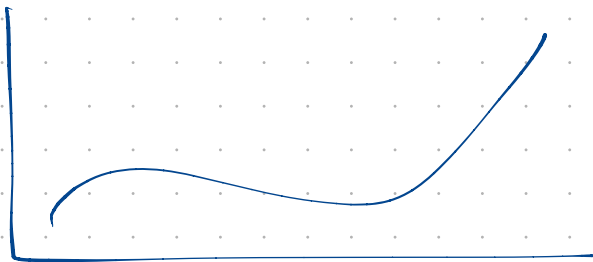




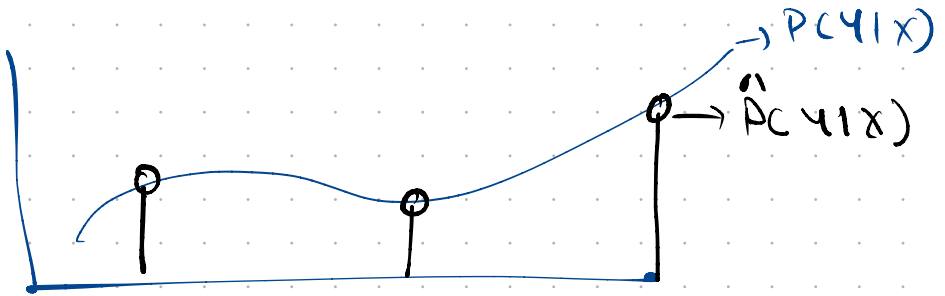
# Supervised Machine Learning

Task 1: Price of apartment  
from size of apartment



Real world  
(latent)  
distribution  
that we  
don't  
know

we don't have access to it. If we did,  
our prob is solved



Database

size (x)

$x_1$   
 $x_2$   
|  
 $x_n$

price (y)

$y_1$   
 $y_2$   
|  
 $y_n$

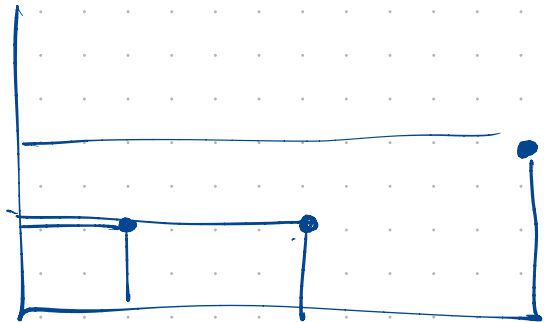
Approximate

$P(Y|X)$  from

$x \in \bar{X} \subseteq \bar{X}$   
inputs space

$y \in \bar{Y} \subseteq \bar{Y}$   
label space

How?

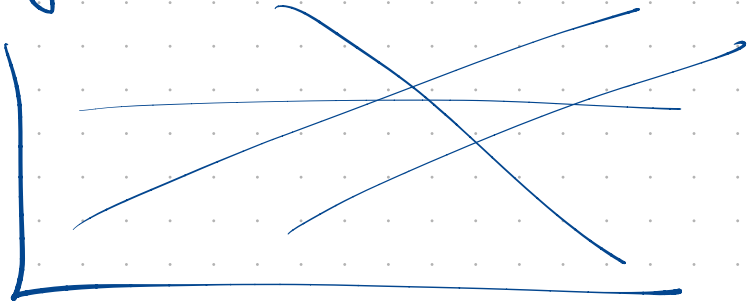


Assume characteristics of the real world distribution.

Say we assume that the actual (unknown) distribution is linear, i.e., it follows

$$\hat{y} = mx + c$$

This gives us two "parameters" to play around with —  $m$ ,  $c$



the  
**Goal**

Find  $m$  and  $c$  such that

$$\hat{y}_i \sim y_i \quad \forall \quad x_i, y_i \in \underbrace{X, Y}$$

↑  
our model's predictions for  $x_i$  ← are as close as possible

→ actual value for that  $x_i$

for all  $x_i, y_i$  in our dataset

How do we do  
that?

Optimizations