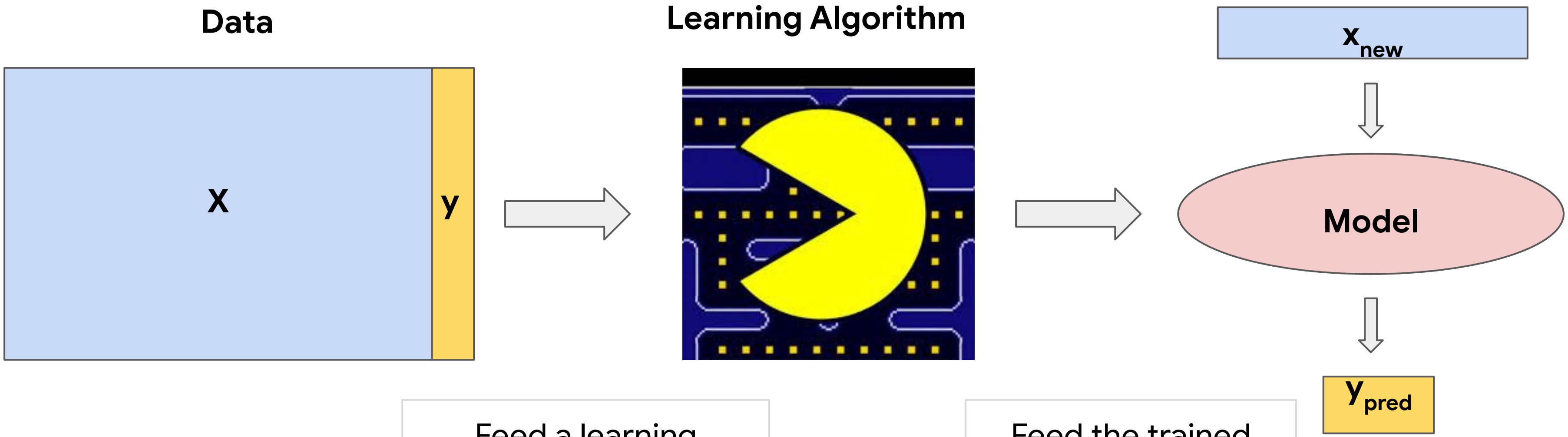
Statistical Learning: Least Squares

Jesse Gronsbell
Department of Statistical Sciences, University of Toronto

Part I: Least Squares Estimation

Now matter what you call it, this is what you do



Feed a learning algorithm your data for ***estimation*** or ***training*** of the model.

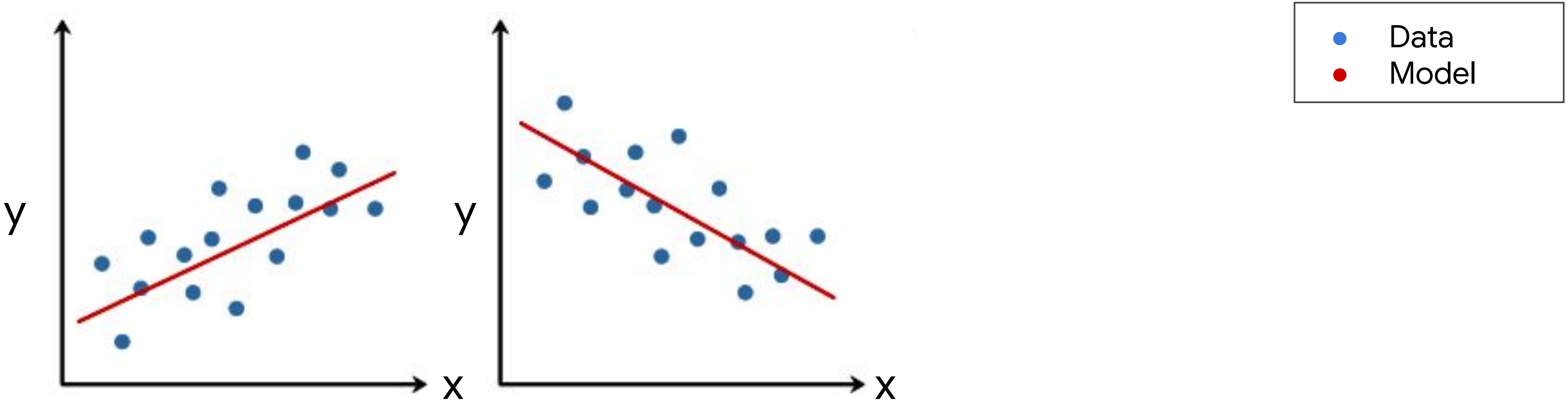
Feed the trained model new data to make a ***prediction*** about the outcome.

