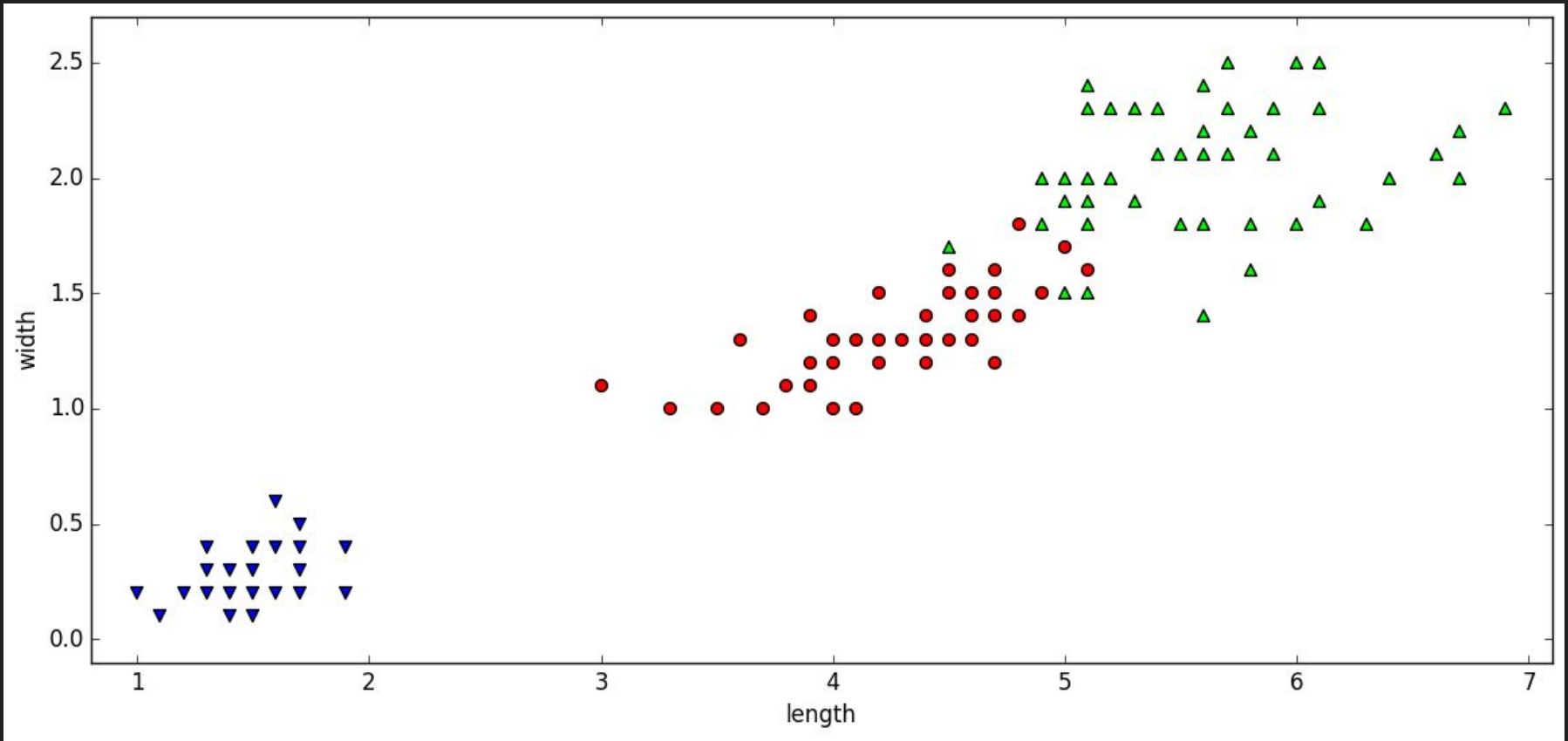
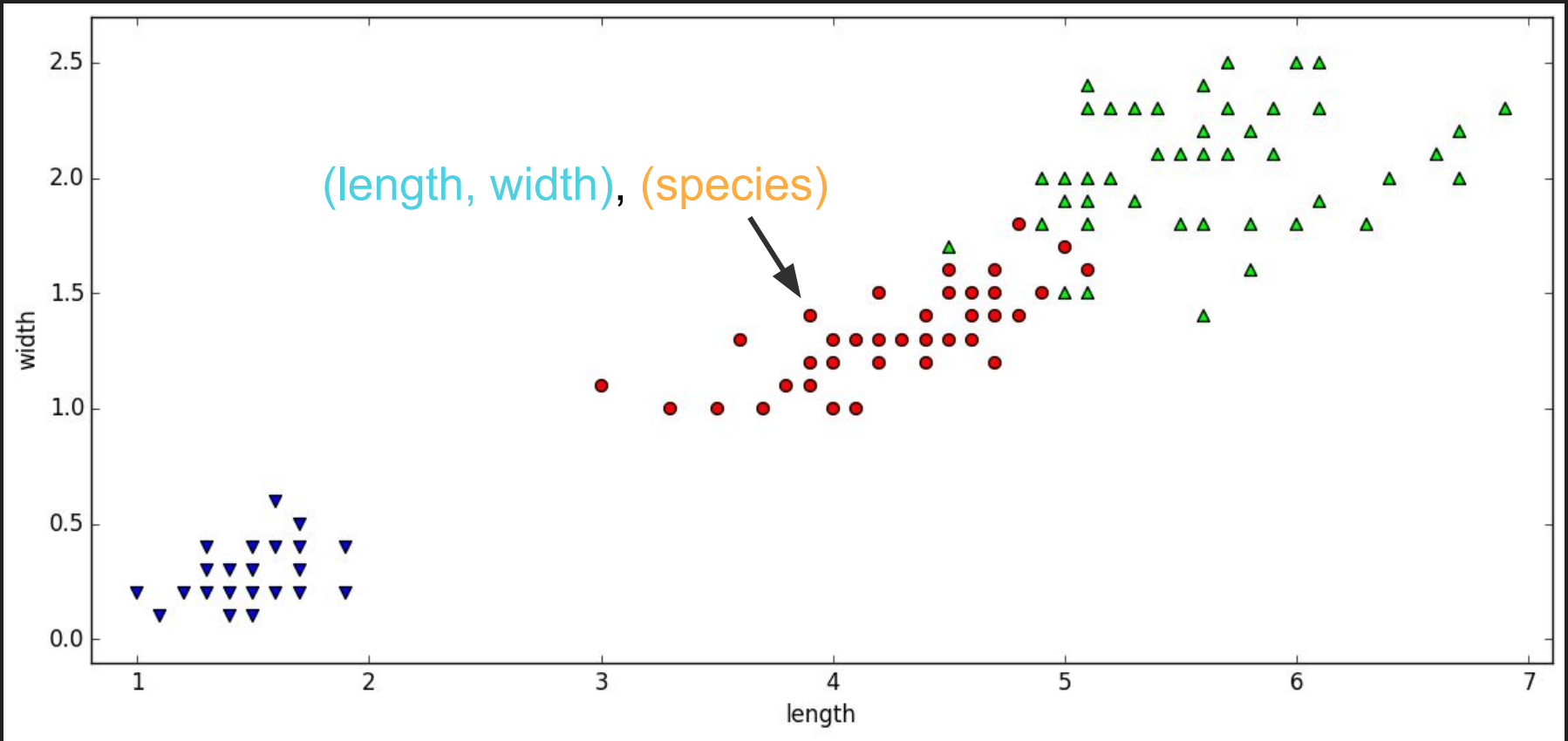
Nearest Neighbour

