IME/USP

MAC 0460 / 5832
Introduction to Machine Learning

• supervised learning • regression • classification •
• Statistics × Machine Learning •

Last class: From a purely computational point of view, ML is a meta-programming approach

 $\Longrightarrow$ generic model $w(0) \to w(1)$ $\Longrightarrow$ `output=eliptic?`

 $\Longrightarrow$ generic model $w(1) \to w(2)$ $\Longrightarrow$ `output=spiral ?`

 $\Longrightarrow$ generic model $w(2) \to w(3)$ $\Longrightarrow$ `output=spiral ?`

Our assumptions

Data

Availability of input-output examples: $\overbrace{(\mathbf{x}^{(n)}, y^{(n)})}^{\text{Data}}$, $n = 1, \ldots, N$

Existence of a plausible relationship between $\mathbf{x}$ and y

input

output (target)

Problems we would like to solve:

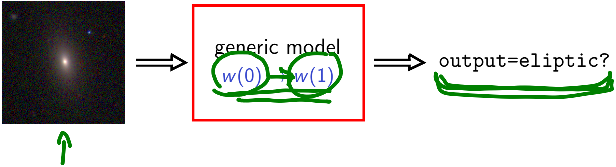
Find a mapping (computer program) that takes $\mathbf{x}$ to y

We will discuss regression and classification

ML techniques used to solve typical regression and classification problems are often called **supervised**

Supervision refers to the fact that the expected output y is known for each observation $\mathbf{x}$ in the input sample set

The expected output y is used in ML algorithms to guide their learning process

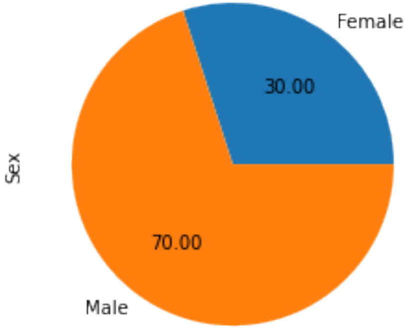


(later we will see that there is also the so called *unsupervised learning techniques*)

Let us use the data collected two days ago ($N = 130$) to formulate regression and classification problems!

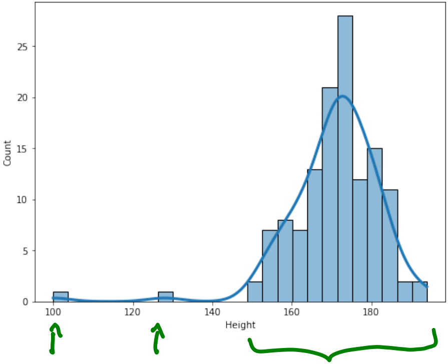
- Sex (Female / Male)
- Age
- Height
- Weight
- Shoe number
- Trouser number