Goal of statistical learning

Use data to estimate a statistical model to capture the behavior of a process so that future **predictions** can be made

Goal of statistical learning

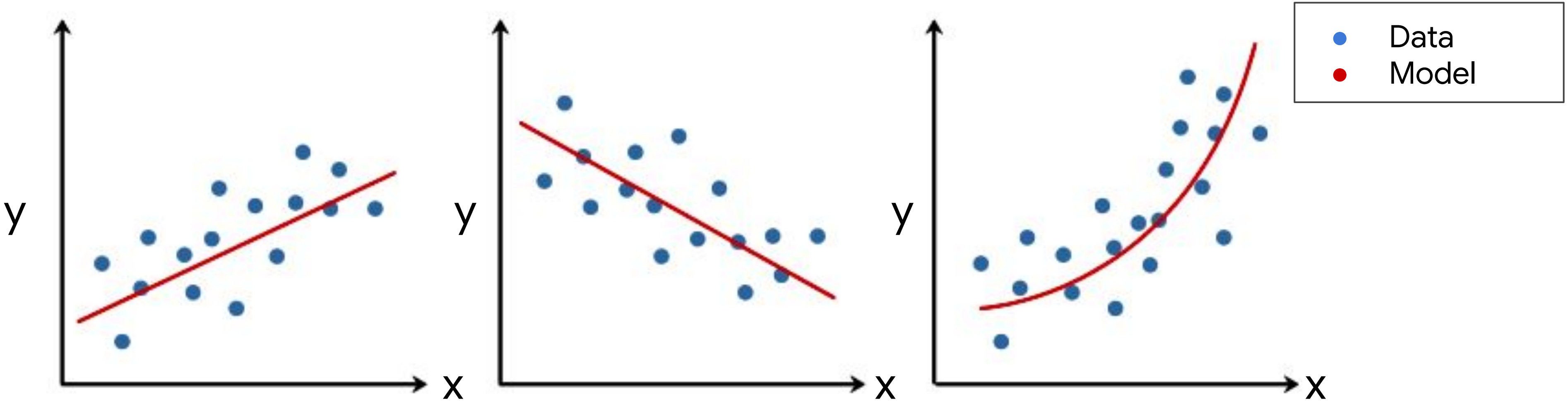
Use data to estimate a statistical model to capture the behavior of a process so that future **predictions** can be made



Linear Model

Goal of statistical learning

Use data to estimate a statistical model to capture the behavior of a process so that future **predictions** can be made



Linear Model

Exponential Model

Key ingredients of a statistical learning problem

1. **Process**: The problem you are interested in modeling

   ex. Can we predict ice cream sales from daily temperature?

Key ingredients of a statistical learning problem

1. **Process**: The problem you are interested in modeling

   ex. Can we predict ice cream sales from daily temperature?

2. **Data**: Samples of both the outcome(s) and the covariate(s)

   ex. samples of (x,y) = (temp, sales) from an ice cream truck in Summer 2021

Key ingredients of a statistical learning problem

1. **Process**: The problem you are interested in modeling

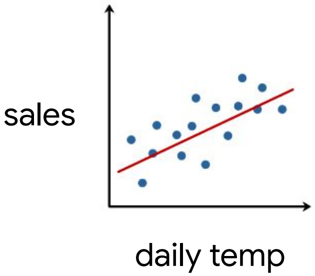
   ex. Can we predict ice cream sales from daily temperature?

