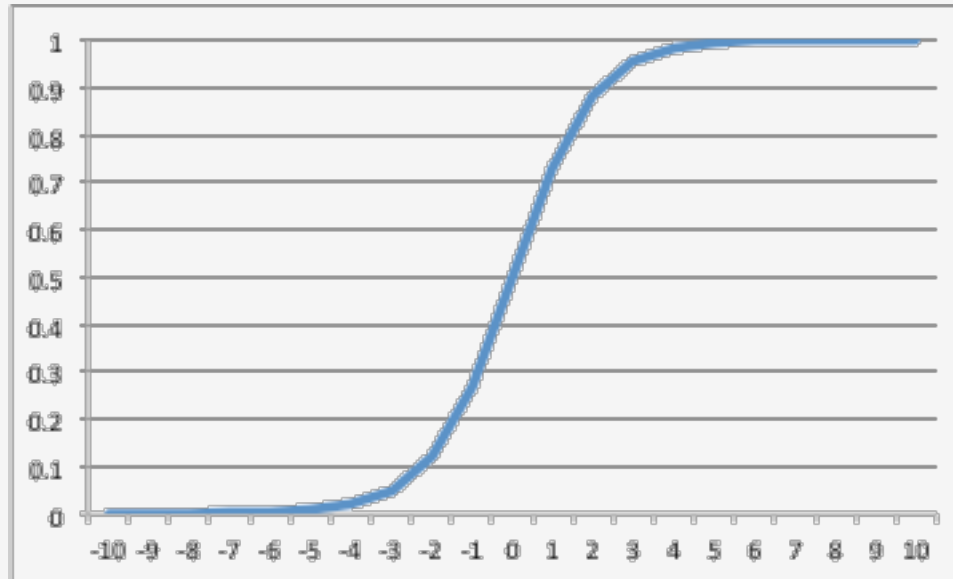
- Note that this model has a relevant property, when  $X$  is positive and very large  $Y=1$  and when  $X$  is negative and very large then  $Y=0$ .
- The response variable (output) is, then, a bounded number between  $[0,1]$ , and so it can be interpreted as a probability.

The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than "Cartagena". The text is set against a background of light blue and orange geometric shapes, including a large, faint "99" watermark.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70



- Of course we can also consider the *multivariate logistic model*:

$$Y = \frac{1}{1 + e^{-\beta_0 - \beta_1 X_1 - \beta_2 X_2 - \dots - \beta_n X_n}}$$

**Cartagena99**

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP: 689 45 44 70

- As mentioned, logistic models can be used in situations where one wants to determine the probability that some event happens ( $Y=1$ ) given some predictor variables.
- For example, one might be interested in calculating the probability that one individual develops some particular symptom given that he has taken some combination of drugs by using some database of symptoms-drugs intake.
- After the model is built and tested, doctors could use it to predict whether or not some combination will provoke some reaction.
- The logistic function has been particularly important in the development of Artificial Neural Networks models since, for some time, it was the most widely employed transfer function

**Cartagena99**

CLASES PARTICULARES, TUTORIAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

# PARAMETRIC AND NONPARAMETRIC MODELS

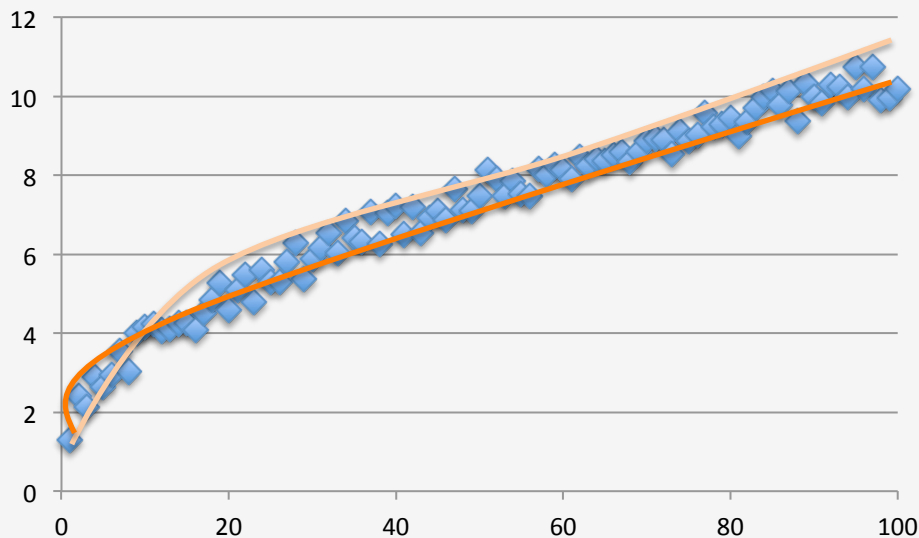
The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than the "Cartagena" part. The text is set against a light blue background with a white swoosh underneath, all contained within a white rectangular box.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- The problem is that all linear models are the same all but all the non-linear models are different, that is, there is a huge number of models that can be tried.
- In the preceding example we could propose any of these two models



$$Y = \alpha t + \beta t^2 + \gamma t^3 + \delta t^4$$

$$Y = \alpha \sqrt{t}$$

**Cartagena99**

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Obviously it would be infeasible to try many different models, even a relatively moderate number of them, we need to employ one single parameterization that can be used regardless of the relationship between Y and X.
- The models that we have seen up to now (univariate and multivariate linear models and logistic model) are *parametric* models, they employ a limited number of parameters which are estimated trying to fit the data.
- In contrast, a *nonparametric model* is the one which can not be characterized by a bounded number of parameters, that is, the number of parameters is not pre-determined.
- Notice that nonparametric models still use parameters, the name is a bit confusing since it can be interpreted that these