Count by Sex

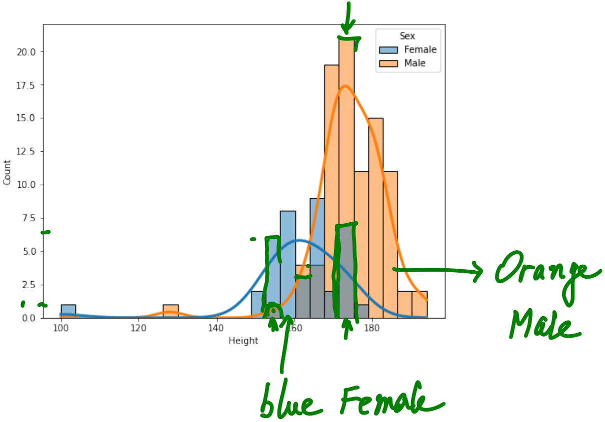


Sex $\in \{\mathrm{Female}, \mathrm{Male}\}$ is a categorical variable

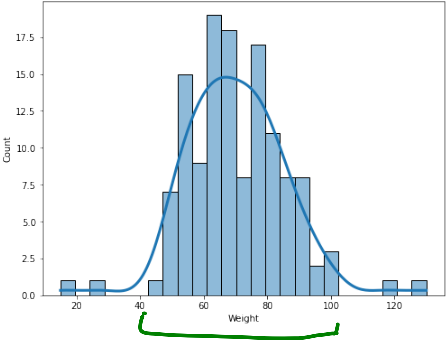
Height histogram



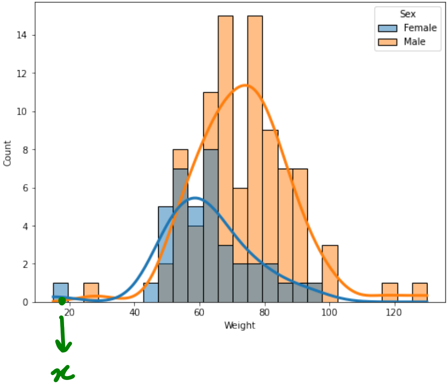
Height histogram, by Sex



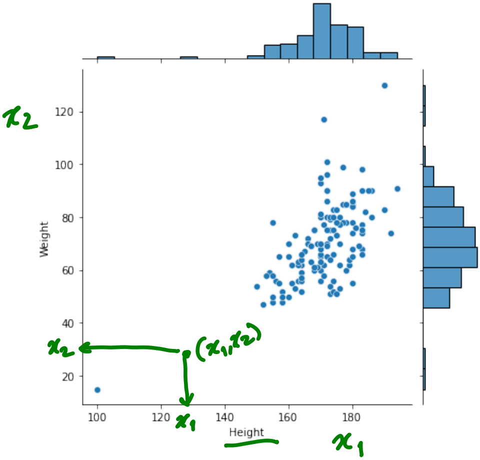
Weight histogram



Weight histogram, by Sex

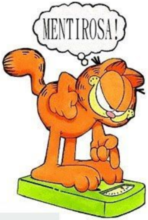


Height × Weight scatterplot



Problem 1

Suppose I need to record the height and weight of a person.



However, my is broken !

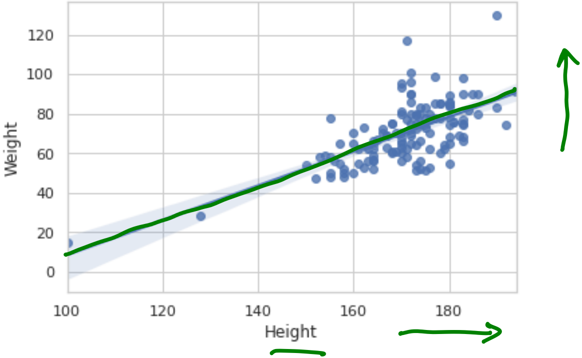
Is it possible to estimate/predict the weight of the person from his/her height only?

Problem 1

x : height (input)
y : weight ?

We would like to find f such that $y = f(x)$

Relation between x and y:



(we have $N = 130$ examples)

y is a continuous variable $\Longrightarrow$ Problem 1 is a **regression** problem

Generic input-output 'mapping' model

We would like to find f such that $y = f(x)$

$f = ??$ How do we choose f ?

For our Problem 1, let us choose $f(x; w, b) = w\,x + b$

w, b are the parameters of the model

$(x_i, y_i) \quad i = 1, \dots, N$

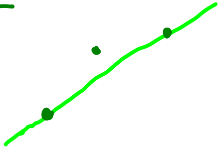
height weight $N = 130$

Problem 1: Stating it again

Given the **training set** $\underbrace{\{(\mathbf{x}^{(n)}, y^{(n)}) : \; n = 1, \dots, N\}}$
and
the family $\underbrace{\mathcal{F} = \{f(x) = wx + b : w, x, b \in \mathbb{R}\}}$,
choose
$f^* \in \mathcal{F}$ (or w and b) such that $\underbrace{f^*(x^{(n)}) = y^{(n)}}$, $n = 1, \dots, N$.

?

Is it possible ?



Do we always have $f^* \in \mathcal{F}$ such that $f^*(x^{(n)}) \cancel{=} y^{(n)}$, $n = 1, \ldots, N$?

NO.

We will be happy with

$f^* \in \mathcal{F}$ such that $f^*(x^{(n)}) \approx y^{(n)}$, $n = 1, \ldots, N$

That means we need some way to

measure how close $f^*(x^{(n)})$ is to $y^{(n)}$

Problem 1: Formalization

y

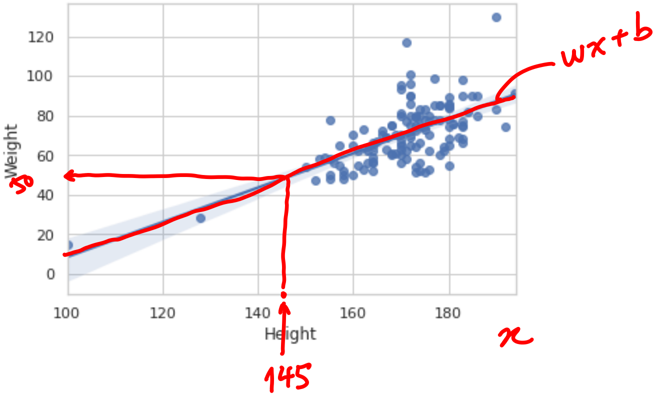
Given the **training set** $\{(\mathbf{x}^{(n)}, y^{(n)}):\ n = 1, \ldots, N\}$
and
the family $\mathcal{F} = \{f(x) = w\,x + b : w, x, b \in \mathbb{R}\}$,
determine

w and b such that $J(w, b) = \sum_{n=1}^{N} l((w\,x^{(n)} + b), y^{(n)})$ is minimum.