2. **Data**: Samples of both the outcome(s) and the covariate(s)

   ex. samples of (x,y) = (temp, sales) from an ice cream truck in Summer 2021

3. **Model:** A function that takes in an x and outputs a value of y

   ex. A linear model: expected sales = 10 + 5*temp



Key ingredients of a statistical learning problem

1. **<u>Process</u>**: The problem you are interested in modeling

   ex. Can we predict ice cream sales from daily temperature?

2. **<u>Data</u>**: Samples of both the outcome(s) and the covariate(s)

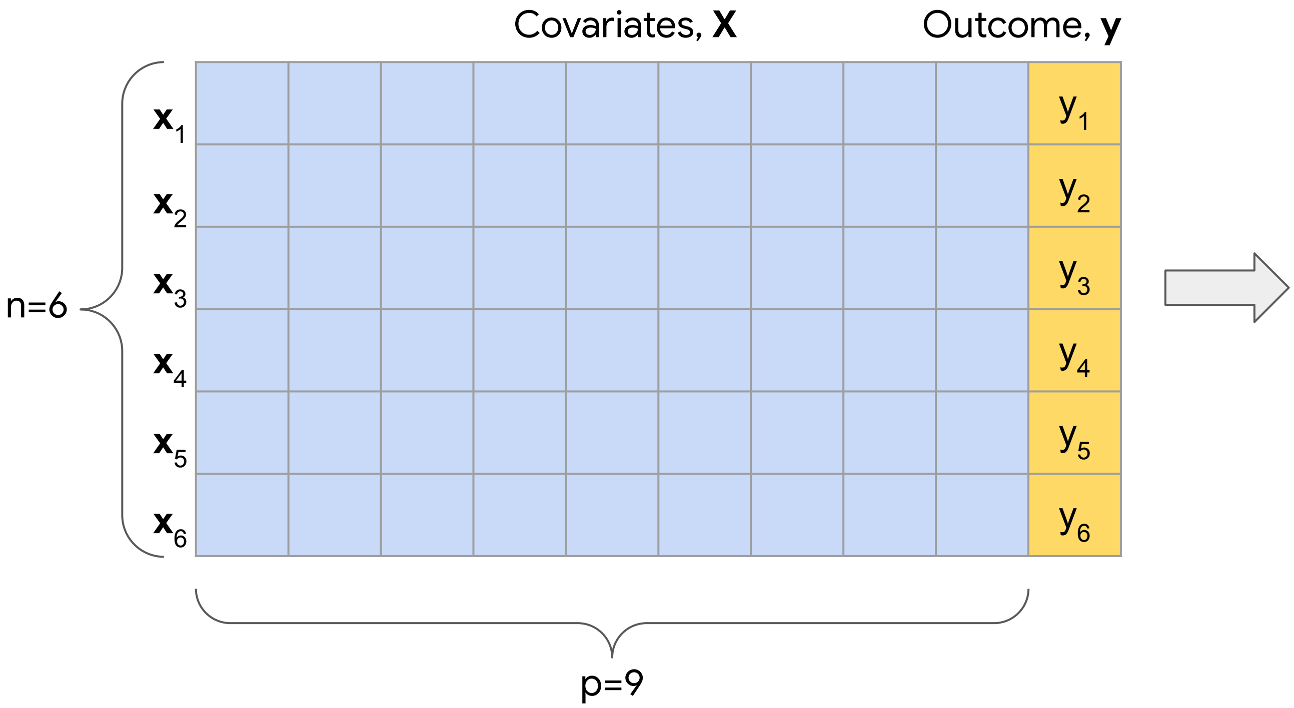
   ex. samples of (x,y) = (temp, sales) from an ice cream truck in Summer 2021

3. **<u>Model</u>:** A function that takes in an x and outputs a value of y

   ex. A linear model: expected sales = 10 + 5*temp

We'll focus on different types of models and how to estimate them!

