

The background features a complex network of thin grey lines and dots, forming a web-like structure. Scattered throughout are various triangles of different sizes and orientations, some with solid grey dots at their vertices. The overall aesthetic is minimalist and technical.

# **Are simple models safe?**

---

Mateusz Sieniawski  
21.05.21

Are these edible?



This seems so



Probably not

# Are these edible?



This seems so



Hmmm but this?



Probably not

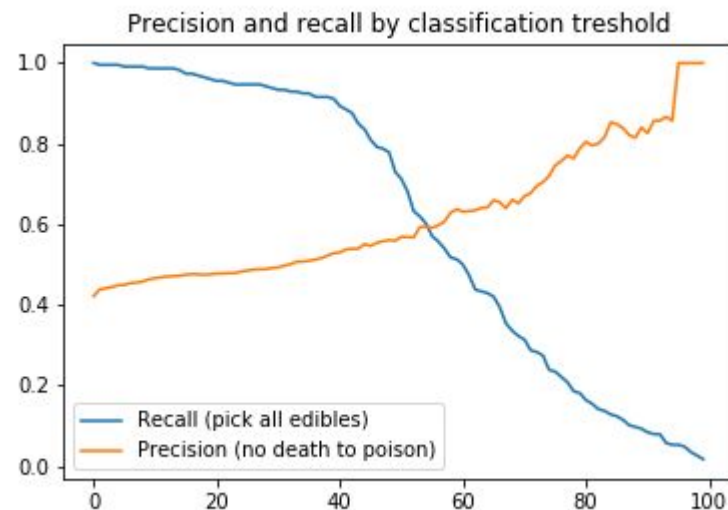
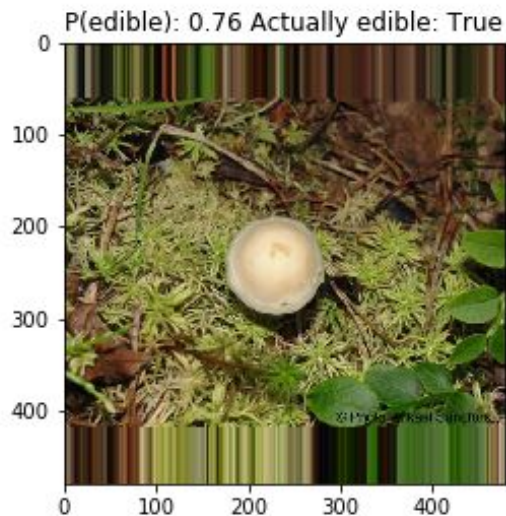
# Some tips from the Internet

1. Avoid mushrooms with white gills, a skirt or ring on the stem and a bulbous or sack like base called a volva. You may be missing out on some good edible fungi but it means you will be avoiding the deadly members of the Amanita family.
2. Avoid mushrooms with red on the cap or stem. Again you will be missing out on some good mushrooms but more importantly you won't be picking poisonous ones.
3. **Finally don't consume any mushrooms unless you are 100% sure of what they are. I know I have already mentioned this but it is by far the most important rule.**



<https://www.wildfooduk.com/articles/how-to-tell-the-difference-between-poisonous-and-edible-mushrooms/>

# Some guys actually implemented CNN classifying edible mushrooms



<https://tuomonieminen.github.io/deep-shrooms/>



# Some tips from the Internet

1. Avoid mushrooms with white gills, a skirt or ring on the stem and a bulbous or sack like base called a volva. You may be missing out on some good edible fungi but it means you will be avoiding the deadly members of the Amanita family.
2. Avoid mushrooms with red on the cap or stem. Again you will be missing out on some good mushrooms but more importantly you won't be picking poisonous ones.
3. **Finally don't consume any mushrooms unless you are 100% sure of what they are. I know I have already mentioned this but it is by far the most important rule.**
4. Some UK mushrooms are easy to identify so quite safe to eat, the Giant Puffball, Beefsteak Fungus, Porcelain Fungus, Cauliflower Fungus, Dryads Saddle, the Hedgehog Fungus ...



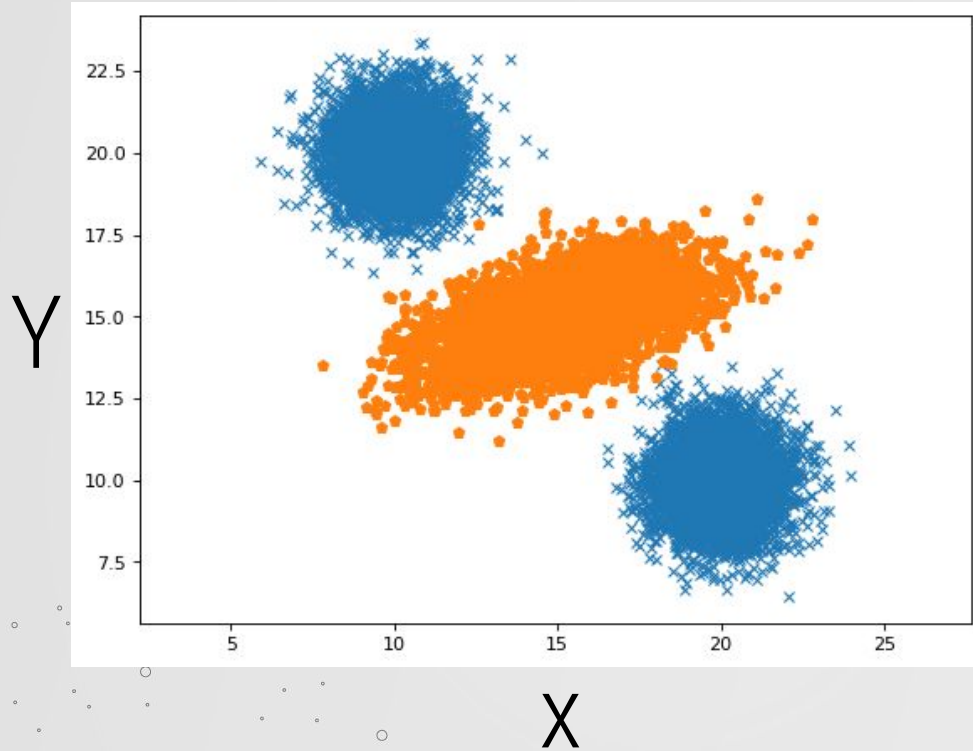
<https://www.wildfooduk.com/articles/how-to-tell-the-difference-between-poisonous-and-edible-mushrooms/>