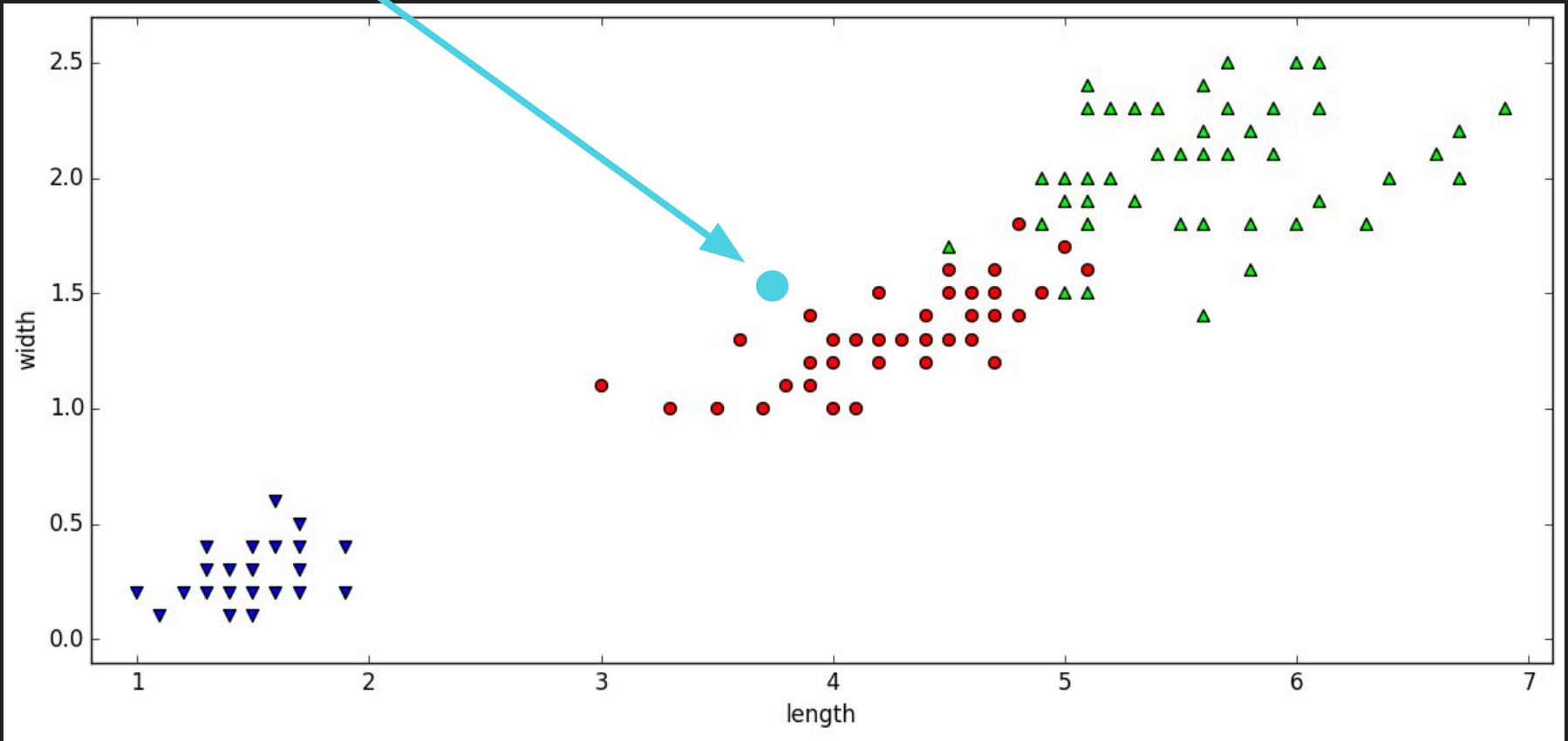
Here is our 2D dataset, with 3 different classes