But first... the data matrix in mathematical notation



$\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_6, y_6)\}$

But first… the data matrix in mathematical notation

				Covariates, **X**					Outcome, **y**
$\mathbf{x}_1$	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	y_1
$\mathbf{x}_2$	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	y_2
$\mathbf{x}_3$	x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	x_{36}	x_{37}	x_{38}	x_{39}	y_3
$\mathbf{x}_4$	x_{41}	x_{42}	x_{43}	x_{44}	x_{45}	x_{46}	x_{47}	x_{48}	x_{49}	y_4
$\mathbf{x}_5$	x_{51}	x_{52}	x_{53}	x_{54}	x_{55}	x_{56}	x_{57}	x_{58}	x_{59}	y_5
$\mathbf{x}_6$	x_{61}	x_{62}	x_{63}	x_{64}	x_{65}	x_{66}	x_{67}	x_{68}	x_{69}	y_6

n=6

p=9

$$\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_6, y_6)\}$$

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{i9})$$

But first… the data matrix in mathematical notation

Covariates, $\mathbf{X}$ Outcome, $\mathbf{y}$

$\mathbf{x}_1$	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	y_1
$\mathbf{x}_2$	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	y_2
$\mathbf{x}_3$	x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	x_{36}	x_{37}	x_{38}	x_{39}	y_3
$\mathbf{x}_4$	x_{41}	x_{42}	x_{43}	x_{44}	x_{45}	x_{46}	x_{47}	x_{48}	x_{49}	y_4
$\mathbf{x}_5$	x_{51}	x_{52}	x_{53}	x_{54}	x_{55}	x_{56}	x_{57}	x_{58}	x_{59}	y_5
$\mathbf{x}_6$	x_{61}	x_{62}	x_{63}	x_{64}	x_{65}	x_{66}	x_{67}	x_{68}	x_{69}	y_6

n=6

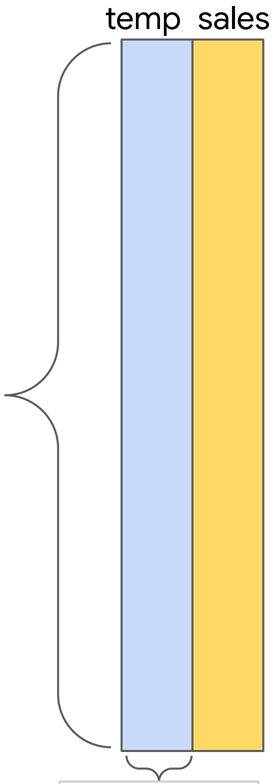
p=9

$$\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_6, y_6)\}$$

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{i9})$$

Looks messy, but not too bad! :)

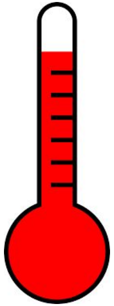
Visualizing the ice cream data matrix



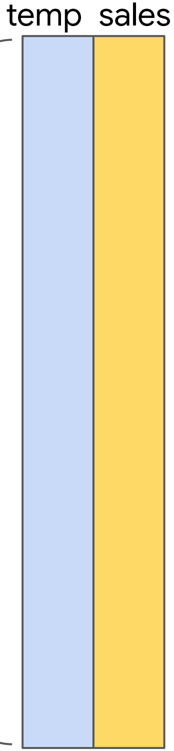
(x_i, y_i) = (temp, sales) for $i = 1, \ldots, n$