# 1. Is linear regression too simple?

---



# XOR - typical counterexample against linear models

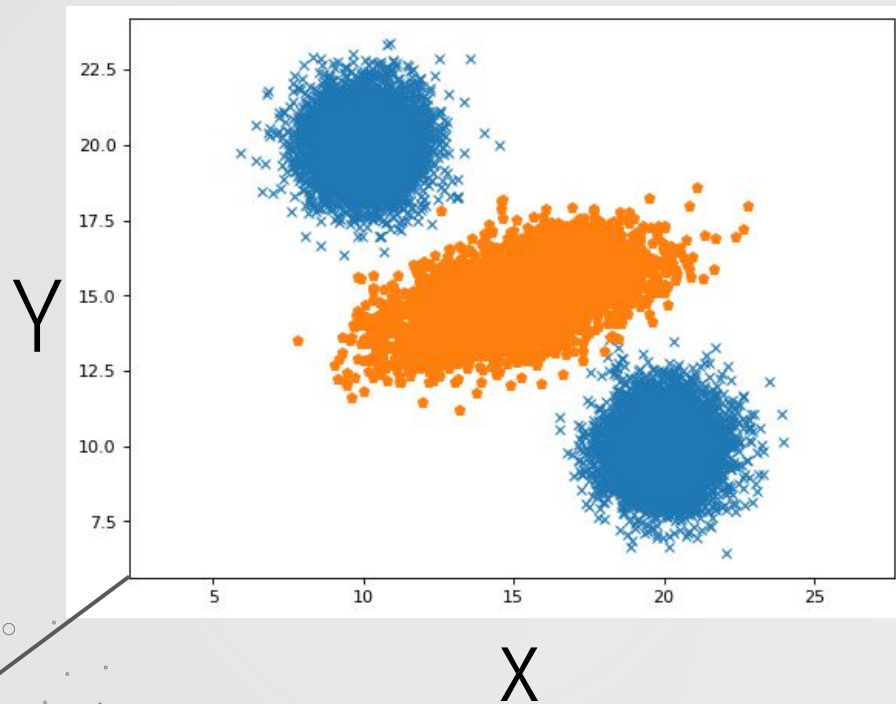


$$R^2 = -0.50$$

$$\begin{bmatrix} 3305 & 0 \\ 1645 & 0 \end{bmatrix}$$



# Clever feature engineering was enough



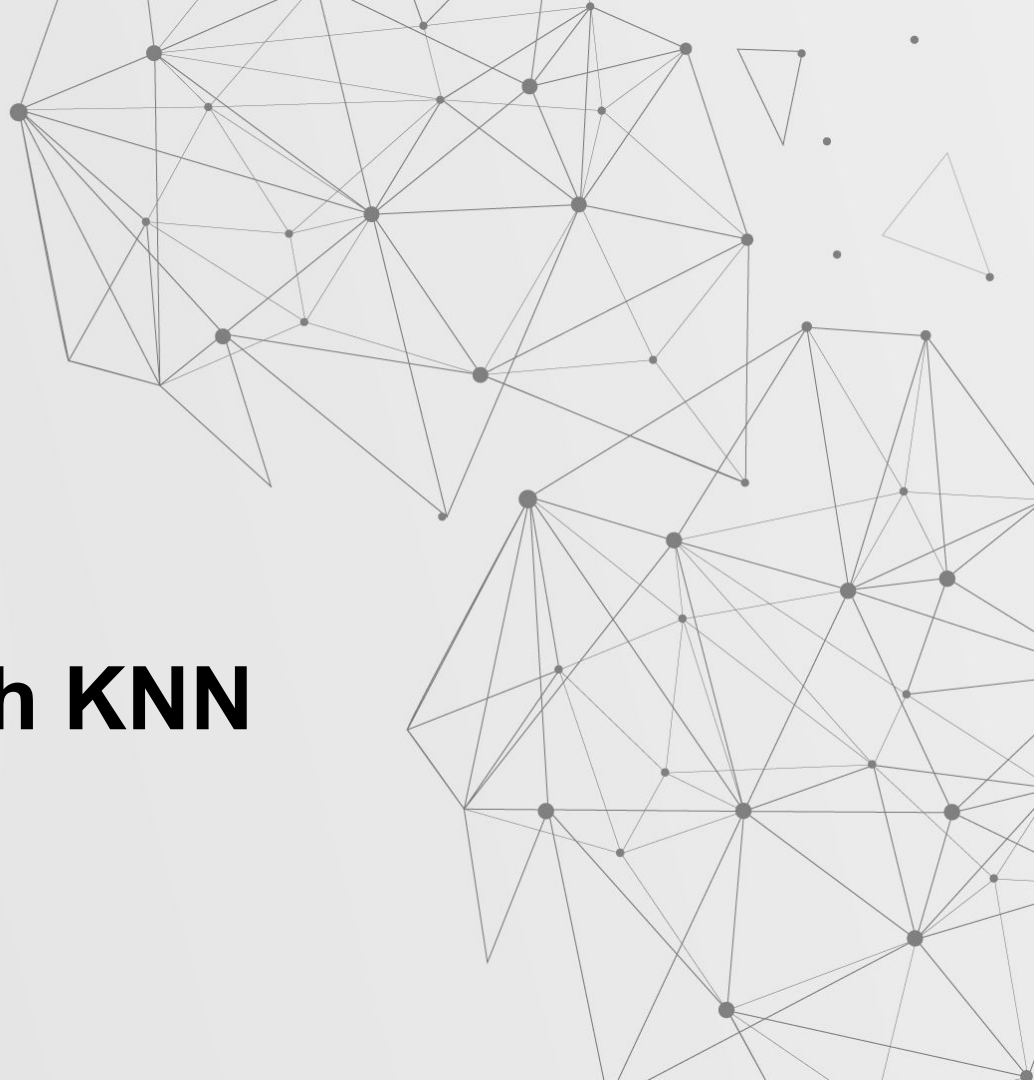
$X*Y$

$$R^2=1.00$$