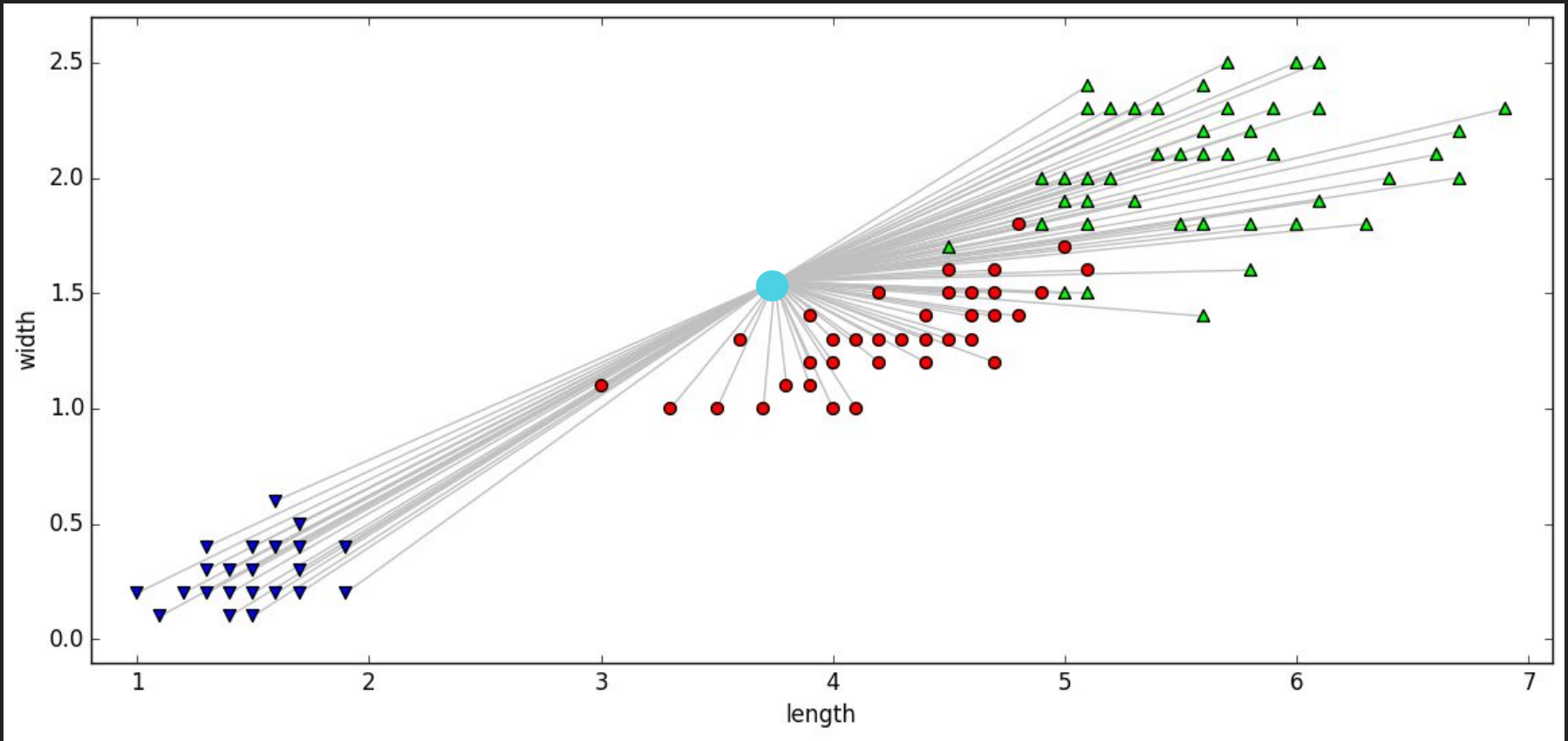
Each datapoint has some features and a class label



Given a new datapoint, how can we determine its class?

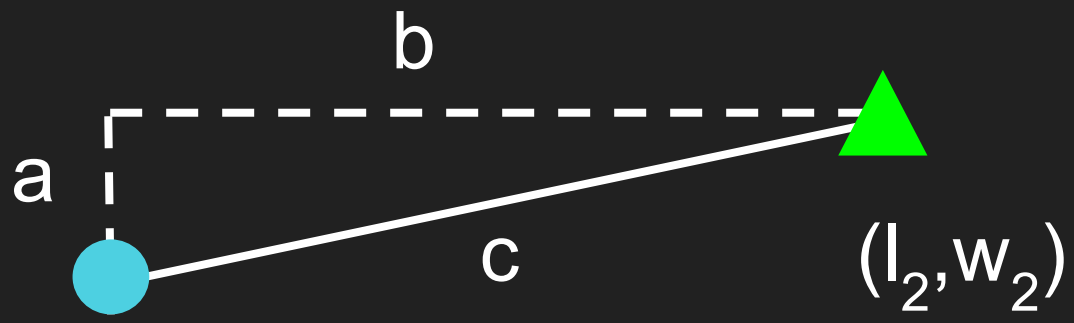


Find its “Nearest Neighbour” in the feature space



Computing “similarity” between two points

Euclidean distance in 2D

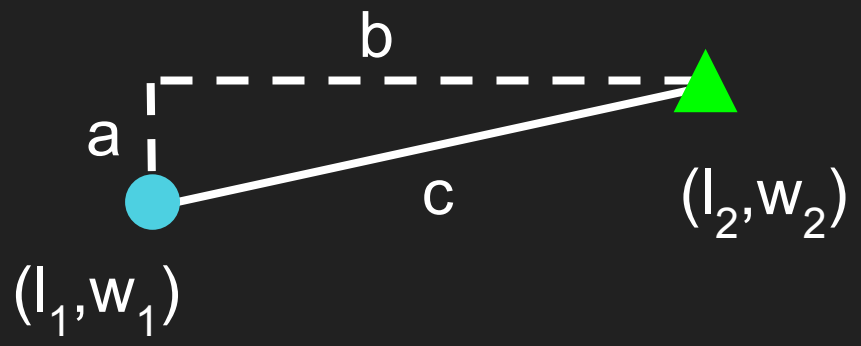


(l_1,w_1)