We could consider $l(a, b) = (a - b)^2$ and then

$$J(w, b) = \sum_{n=1}^{N} \Big((w\,x^{(n)} + b) - y^{(n)}\Big)^2$$

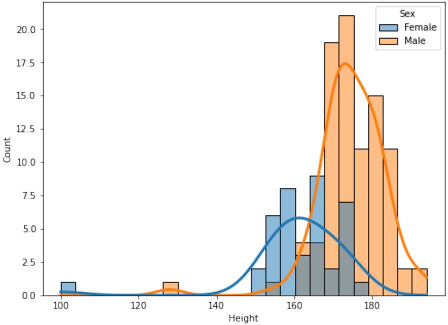
Once you have an optimal function f^*, you can use it to predict weight y from height x



$x = (x_1, x_2, \ldots, x_d)$

Problem 2

Suppose we know the height of a person. Can we guess correctly if this person is Female or not ?



Problem 2

input

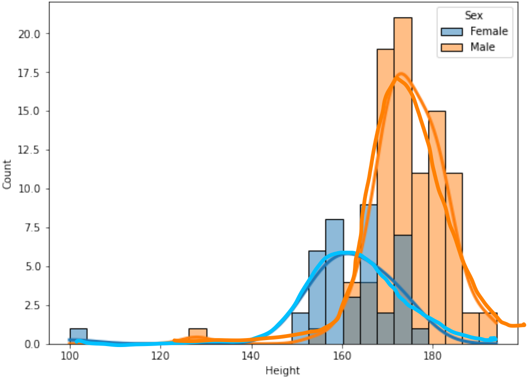
categorical variable

x : height
$y \in \{\text{Female}, \text{Male}\}$

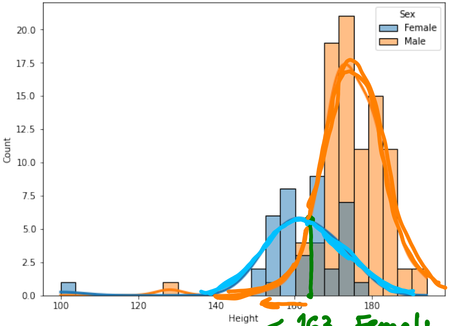
We would like to find f such that $y = f(x)$

Output is categorical $\Longrightarrow$ a classification problem

Relation between x and y: ?



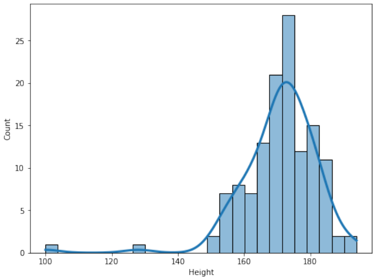
Given x and you have been asked to guess if the person is Female or Male, which would be your answer?

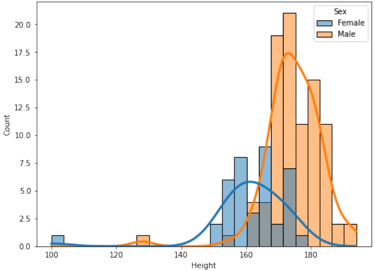


< 163 Female
> 163 Male

Is there a simple rule that will lead you to guess correctly most of the times ?

Probability distribution





$p(x)$

$p(x|y = \text{'Female'})$
$p(x|y = \text{'Male'})$

$y \in \{Female, Male\}$

$P(y = Male) = 0,7$

$P(y = Female) = 0.3$

Statistical approach

Bayes' Theorem

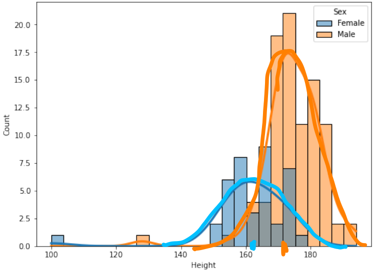
$$\boxed{p(y|x)} = \frac{p(y)\boxed{p(x|y)}}{\cancel{p(x)}}$$

If you know the distributions, you have the winning rule:

$$\boxed{y^* = \arg\max_y \{p(y|x)\}}$$

$p(y = Female/x) > p(y = Male/x) \Rightarrow y^* = F$

$<$ $\Rightarrow y^* = M$



In general:

$$p(y = \text{'Female'}) = p(y = \text{'Male'})$$

In QT1 dataset:

$$\hat{p}(y = \text{'Female'}) = 0.3$$
$$p(y = \text{'Male'}) = 0.7$$

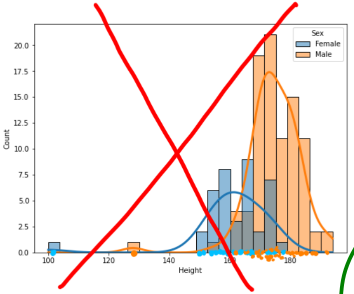
$$p(x|y = \text{'Female'}) \approx N(161, \sigma_F)$$