$$\begin{bmatrix} 3302 & 3 \\ 0 & 1645 \end{bmatrix}$$

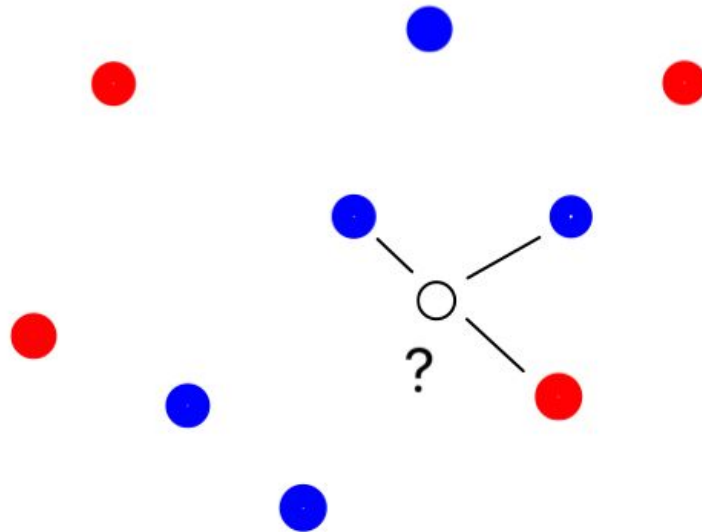
## 2. The problem with KNN

---



# A quick reminder on KNN

$k = 3$



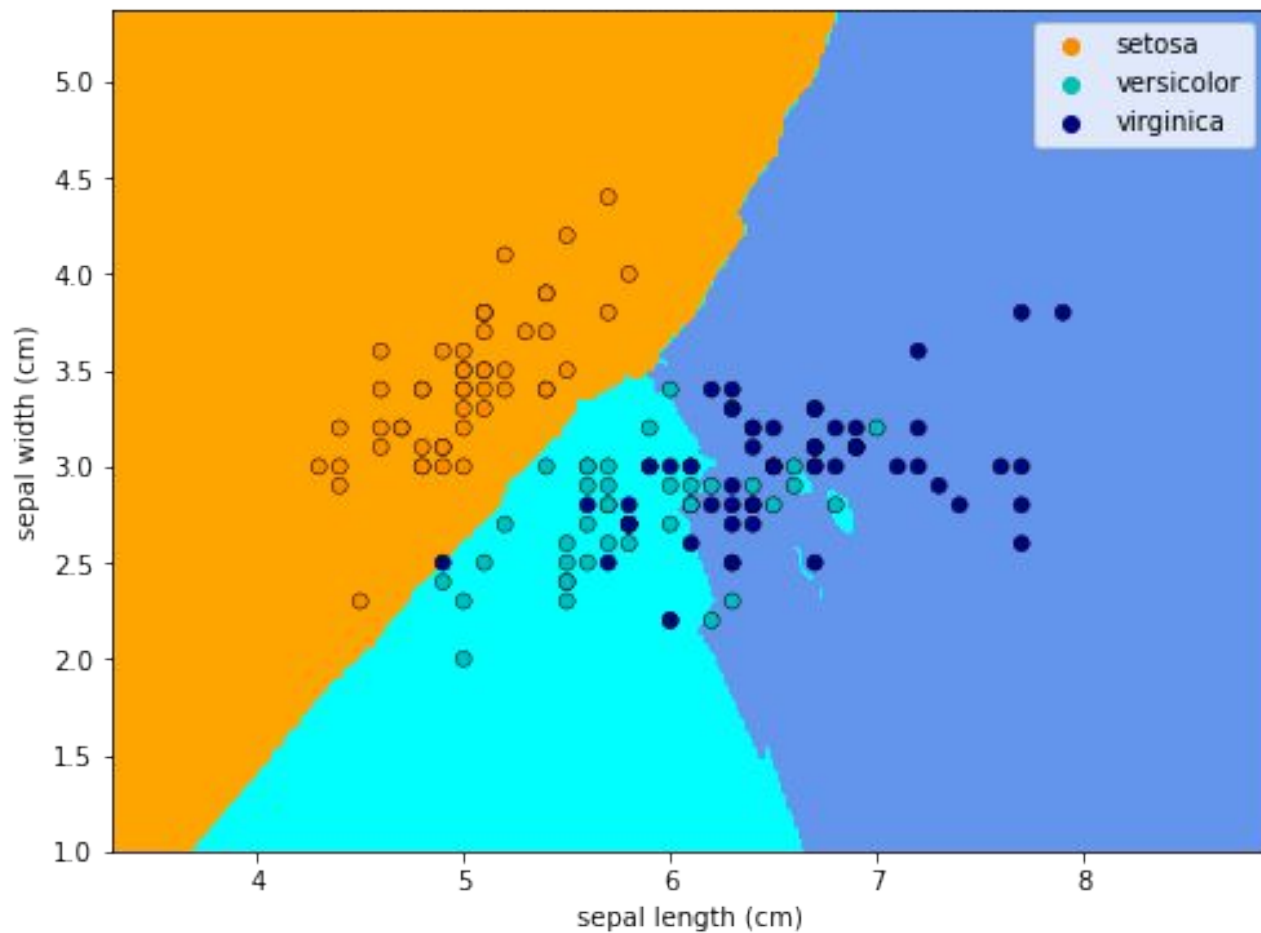


Kosaciec

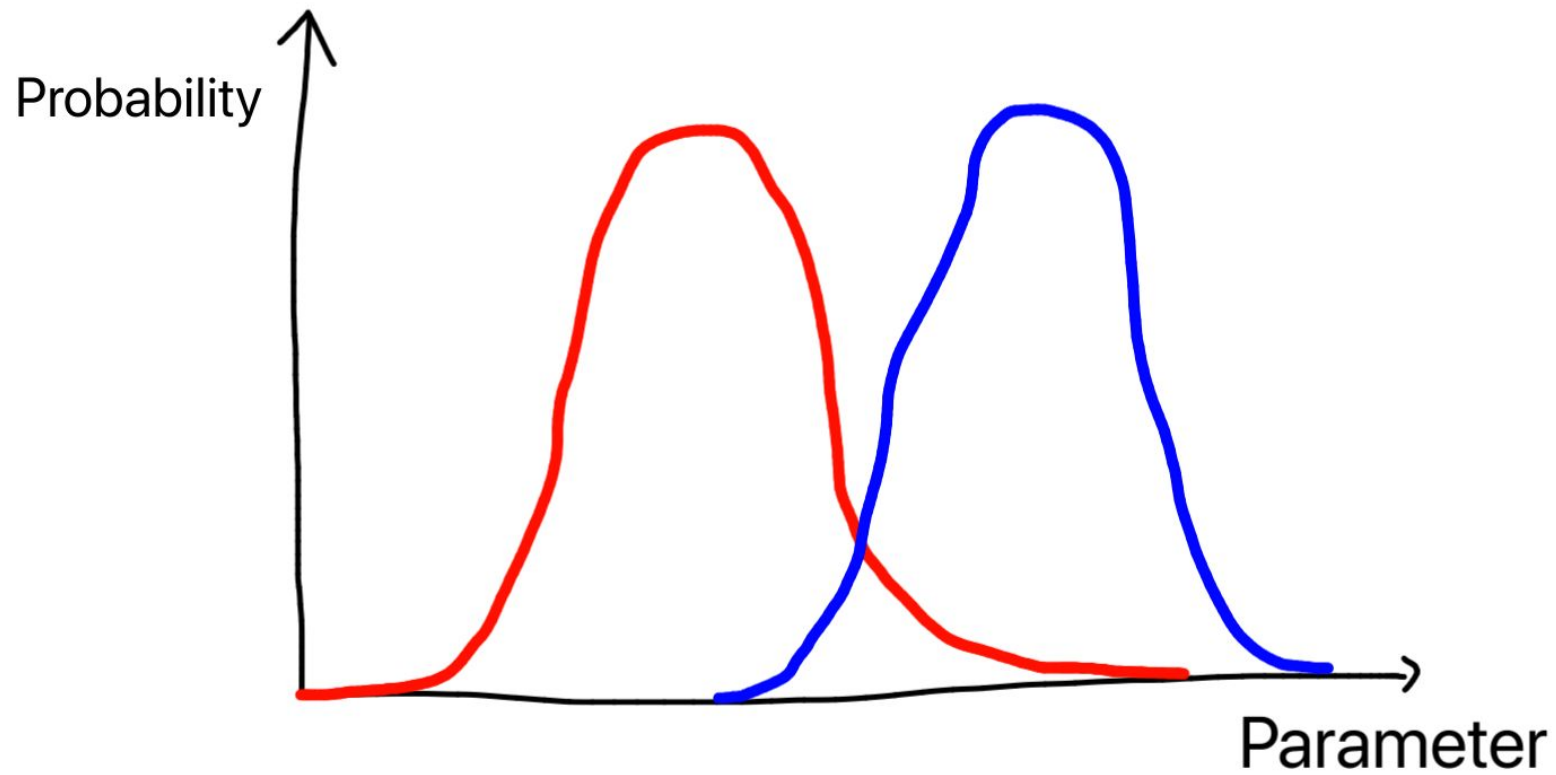


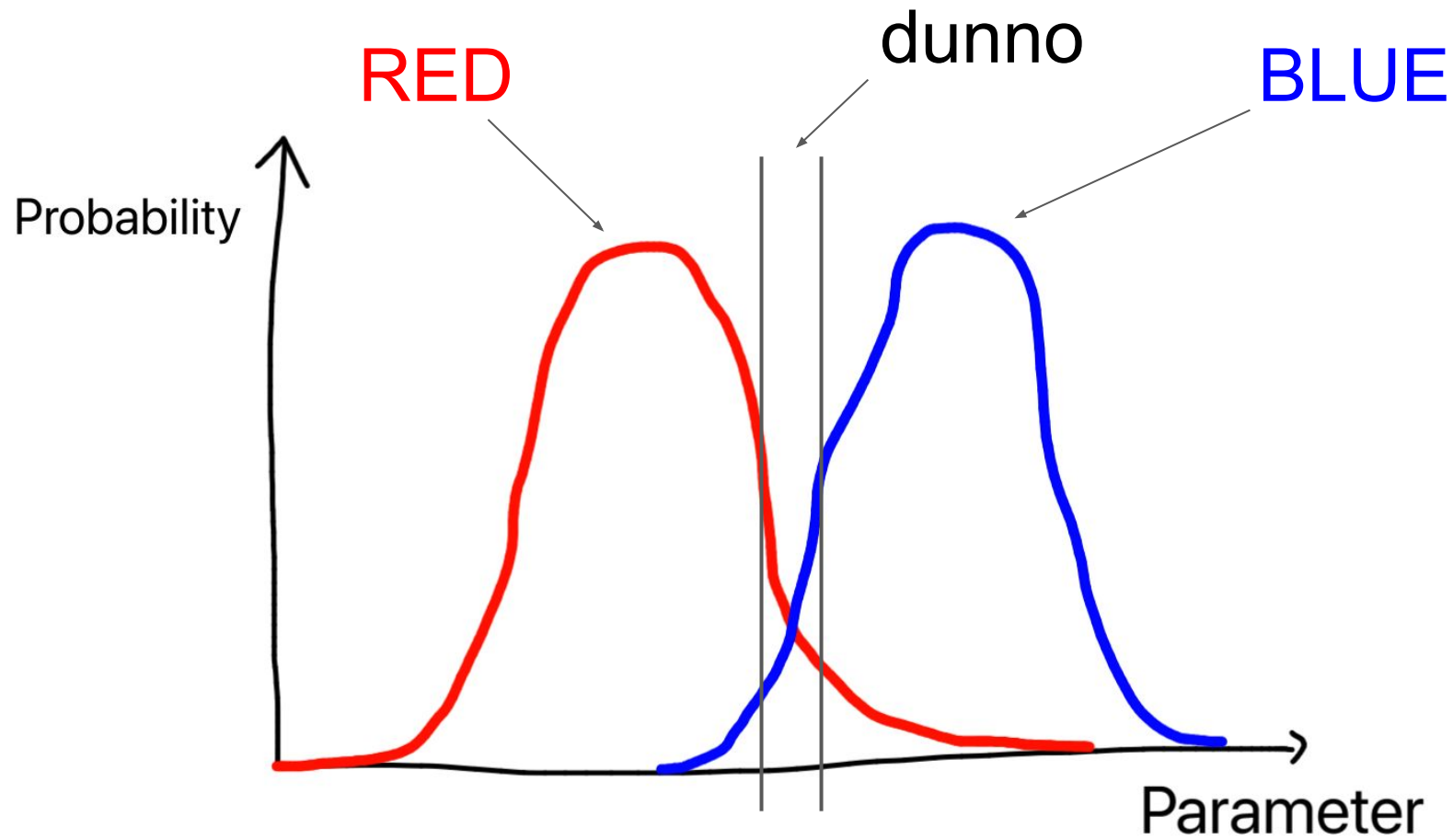
Kosaciec (ang. iris)

KNN Classification on Iris dataset ( $k = 15$ )