Each of the *n* rows is a day with a (temp, sales) pair

$p = 1$



Visualizing the ice cream data matrix

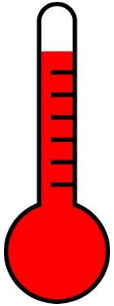


$(x_i, y_i) = (\text{temp}_i, \text{sales}_i)$ for $i = 1, \ldots, n$

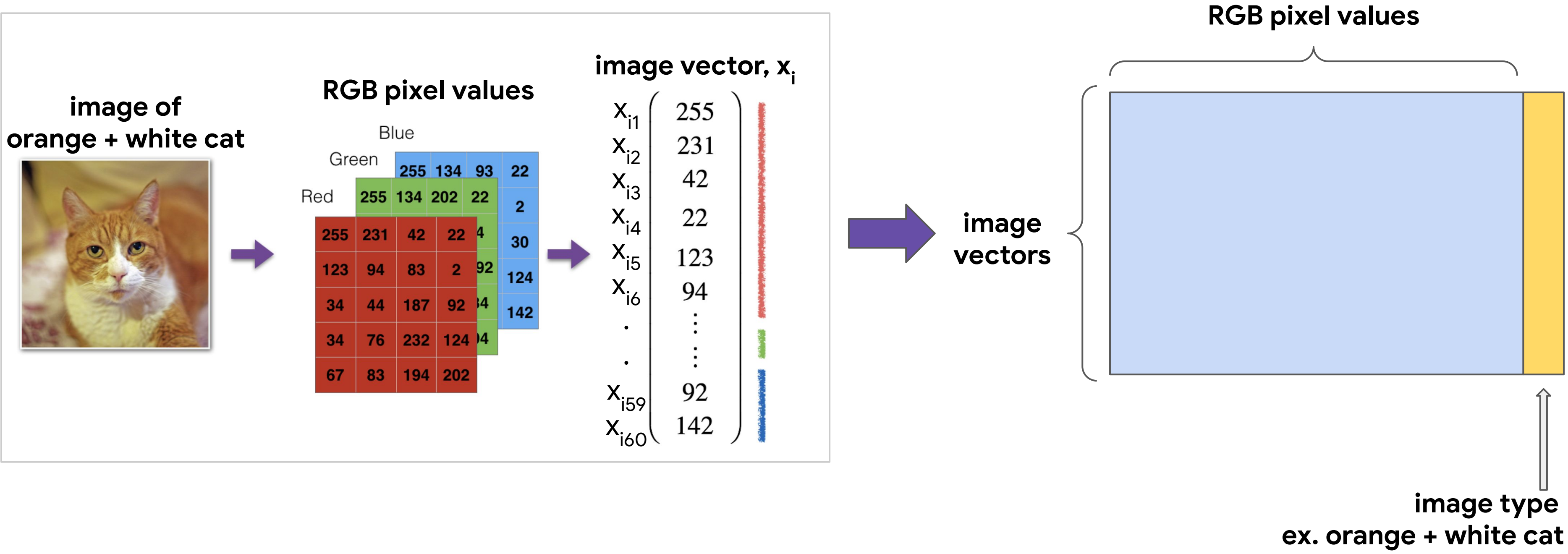
(x_1, y_1) , (x_2, y_2), …, (x_n, y_n)

Each of the *n* rows is a day with a (temp, sales) pair

p = 1



Looking ahead: Getting images into a data matrix

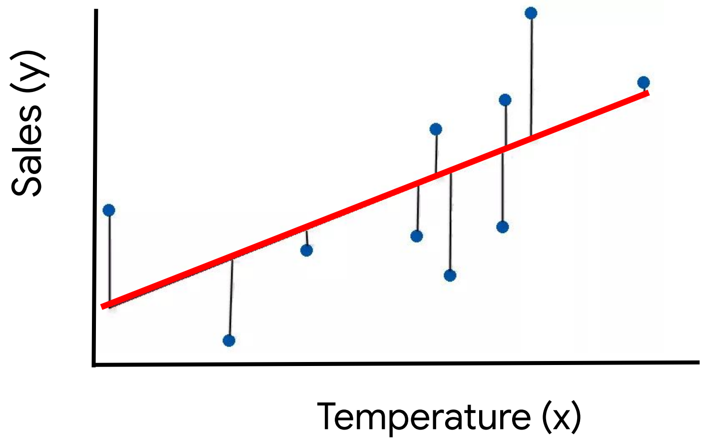


https://towardsdatascience.com/classifying-cat-pics-with-a-logistic-regression-model-e35dfb9159bb

Back to the ice cream example!