Cartagena99

CLASES PARTICULARES, TUTORIAS TECNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Some nonparametric models have the important property that they can approximate any function to any desired level just given enough complexity, this is called the *universal approximation property*.
- For example, assume that we want to find the relationship between one variable  $Y$  and two variables  $X_1, X_2$ , that is:

$$Y = f(X_1, X_2)$$

- Assume that we do not know the specific functional form of  $f$  so that e.g. we do not know whether  $f$  is linear, logistic etc.
- We may use a nonparametric model so that regardless which is  $f$  it can replicate the relationship between  $Y$  and  $X$ .

• One candidate is a polynomial of order  $n$ , where  $n$  is indeterminate:

**Cartagena99**

CLASES PARTICULARES, TUTORIAS TECNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- It can be demonstrated that there exist parameters

$$\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \hat{\alpha}_5, \dots$$

such that

$$Y = f(X_1, X_2) \approx \hat{\alpha}_1 X_1 + \hat{\alpha}_2 X_2 + \hat{\alpha}_3 X_1^2 + \hat{\alpha}_4 X_2^2 + \hat{\alpha}_5 X_1 X_2 + \dots$$

to any desired level.

- Notice that to the extent that we employ nonparametric models with the approximation property it is unimportant which is the specific functional form that we are trying to find, since the nonparametric models will “behave” exactly the same as the functional form of interest.

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP: 689 45 44 70



- Nonparametric models have a number of advantages against parametric ones being the most important the possibility to use a single model to find any particular relationship.
- Nevertheless they have several disadvantages too:
  - Their complexity needs to be controlled to avoid overfitting (more on this latter)
  - The computational load is generally important
  - Models are not easily understandable
  - Formal testing (e.g. significance of the parameters) is difficult or mostly impossible.

The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than the "Cartagena" part. The text is set against a light blue and white background with a subtle wave-like pattern.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

# BIAS AND VARIANCE

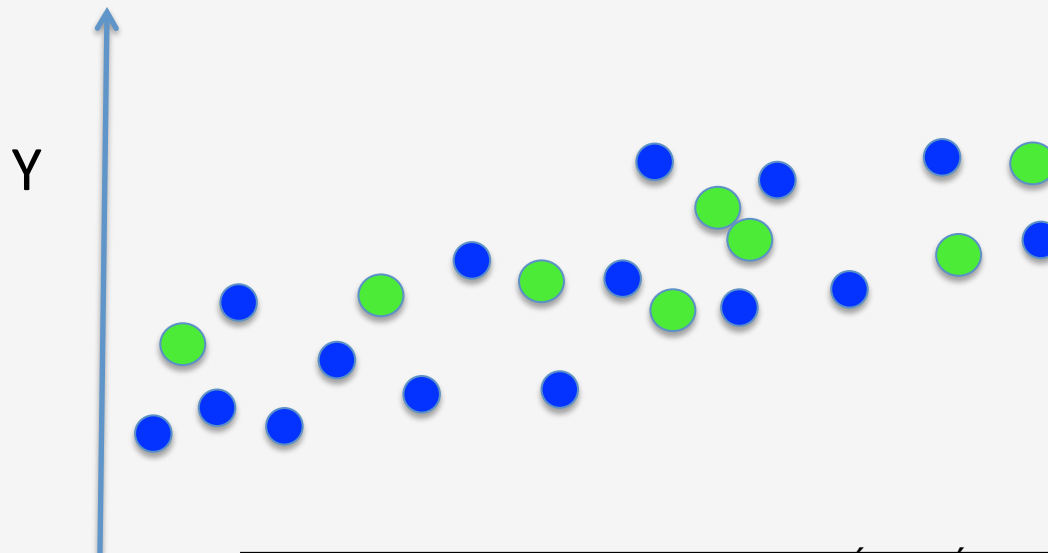
The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than the rest of the text. The logo is set against a light blue and orange background that resembles a stylized wave or a banner.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- As we have seen, models may have many different parameterizations and they will differ on how well they fit the data.
- For example, in the following example, let the **blue** dots represent the training data and the **green** ones the testing data.