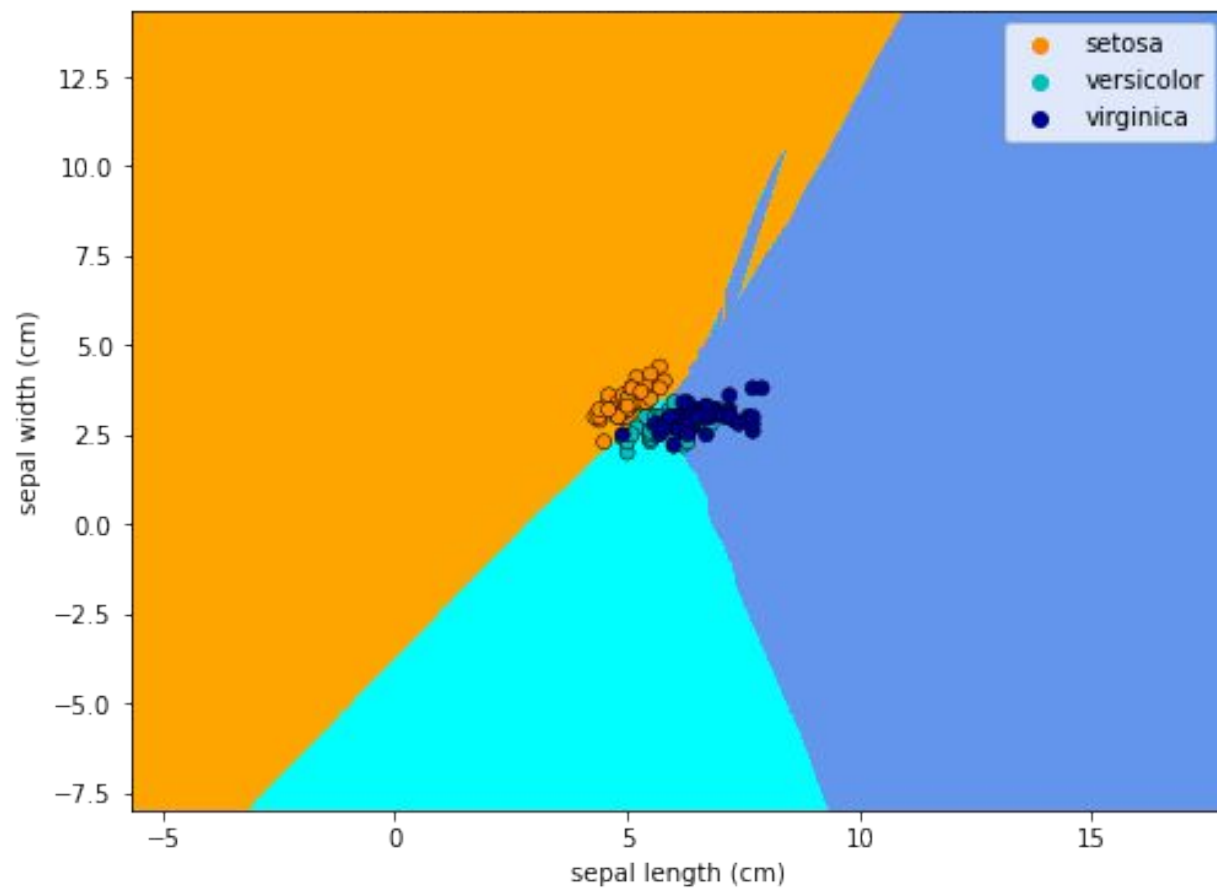




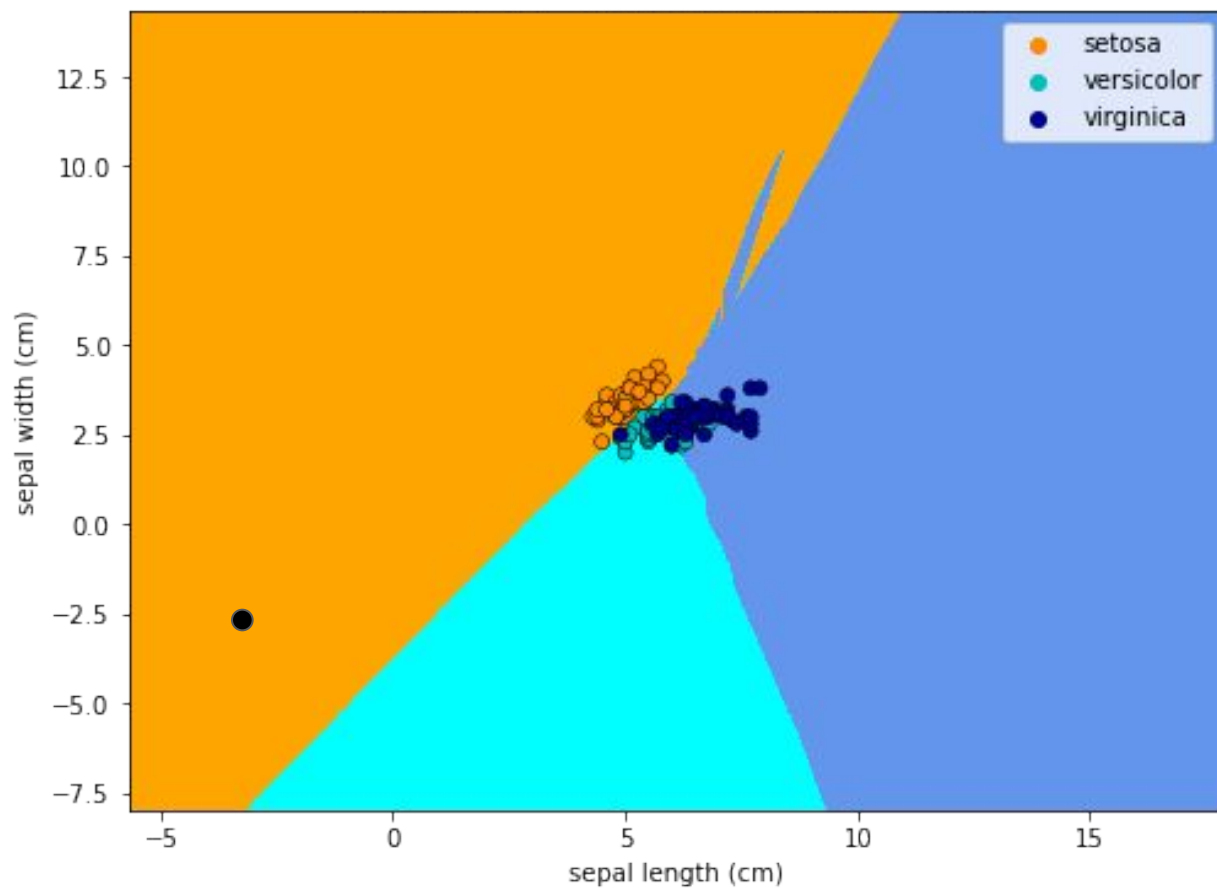




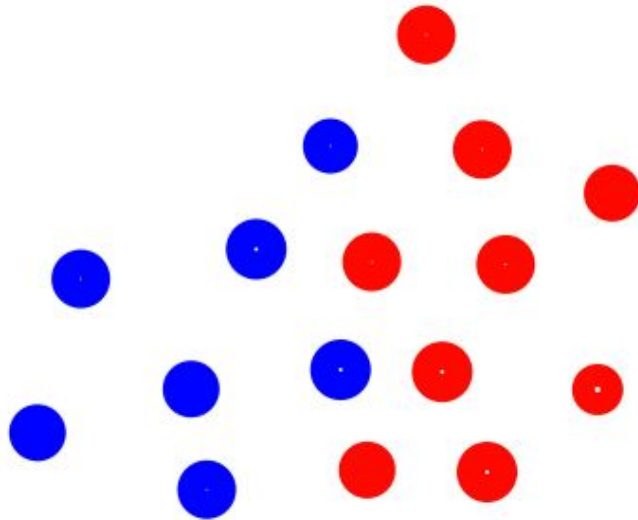
KNN Classification on Iris dataset ( $k = 15$ )



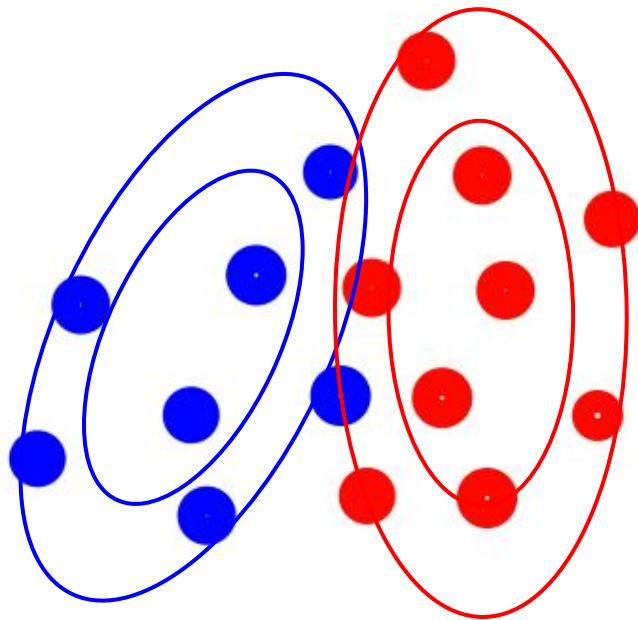
KNN Classification on Iris dataset ( $k = 15$ )