$c^2 = a^2 + b^2$ i.e Pythagoras' theorem

Euclidean distance in 2D

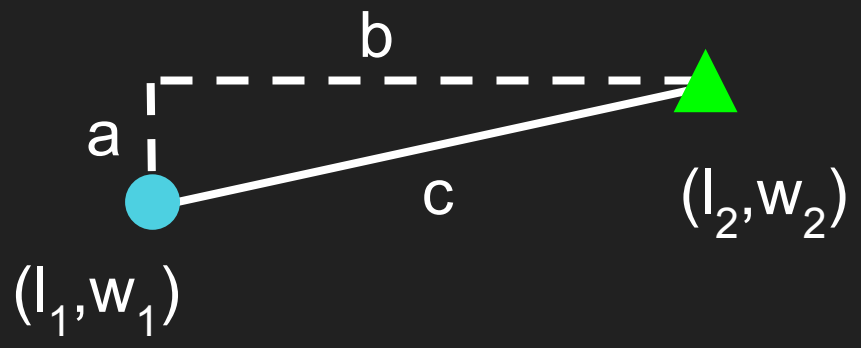
$$c^2 = a^2 + b^2$$



Euclidean distance in 2D

$$c^2 = a^2 + b^2$$

$$c = \sqrt{a^2 + b^2}$$

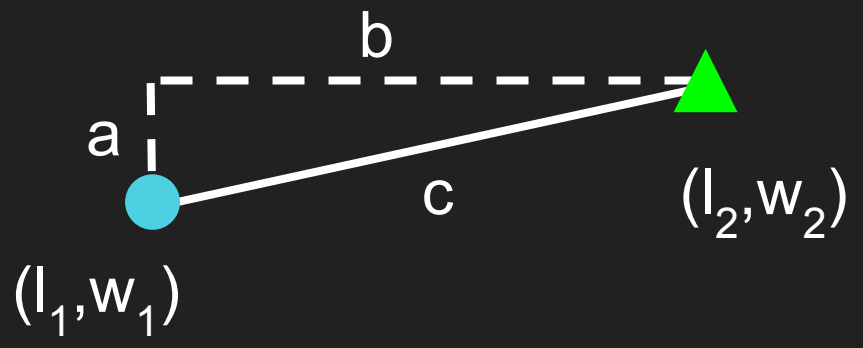


Euclidean distance in 2D

$$c^2 = a^2 + b^2$$

$$c = \sqrt{a^2 + b^2}$$

$$c = \sqrt{(l_1 - l_2)^2 + (w_1 - w_2)^2}$$

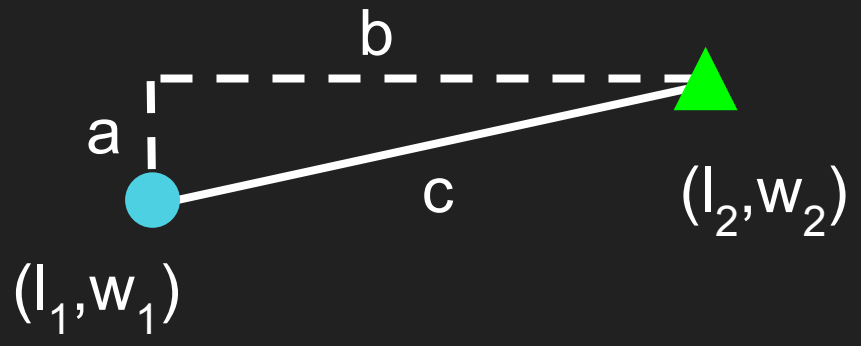


Euclidean distance in 2D

$$c^2 = a^2 + b^2$$

$$c = \sqrt{a^2 + b^2}$$

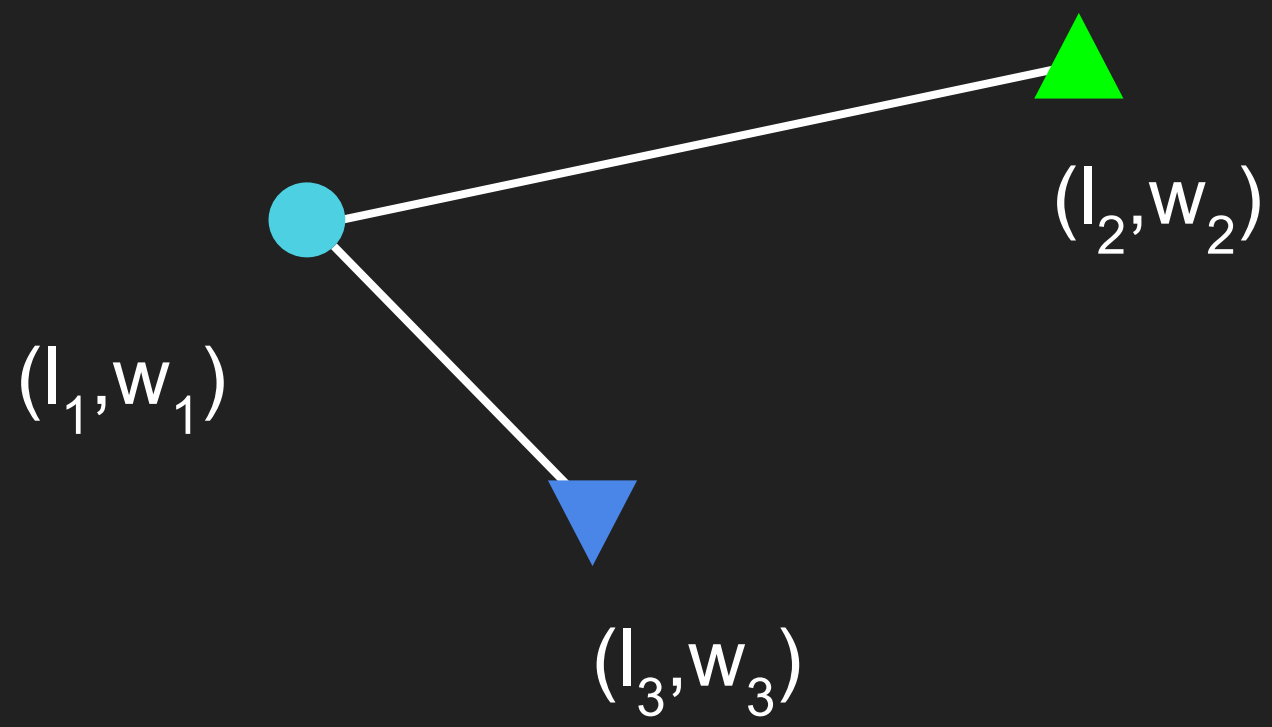
$$c = \sqrt{(l_1-l_2)^2 + (w_1-w_2)^2}$$