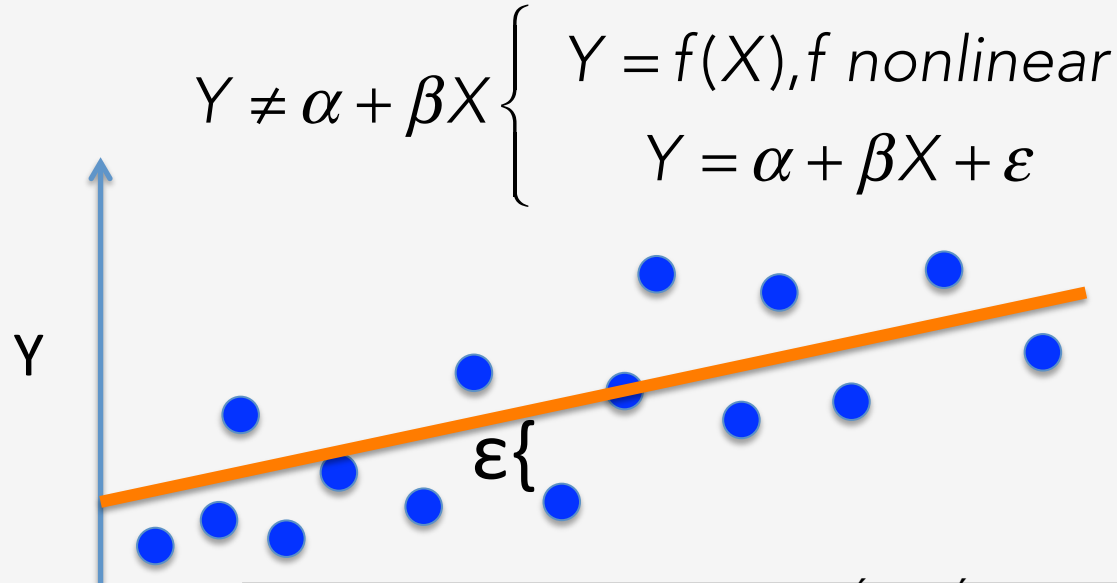
Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Let us suppose that we fit a linear model by minimizing the squared distance to the blue points.
- The linear model will fail to capture exactly the relationship between  $X$  and  $Y$  in the training data because not all the data fall along a line: relationship might be nonlinear or affected by noise.



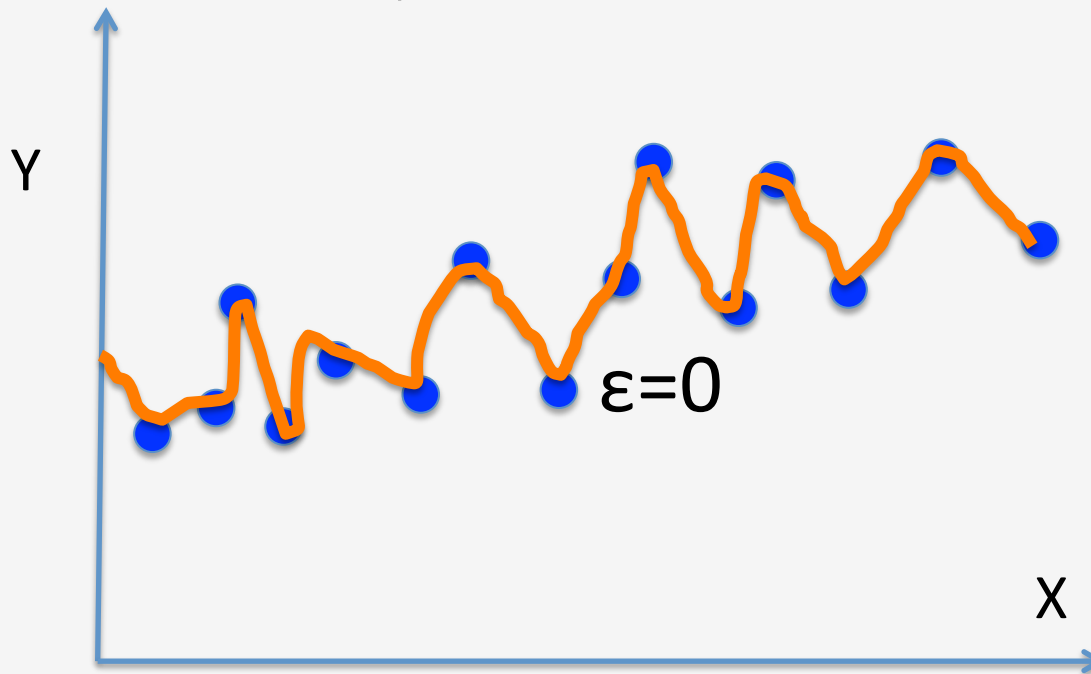
Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
 CALL OR WHATSAPP: 689 45 44 70

- Now, assume that we we employ a nonlinear/nonparametric model that, with enough complexity, could perfectly approximate the data points:



- The curve passes exactly over the **training data**.

**Cartagena99**

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- The second model is able to fit the training data because it is more flexible, it can exactly represent the training data it is said that it has a *low bias*.
- The first model has a lot of restrictions on the shape of its functional form this makes it unable to capture the training data, it has a *high bias*.
- In intuitive terms, we may consider bias as an over-simplification of the hidden relationship between X and Y.
- Notice that bias is easily avoidable: we can just increase the complexity of the model to reduce it, nevertheless, this will have consequences as we will see now.