Clever idea

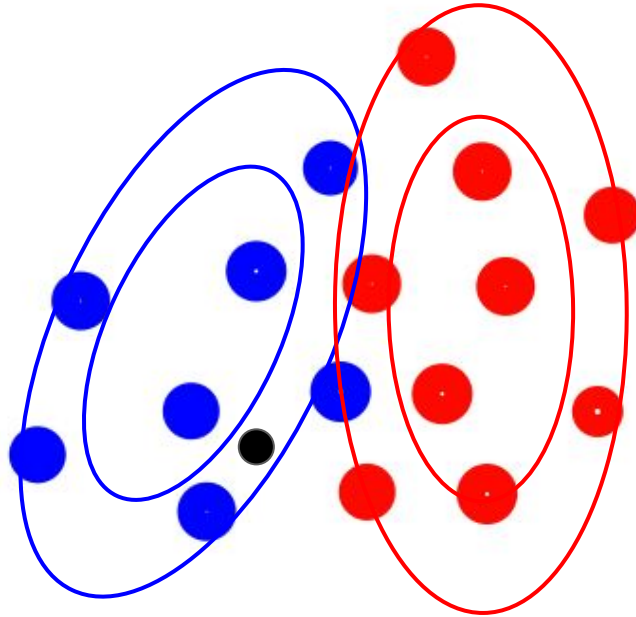


Clever idea - fit the Gaussian function

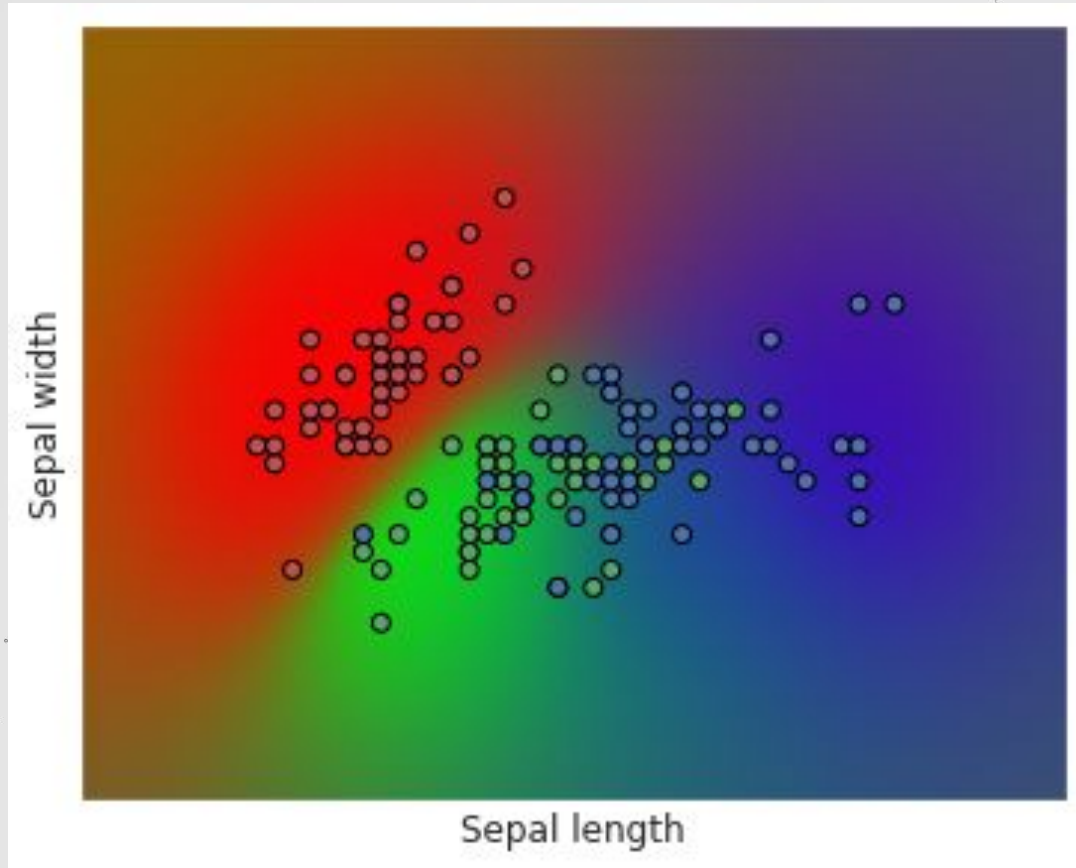




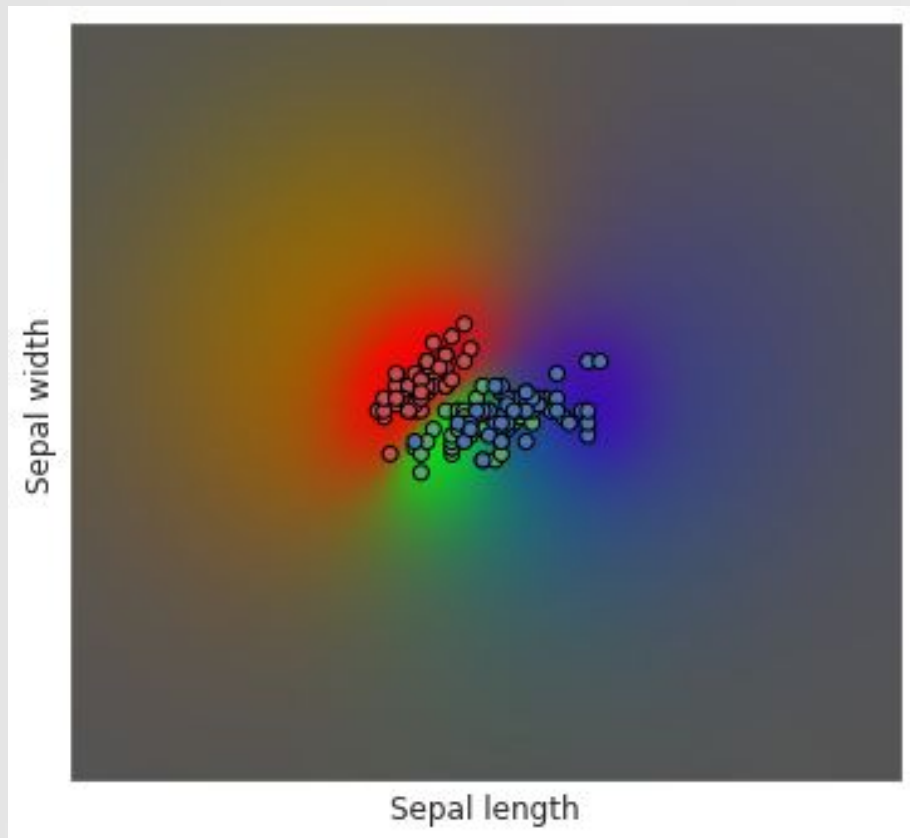
Clever idea - fit the Gaussian function



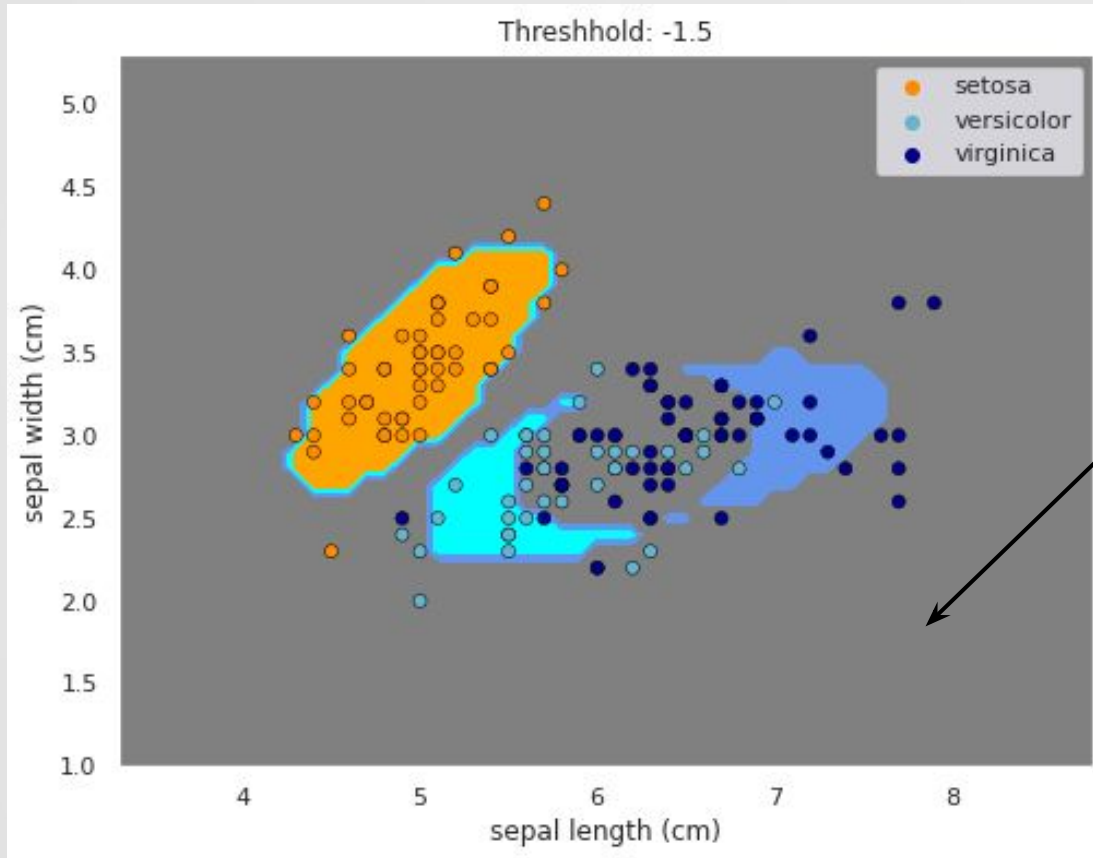
# Fit in practice



# And zoom out

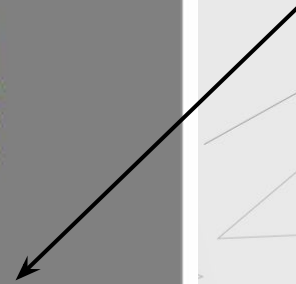


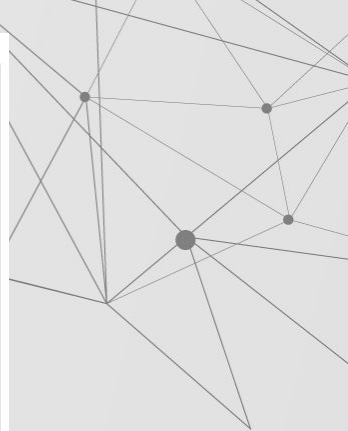
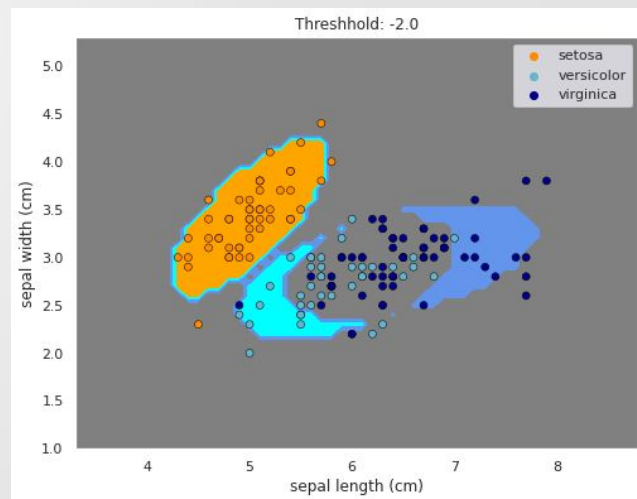
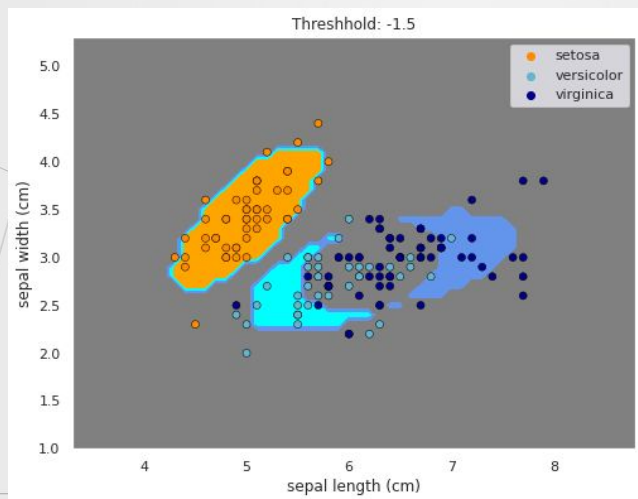
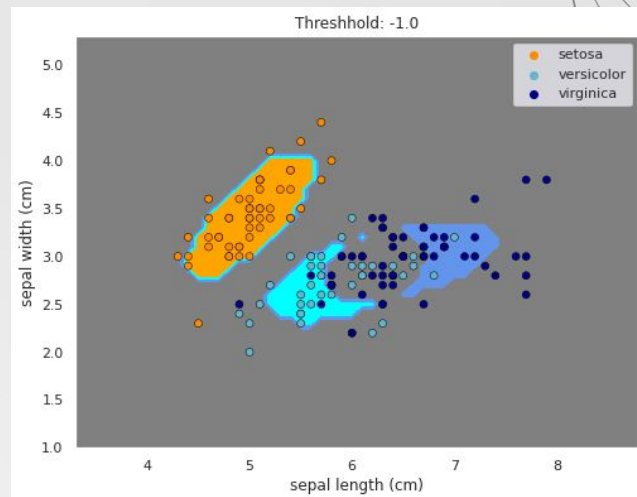
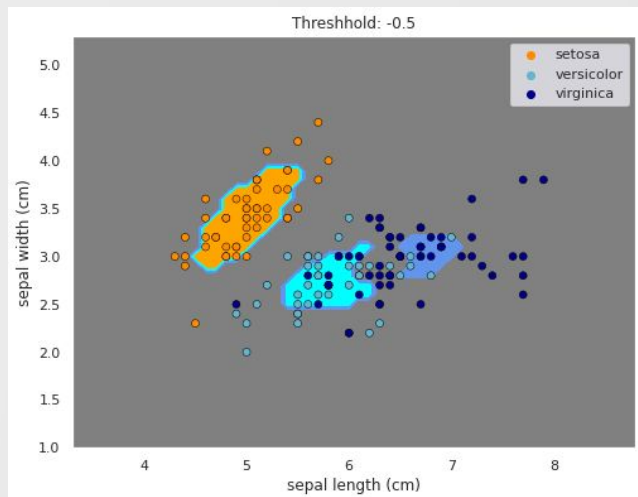
# Let's add an threshold

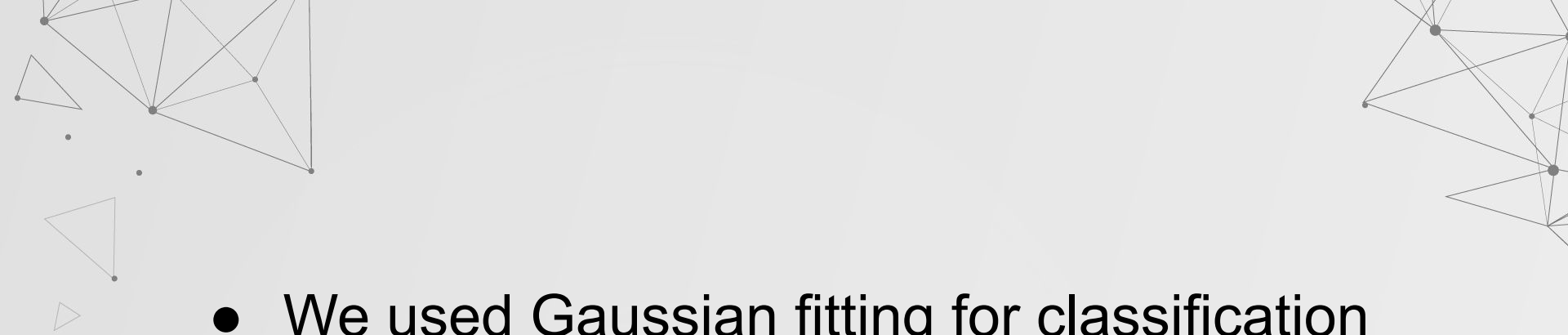


grey  
=

“prefer not to  
answer”





- 