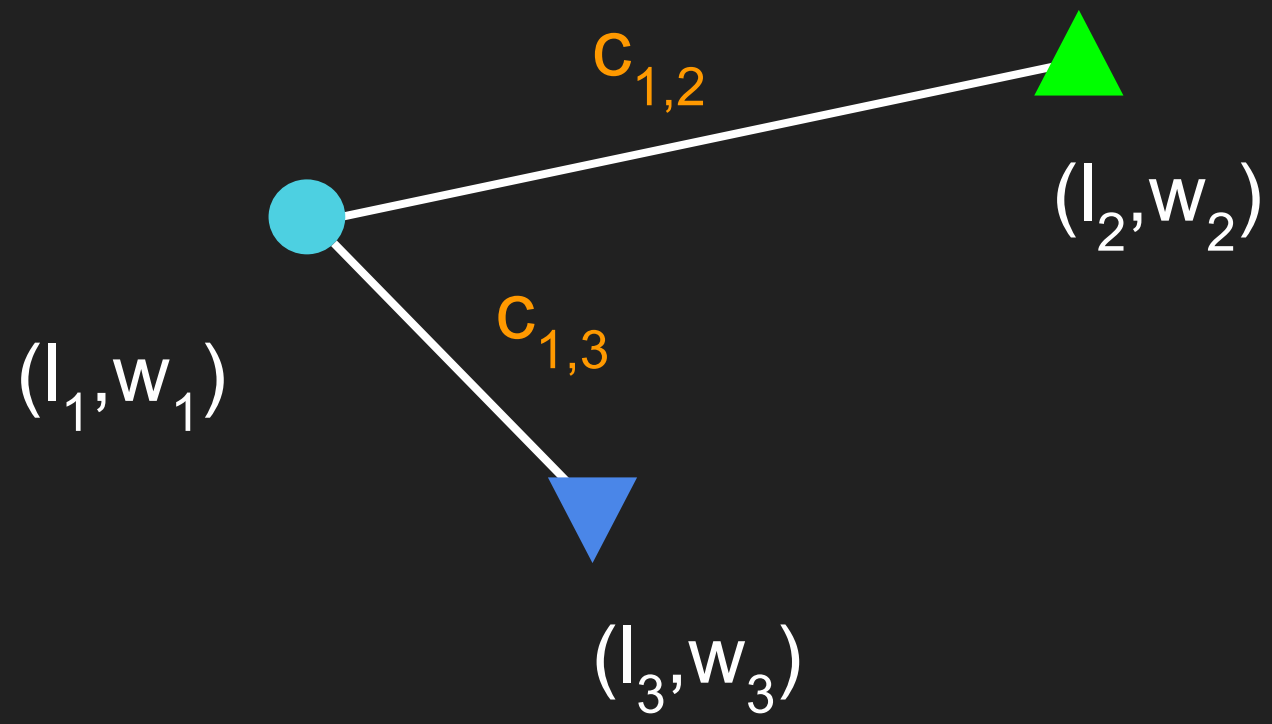
In R:

```
p1 <- c(0, 0)
p2 <- c(1, 1)
distance <- sqrt(sum((p1-p2)^2))
```