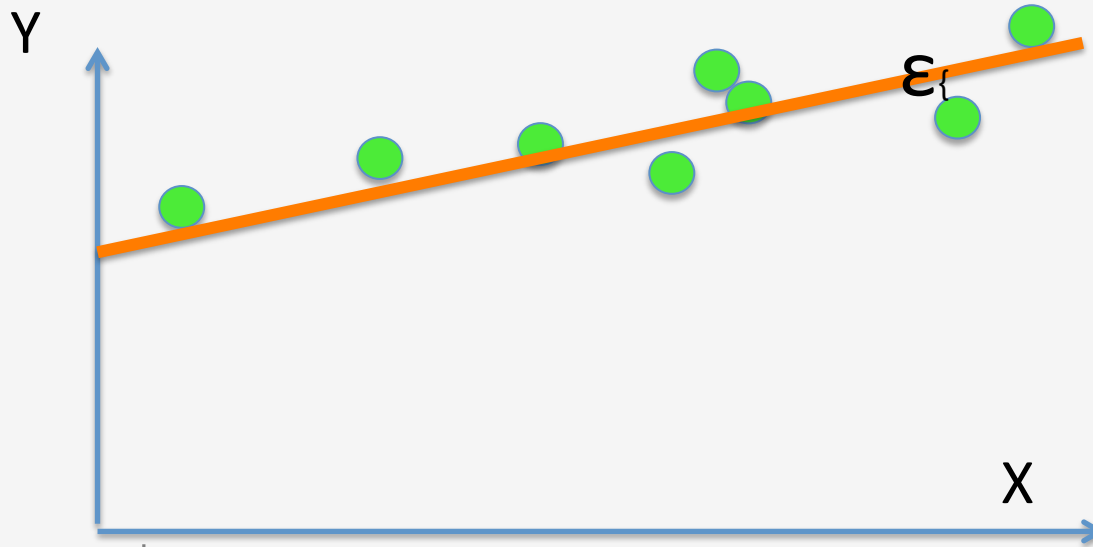
The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than the "Cartagena" part. The text is set against a light blue and white background with a subtle wave-like pattern. Below the text, there is a horizontal orange and yellow gradient bar.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

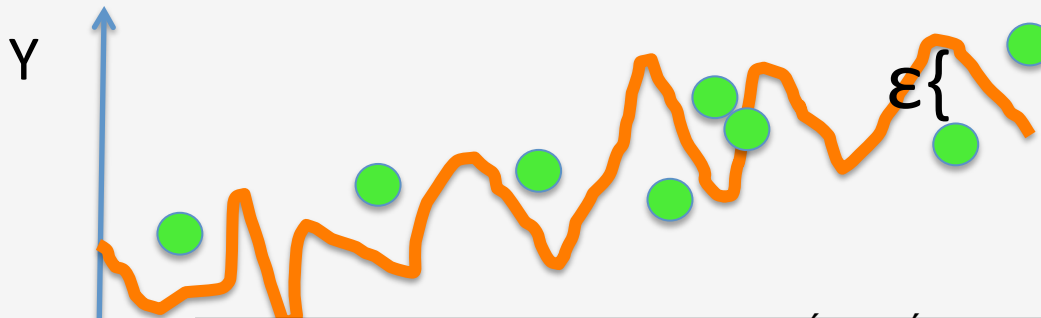
- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Now, let us focus on the testing data set, in the first model we have:



- And in the second:



**Cartagena99**

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- Notice that the errors on the testing set are higher for the second model than for the first one.
- We have to conclude that the first model is better on the testing set!!!.
- When we use the same models on new data the aparent flexibility of the second –more complex- model transforms into a drawback: it is perfect for the training data but this makes it being imperfect for the testing data.
- When a model exhibits a big difference between data and prediction on new data it is said to have a high variance.
- By contrast, the first model made few assumptions on the data, it

The logo for Cartagena99, featuring the text "Cartagena99" in a stylized font with a blue and orange gradient background.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70