- The top-left and top-right corners of the slide feature decorative geometric patterns. These consist of thin, light-gray lines connecting small black dots to form various triangular and polygonal shapes. Some dots are isolated, while others are part of larger, more complex networks of lines.
- We used Gaussian fitting for classification
  - We are more aware where how certain our model is about its predictions
  - **We also got outlier detection for free**



# 3. Decision trees, but better

---



# A quick reminder on decision trees

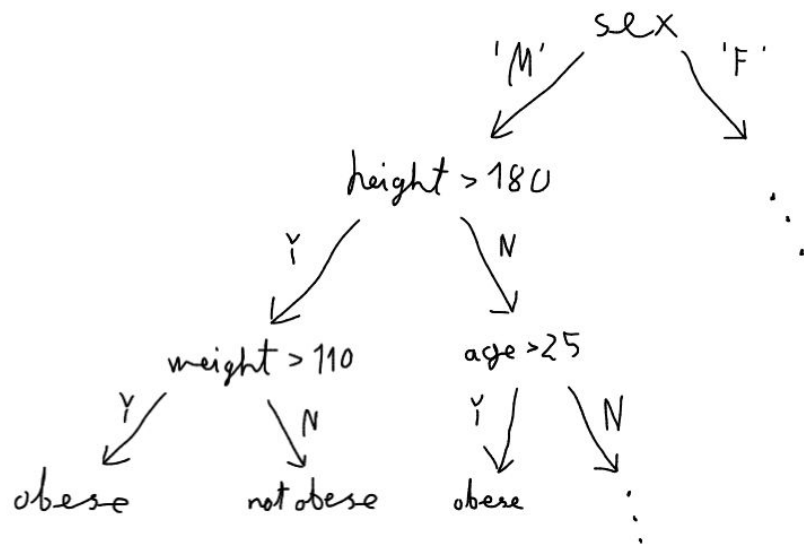


[age, sex, height, weight]  $\longrightarrow$  obese?

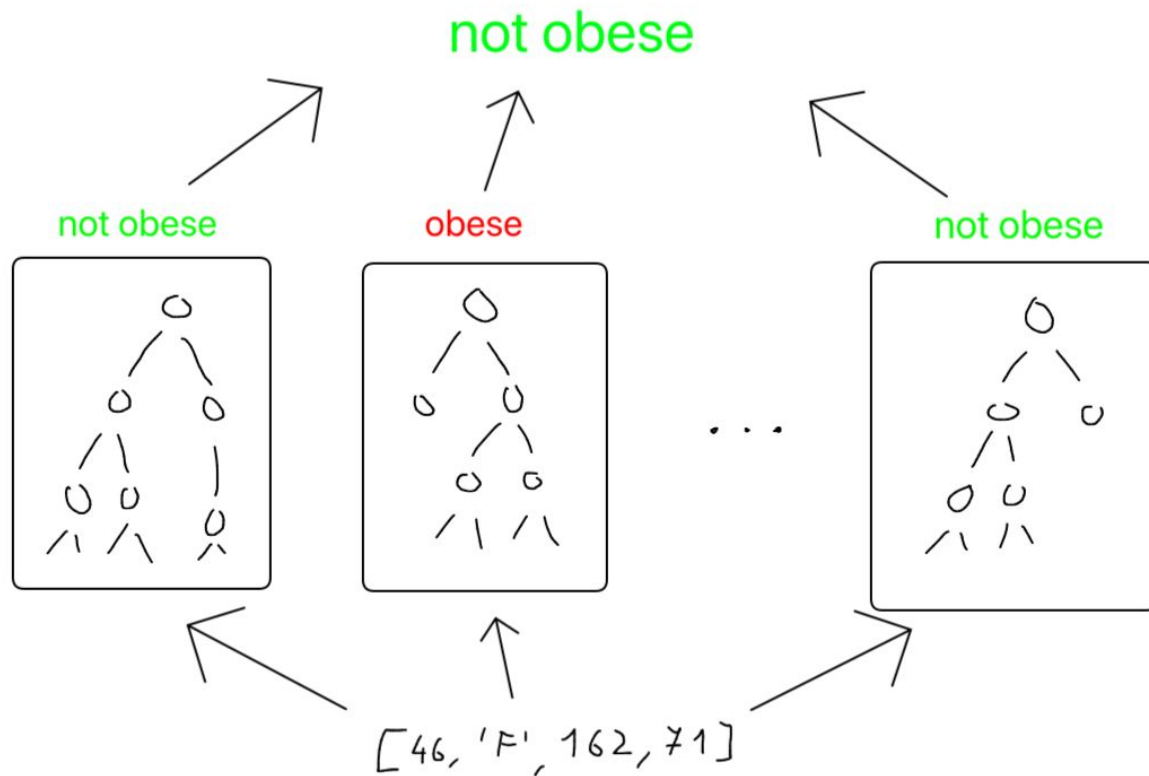
[34, 'M', 187, 82]  $\longrightarrow$  NO

[17, 'F', 150, 92]  $\longrightarrow$  YES

$\vdots$



# A random forest is many decision trees

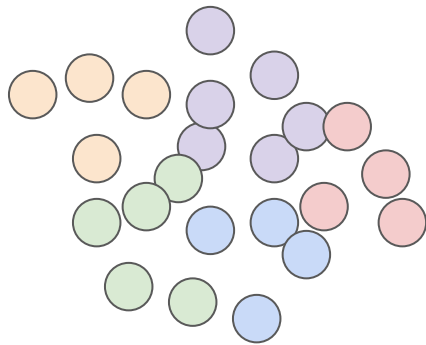


Goal:

Similarly to KNN, we want to have a random forest that tells us when it's not sure

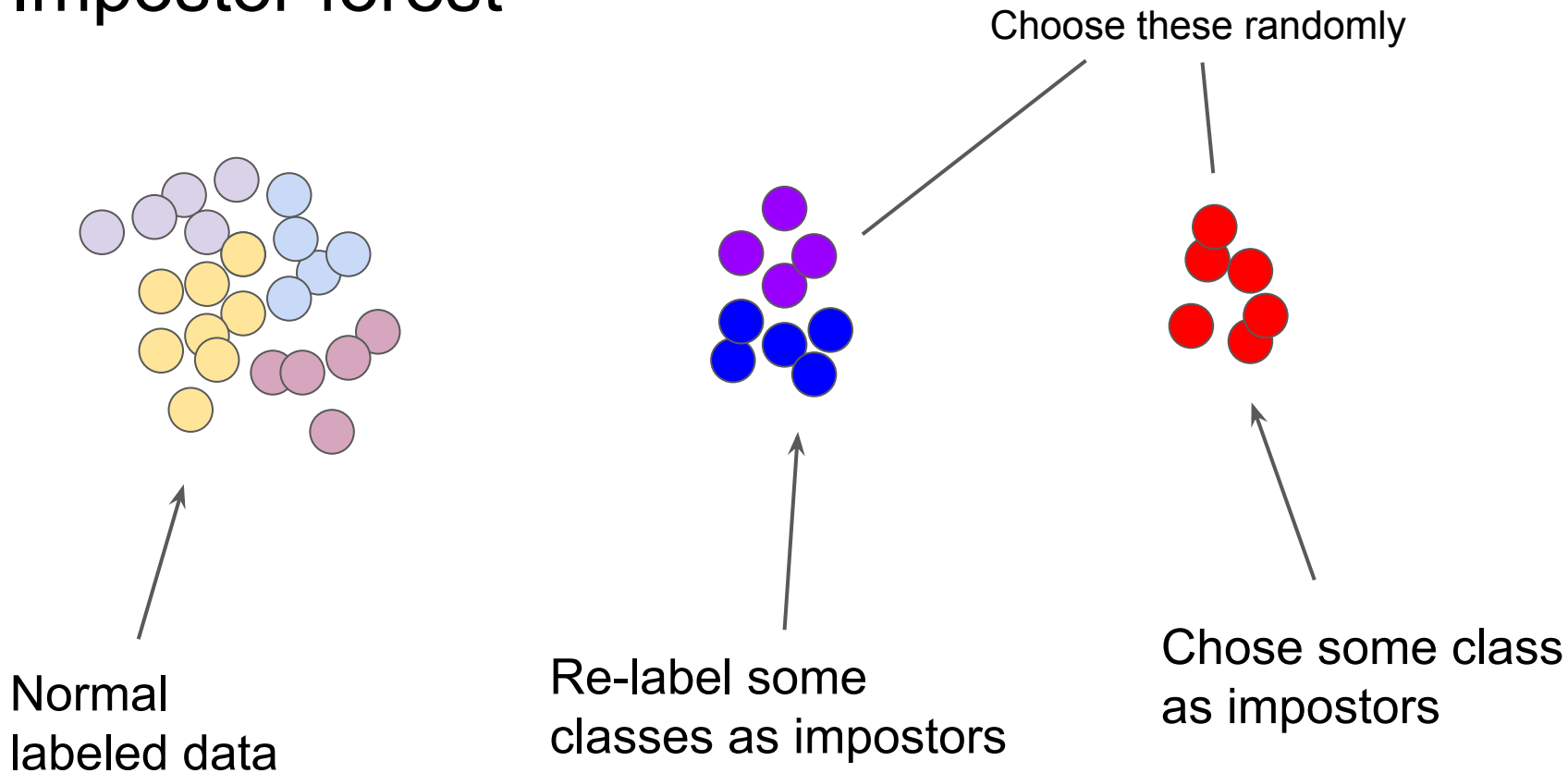


# Impostor forest



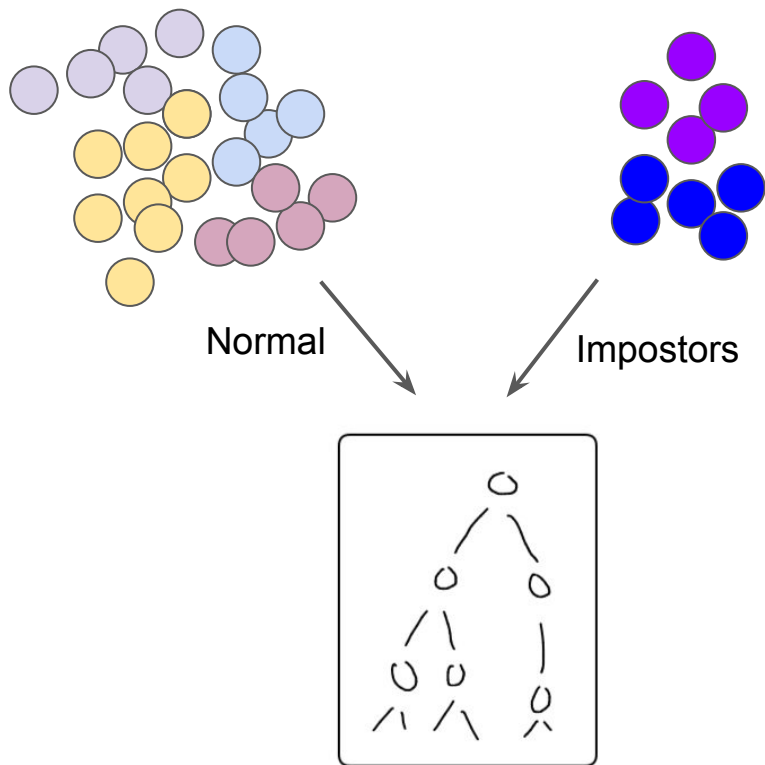
Labeled data

# Impostor forest

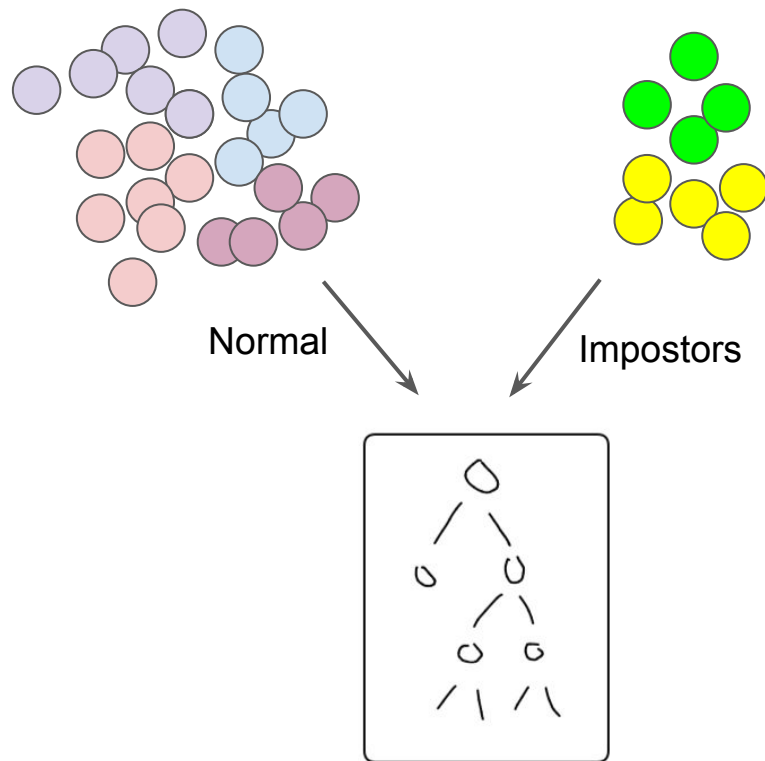




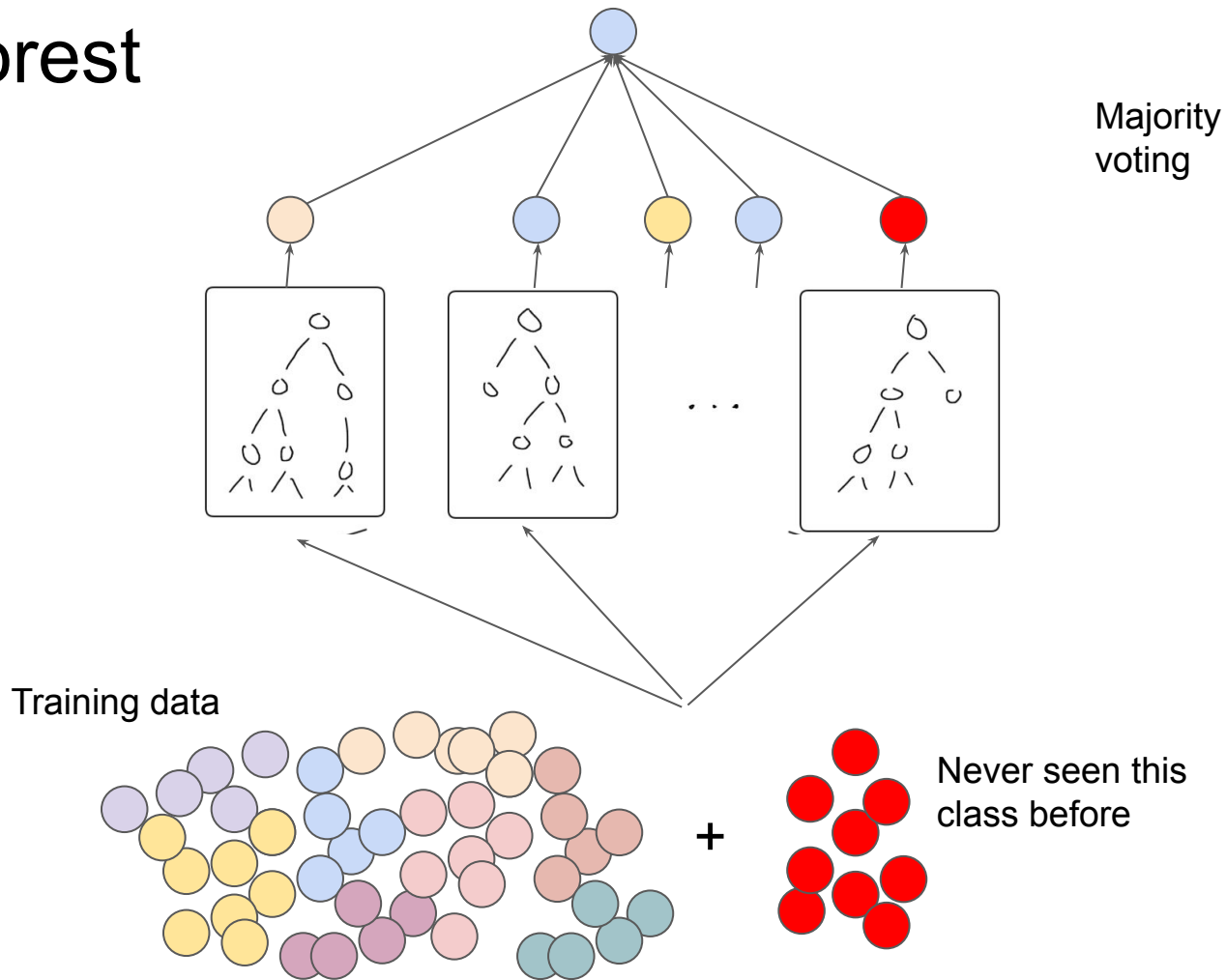
# Impostor forest



...



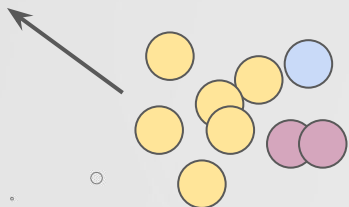
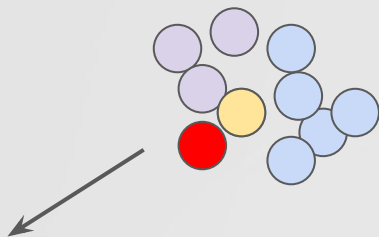
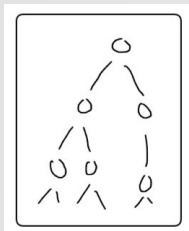
# Impostor forest





# Outlier detection

What about outliers in training data?

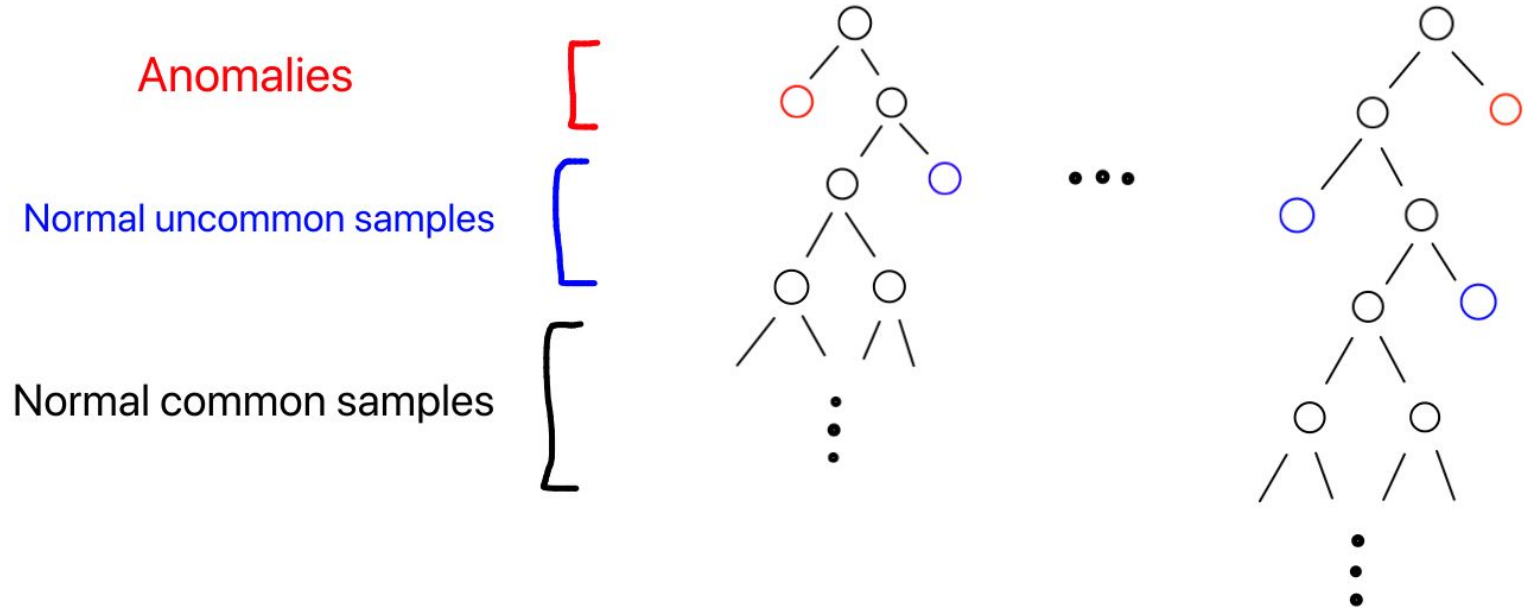