Choose the closest point

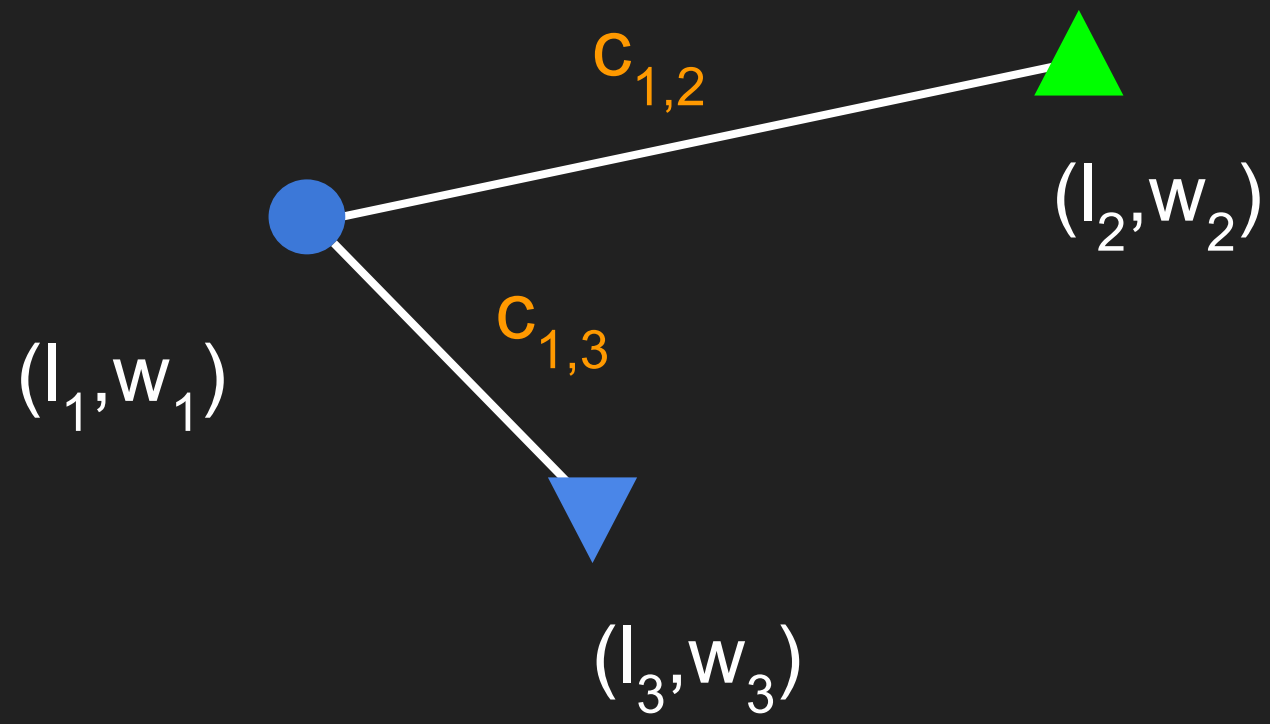


Choose the closest point

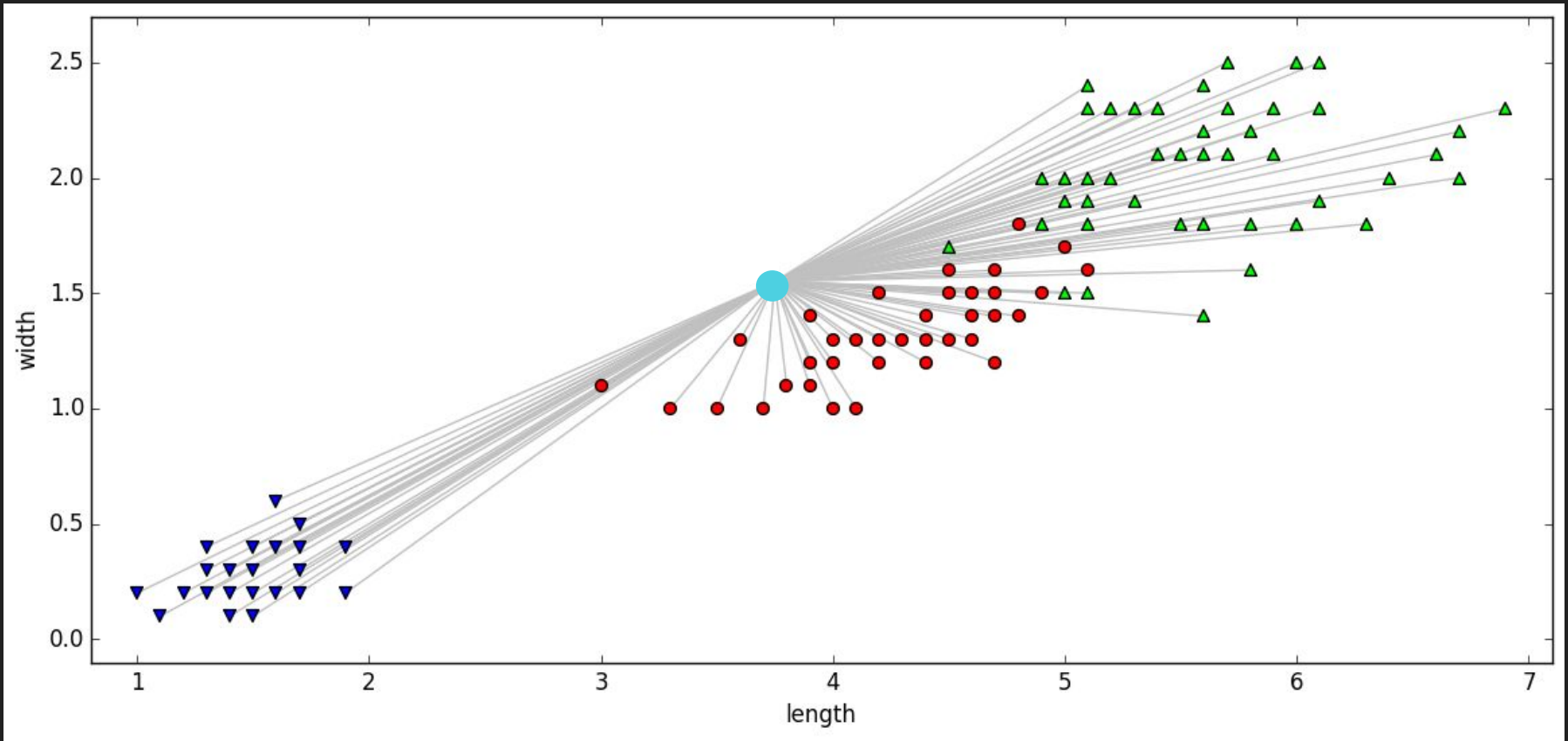


Choose the closest point

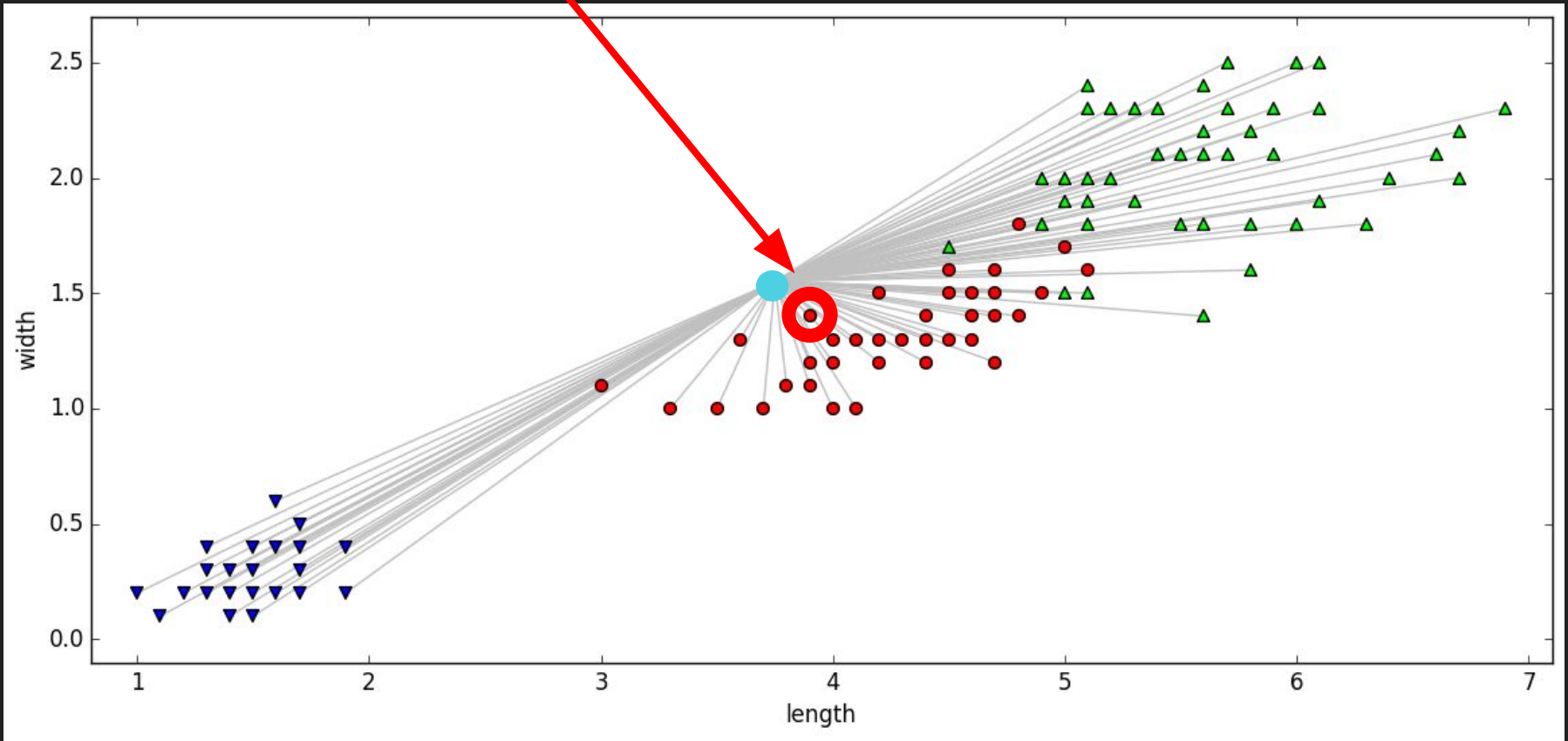
$c_{1,3} < c_{1,2}$