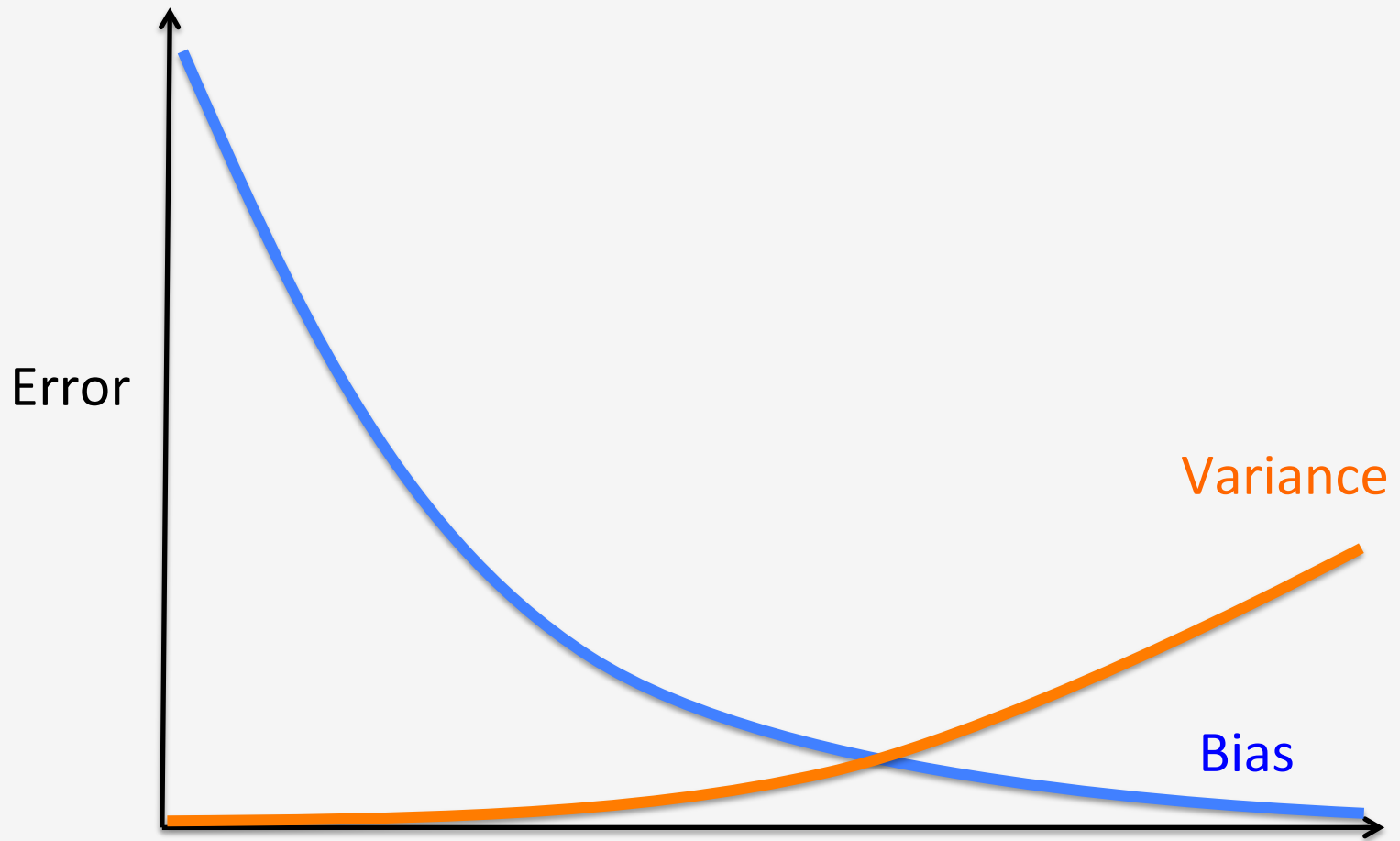
- So:
  - First model: high bias & low variance
  - Second model: low bias & high variance
- This is known as the *bias-variance dilemma*: flexible models have a lower bias but a higher variance than more rigid models.
- This dilemma is unavoidable, it is impossible to have models with low bias and also with a low variance, there is a tradeoff between these two.
- A good model is the one which has a good bias/variance relationship, it correctly captures the relationship in the training

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- The bias-variance dilemma is one of the most important aspects to be taken into account when building models under the ML nonparametric paradigm.
- One should not be tempted by complex models who provide a “perfect” explanation of the past because these models most of the times will provide bad forecasts.
- Of course, models which are bad when capturing relationships on the training data will continue being bad to make forecasts.
- Good models represent the training data but do not over-represent the training data.

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, dark green font. The '99' is significantly larger and more prominent than the 'Cartagena' part. The text is set against a light blue and white background with a subtle wave-like pattern. Below the text, there is a horizontal orange and yellow gradient bar.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- The error of any model can be decomposed in terms of the bias and variance concepts that we have just seen, omitting the mathematical details, it can be demonstrated that, for any model:

$$\text{Error} = (\text{Bias}^2 + \text{Variance}) + \text{Irreducible error}$$

- The above expression tell us that there are two kinds of errors:
  - One that we can call *model error* (bias + variance errors) which is due to the use of an unapropriate model and that can be reduced by correctly choosing the right one
  - Another that we call *unavoidable error* which is due to the stochastic relationship between input and output variables

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, dark green font. The '99' is significantly larger and more prominent than the 'Cartagena' part. The text is set against a light blue background with a white swoosh underneath.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

# UNDERFITTING AND OVERFITTING

The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than the "Cartagena" part. The text is set against a light blue background with a white swoosh underneath, all contained within a white rectangular box.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- The bias-variance dilemma is a direct consequence of the non-perfect (*stochastic*) relationship between explanatory and explained variables.

$$Y = f_{\theta}(X) + \varepsilon$$

- We need models with enough flexibility to fit the relationship that links the variables but it would be useless trying to fit the random component because, by definition, it is random and so it is unpredictable.
- When the model is too flexible it acts as a “database” that simply “remembers” the training data.
- Notice that the training data includes a random component that

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, dark green font. The '99' is significantly larger and more prominent than the 'Cartagena' part. The text is set against a light blue background with a white swoosh underneath.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- For example, assume the true relationship:

$$Y = 2X + \epsilon$$

- If we had the following database:

"true" Y	Y	X	$\epsilon$
4	3.6	2	-0.40
-2	-1.8	-1	0.20
6	6.6	3	0.60
0	-0.3	0	-0.30
-4	-3.7	-2	0.30
8	8.2	4	0.20