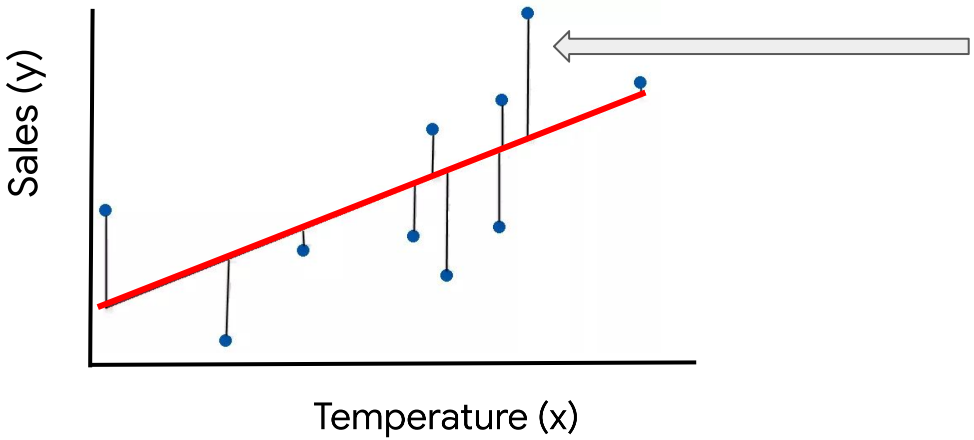
Model estimation: finding reasonable parameter values

<u>Problem</u>: Use the data on sales and temperature to estimate the unknown **parameters *m*** and ***b*** in the **linear model, expected sales = m*temperature + b**



Model estimation: finding reasonable parameter values

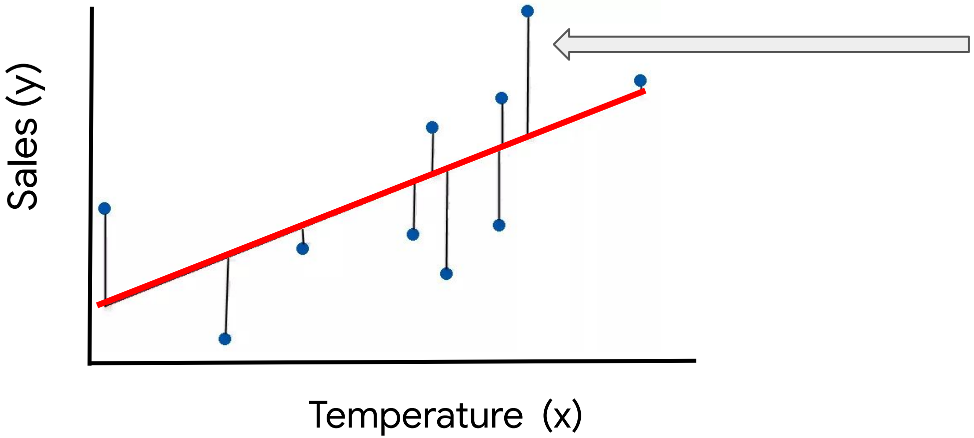
Problem: Use the data on sales and temperature to estimate the unknown **parameters *m*** and ***b*** in the **linear model, expected sales = m*temperature + b**



Error:
$[y_i - (mx_i + b)]^2$

One solution: Minimize the overall distance between the data and the linear model (i.e. ***the error***)

Solution: Least Squares (LS) estimation



Error:

$[y_i - (mx_i + b)]^2$

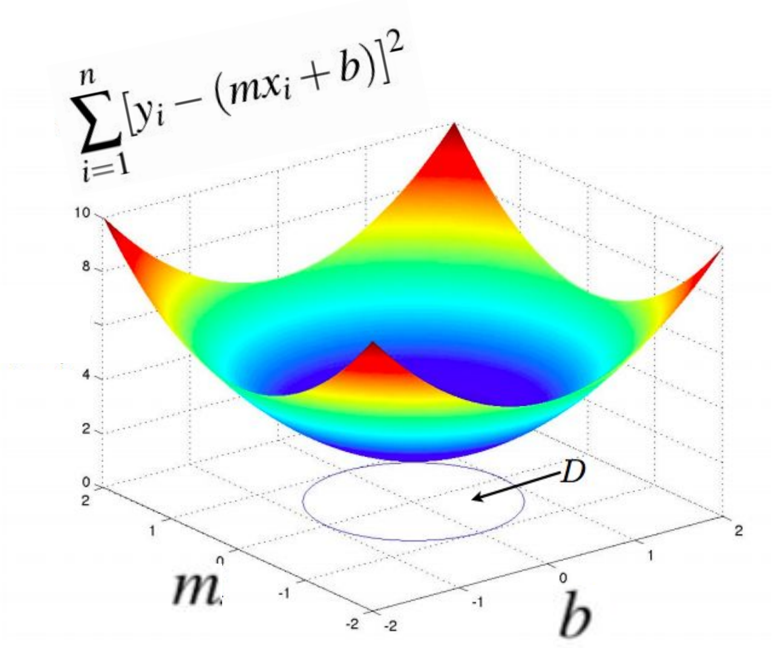
Find ***m*** and ***b*** so that the sum of the errors is as small as possible:

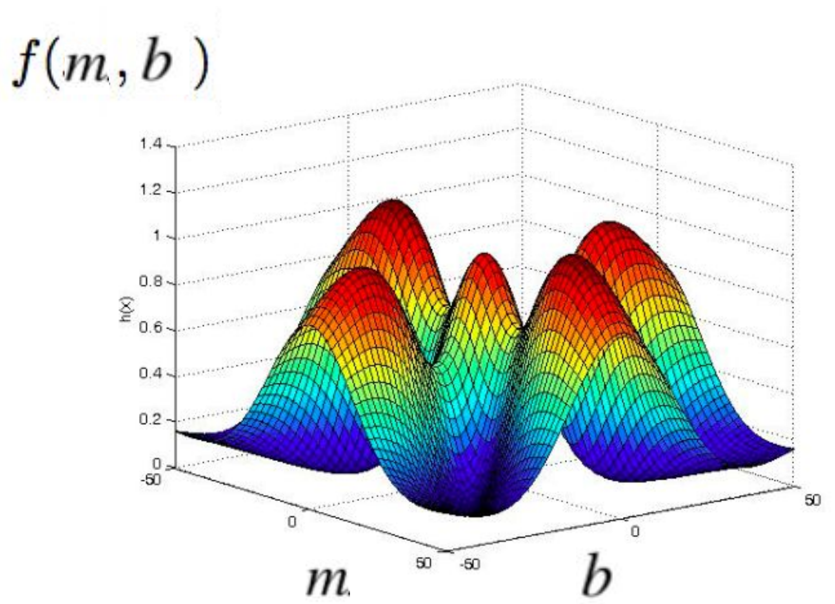
$$\sum_{i=1}^{n}[y_i - (mx_i + b)]^2 = [y_1 - (mx_1 + b)]^2 + [y_2 - (mx_2 + b)]^2 + \ldots + [y_n - (mx_n + b)]^2$$

But optimization isn't always easy!

There are simple scenarios :)

And others not so much :(





Next time

- How do we actually compute or *estimate* m and b?

- More on least squares

- Linear and logistic regression