Find its “Nearest Neighbour” in the feature space



“Nearest Neighbour” in the feature space



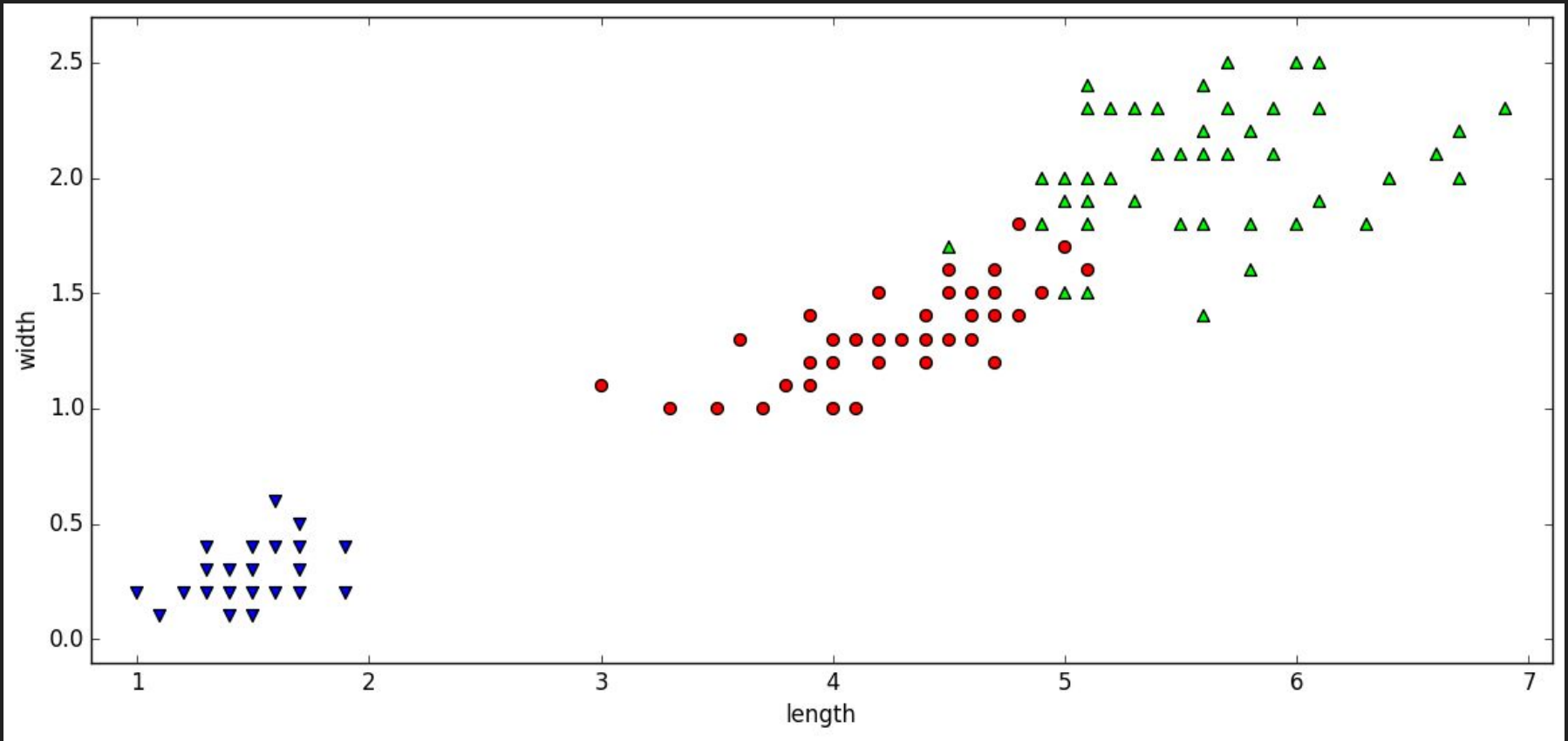
Nearest Neighbour Algorithm

Given a test point **x**