- We would not be interested in "learning" the random component  $\epsilon$  since, in another example we may have

"true" Y	Y	X	$\epsilon$
4	4.3	2	0.30
-2	-2.1	-1	-0.10

CLASES PARTICULARES, TUTORIAS TECNICAS ONLINE  
 LLAMA O ENVIA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
 CALL OR WHATSAPP:689 45 44 70

Cartagena99

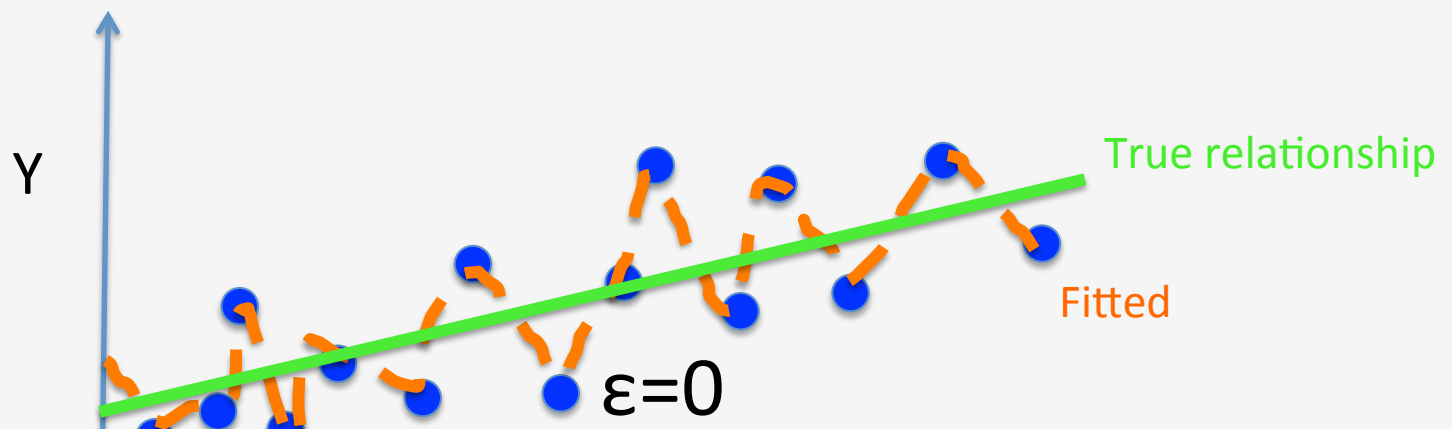
8

7.6

4

-0.40

- The random component does not tell us anything about the true relationship between X and Y.
- If we employ a model which is too complex so that it exactly represents the training data we are forcing it to “remember” random factors that will be useless for prediction.
- We are *overfitting* the data, we should have employed a simpler model.



Cartagena99

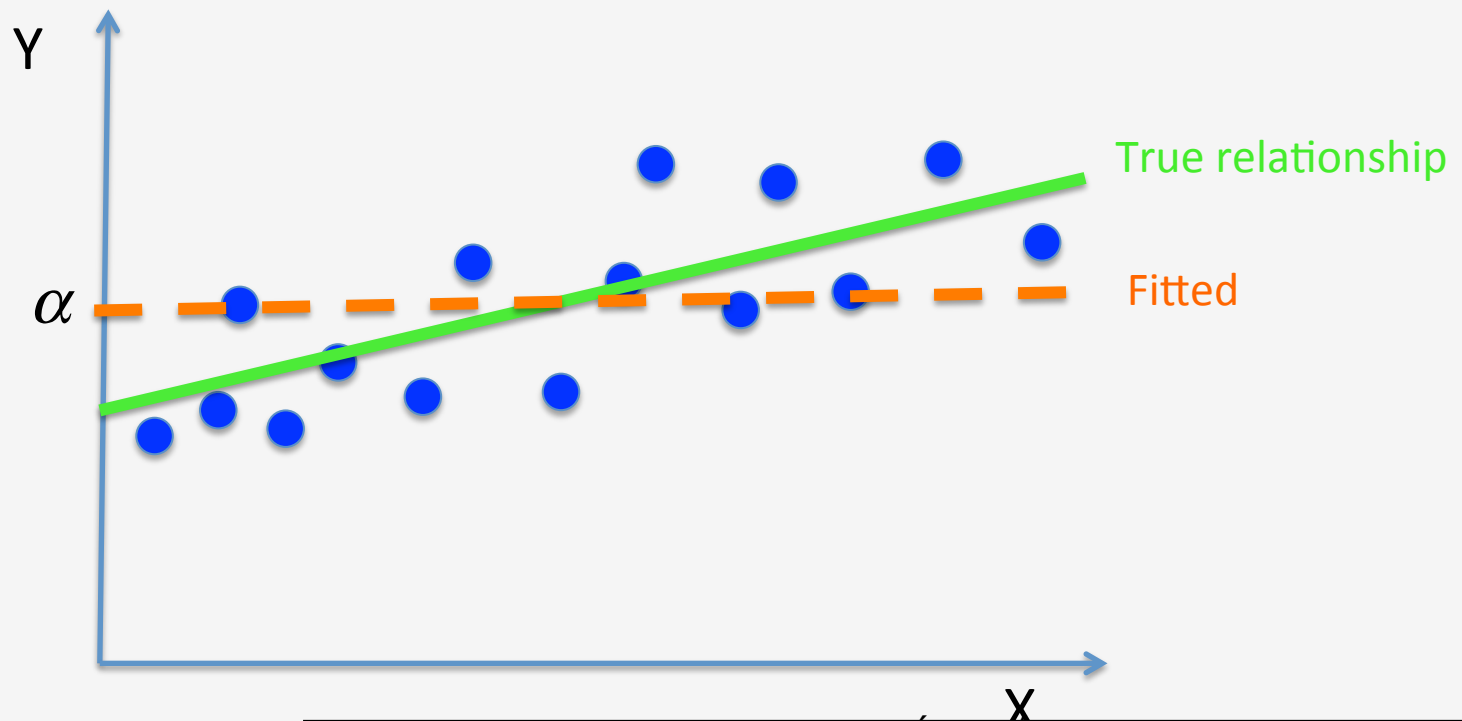
CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
 CALL OR WHATSAPP:689 45 44 70