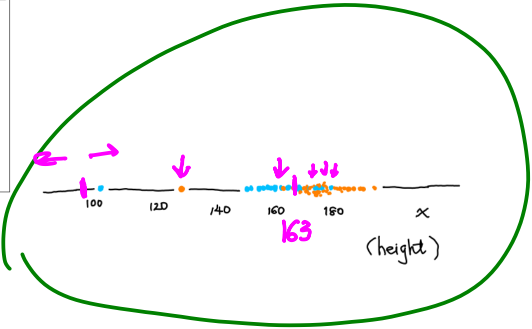
$$p(x|y = \text{'Male'}) \approx N(172, \sigma_M)$$

$$p(y|x) = \frac{p(y)p(x|y)}{p(x)}$$

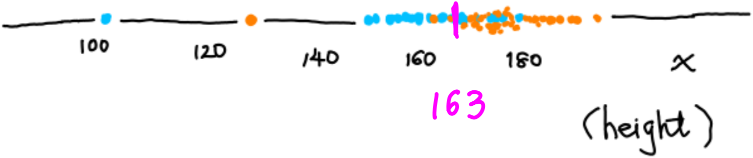
What if you do not have the distributions ?

You only have the observations.



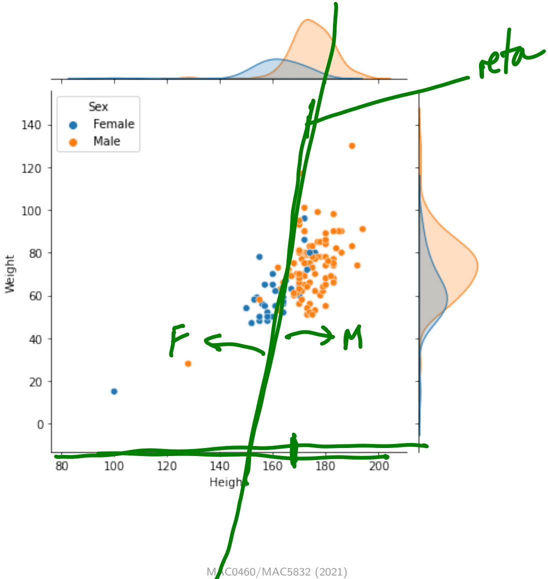


If you have only observations $(x^{(n)}, y^{(n)})$, what can you do to guess the correct y for any given x ?



Now, consider a two dimensional input $\mathbf{x} = (x_1, x_2)$

x_1: height $\quad$ x_2: width



I could use a straight line $w_1x_1 + w_2x_2 + b = 0$ to divide the space $\mathbb{R}^2$ (height $\times$ weight) in two regions and declare:

For a given $\mathbf{x} = (x_1, x_2)$,

$$w_1x_1 + w_2x_2 + b > 0 \Longrightarrow y = \textit{Female}$$

$$w_1x_1 + w_2x_2 + b < 0 \Longrightarrow y = \textit{Male}$$

Decision frontiers: Polinomials instead linear functions ?

Again, which family of mappings should we choose?
How to determine the optimal values of the parameters ?

Statistics $\times$ Machine Learning ?

Back to Bayes' Theorem

$$\boxed{p(y|x) = \frac{p(y)p(x|y)}{p(x)}}$$

Approaches:

1) Focus on the distribution of $\mathbf{x}$

$\Longrightarrow$ Generative approaches

Estatística

2) Focus on the expected output $p(y|\mathbf{x})$:

$\Longrightarrow$ discriminative approaches

$\Longrightarrow$ distribution-free learning

ML

Today's Summary

- Supervised learning ✓
- Regression problems ✓
- Classification problems ✓
- Probability distribution + Bayes' theorem $\Longrightarrow$ optimal classification
- No distribution $\Longrightarrow$ decision frontiers in the input space
- Learning: choose a family of mappings + solve an optimization problem

Next classes

We will learn some basic algorithms

- Perceptron for binary and linearly separable classification problems

- How to solve the MSE regression problem
  Analytical solution
  Iterative algorithm (employing gradient descent technique)

Tasks

Lecture 1: intro + perceptron
(slides 10 to 13 of Lecture 1; Section "1.1.2 A simple learning model" of the textbook);

Lecture 3: analytical solution of the linear regression problem
(slides 10 to 17 of Lecture 3; Section 3.2.1 of the textbook)