training

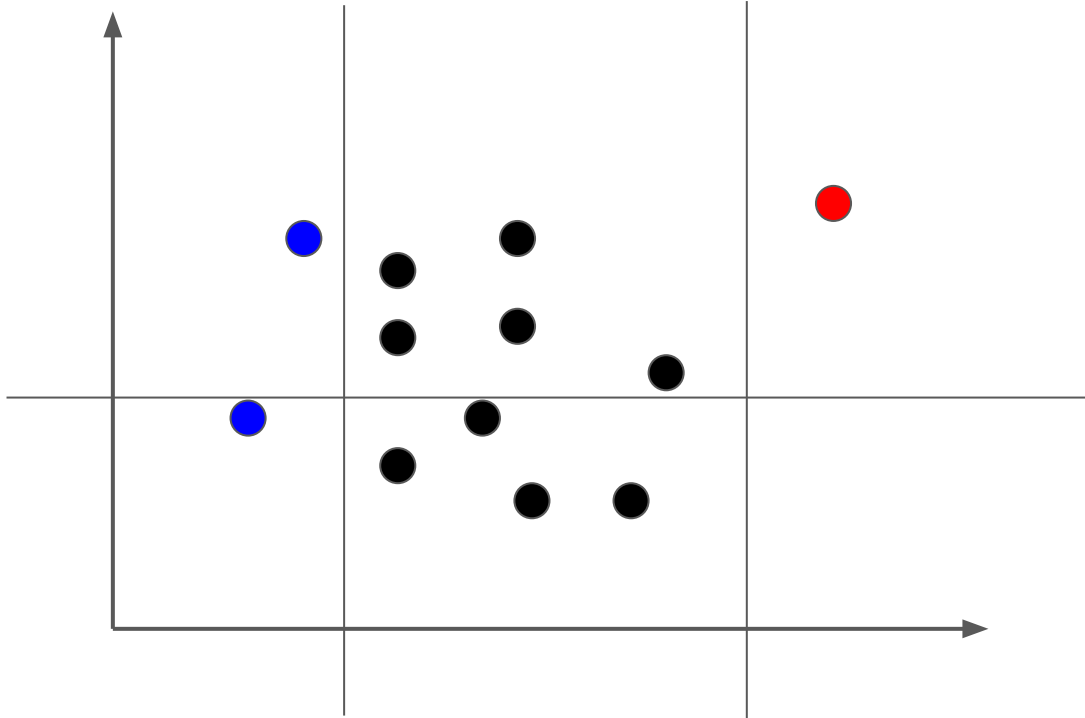
running



# Isolation forest



# Outliers are found in few tree cuts



# Resources:

- Vincent Warmerdam PyData talks:
  - [https://www.youtube.com/watch?v=68ABAU\\_V8qI&t=370s&ab\\_channel=PyData](https://www.youtube.com/watch?v=68ABAU_V8qI&t=370s&ab_channel=PyData)
  - [https://www.youtube.com/watch?v=Z8MEFI7ZJIA&ab\\_channel=PyData](https://www.youtube.com/watch?v=Z8MEFI7ZJIA&ab_channel=PyData)
- Isolation forest paper:
  - <https://ieeexplore.ieee.org/abstract/document/4781136>



**THANKS!**