- Alternatively, the model employed can be too simple and this will make it failing to capture the relationship between Y and X
- For example we may propose a trivial model  $Y = \alpha$

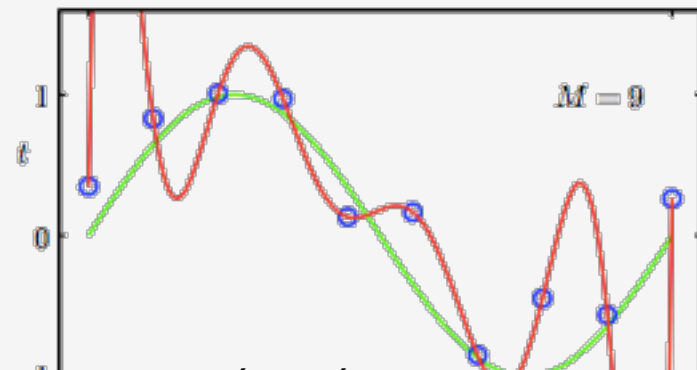
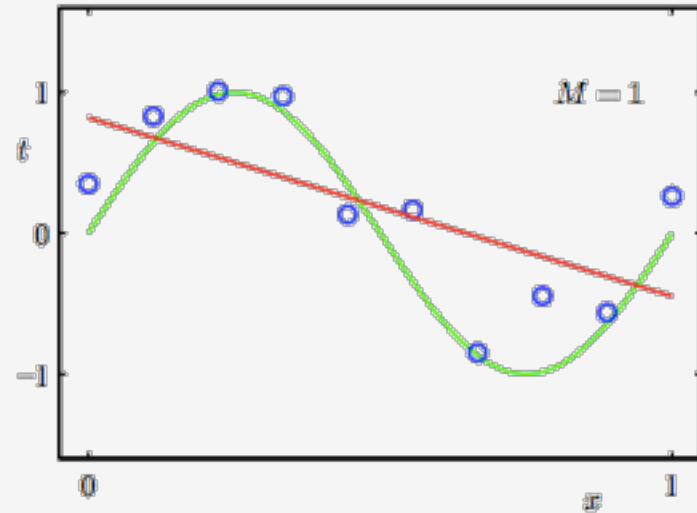
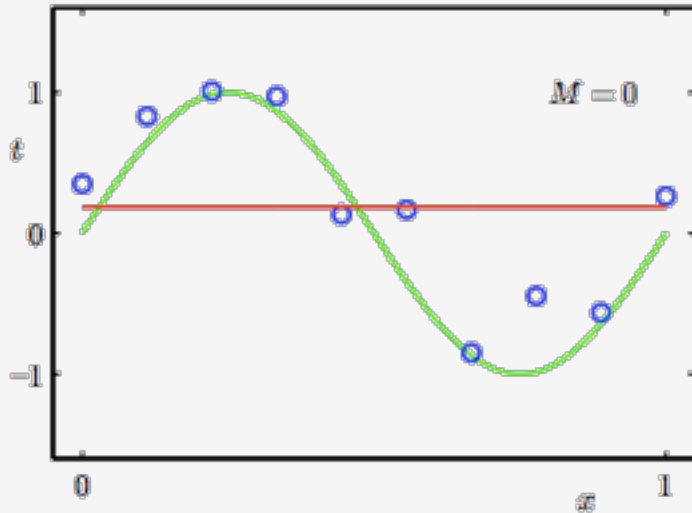


Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70



**Cartagena99**

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
 CALL OR WHATSAPP: 689 45 44 70

- Models which underfit the training data will have a poor performance on the testing data, they are “too simple”.
- Models which overfit the training data will have a poor performance on the testing data, they are “too complex”.
- There will be a model with an optimal complexity, neither too simple nor too complex that will capture the true relationship between X and Y.
- Notice that the concepts of overfitting and underfitting are closely related to the bias-variance dilemma:
  - Models with a high bias underfit data

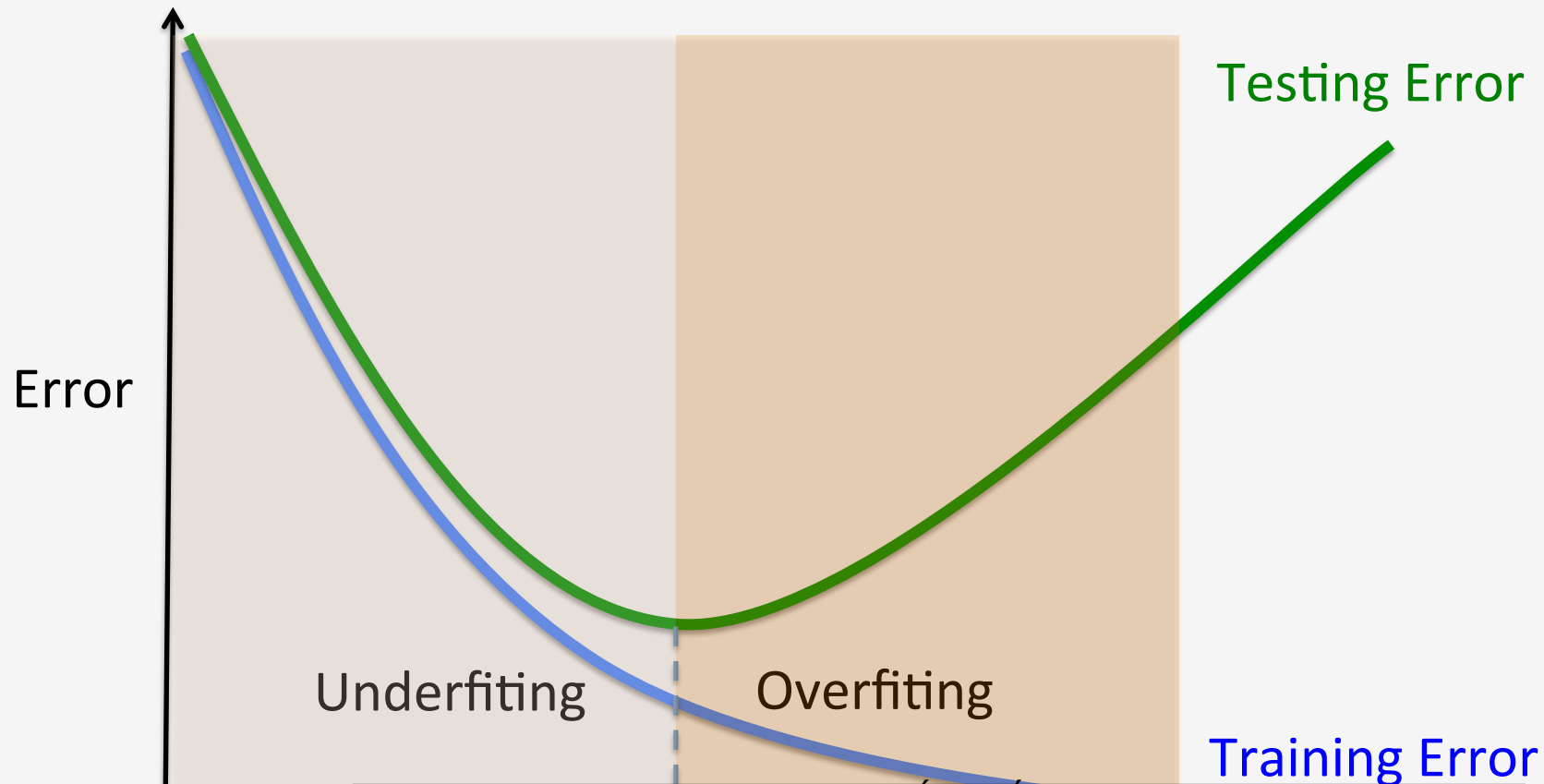
The logo for Cartagena99 features the text 'Cartagena99' in a stylized, dark green font. The '99' is significantly larger and more prominent than the rest of the text. The logo is set against a light blue and white background with a subtle wave-like pattern.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- In general terms we will have the following shape of the *learning curve*:



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVIA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- The optimal balance between bias and variance is mostly context or problem dependent: in some cases models are more prone to have a high variance, they easily overfit the data leading to bad forecasts.
- In other cases, one does not need to worry so much for controlling complexity since the model is relatively robust in terms of the bias-variance trade off.
- As we will see later, there are a number of techniques to control the complexity of the models to attain a reasonable balance between representation and forecasting power.

The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than the "Cartagena" part. The text is set against a light blue background with a white swoosh underneath, all contained within a white rectangular box.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

- The optimal balance between bias and variance is mostly context or problem dependent: in some cases models are more prone to have a high variance, they easily overfit the data leading to bad forecasts.
- In other cases, one does not need to worry so much for controlling complexity since the model is relatively robust in terms of the bias-variance trade off.
- As we will see later, there are a number of techniques to control the complexity of the models to attain a reasonable balance between representation and forecasting power.

The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than the rest of the text. The logo is set against a light blue and orange background with a wavy, abstract design.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70