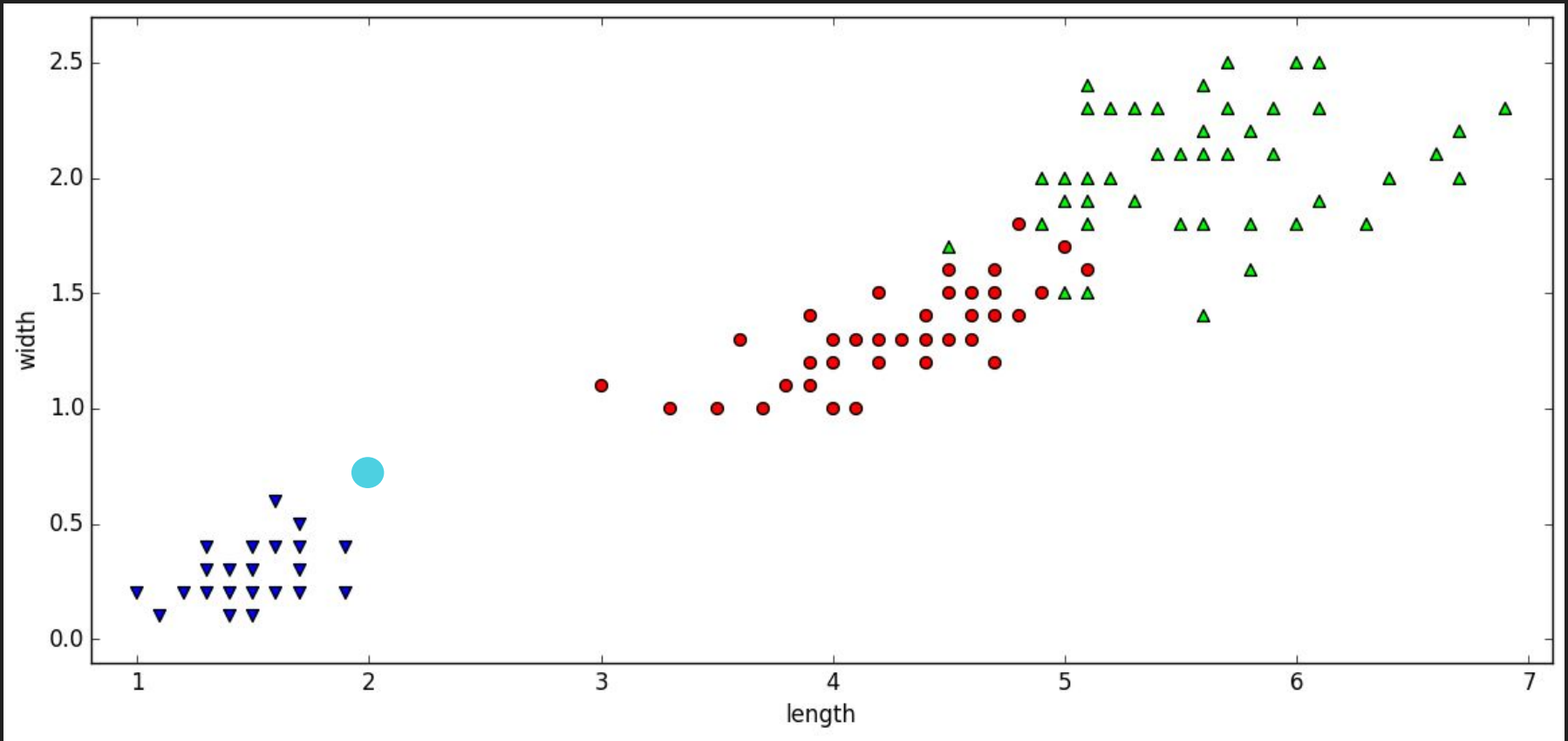
Compute the distance between **x** and every other datapoint

The class of **x** is set as the same as the closest datapoint

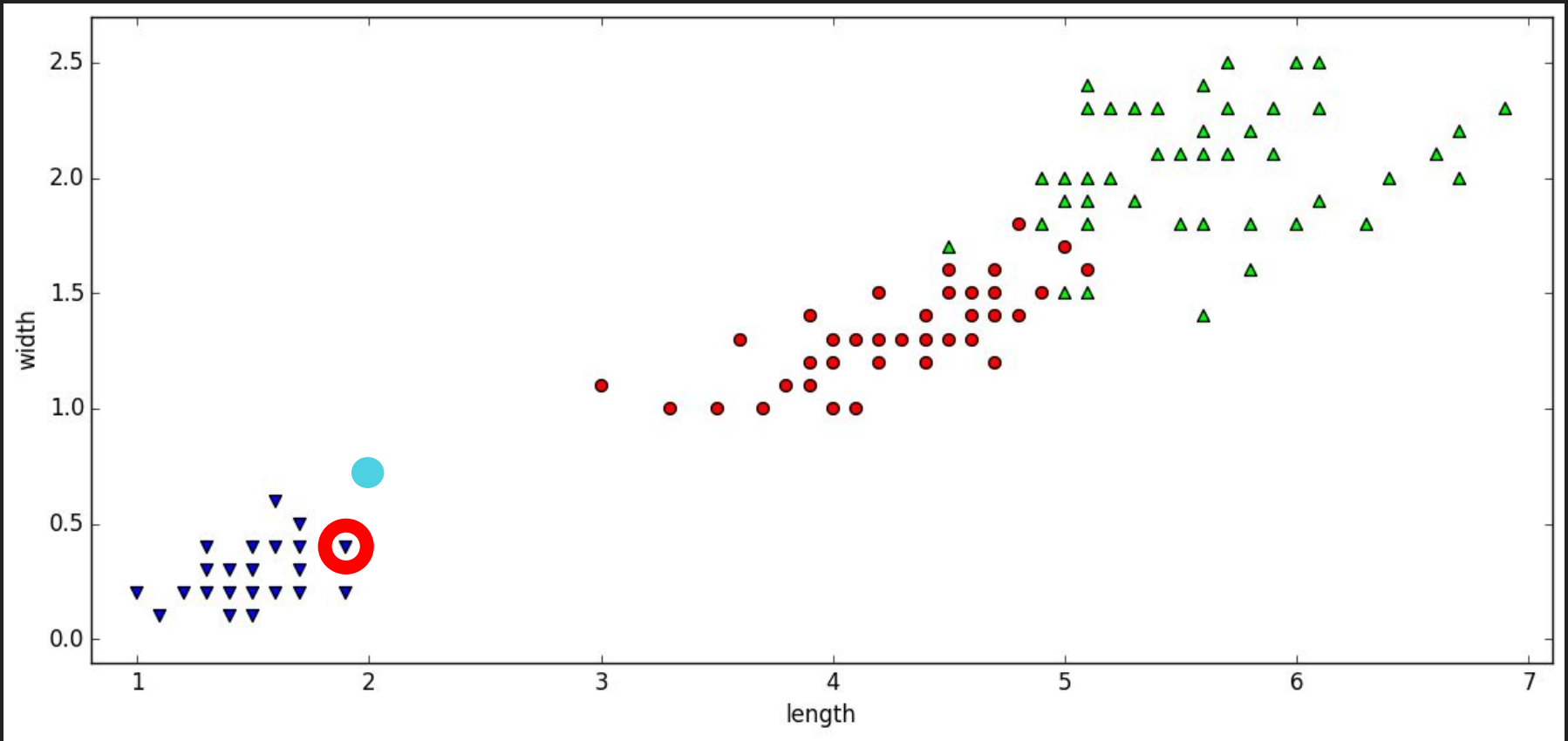
Again our 2D dataset



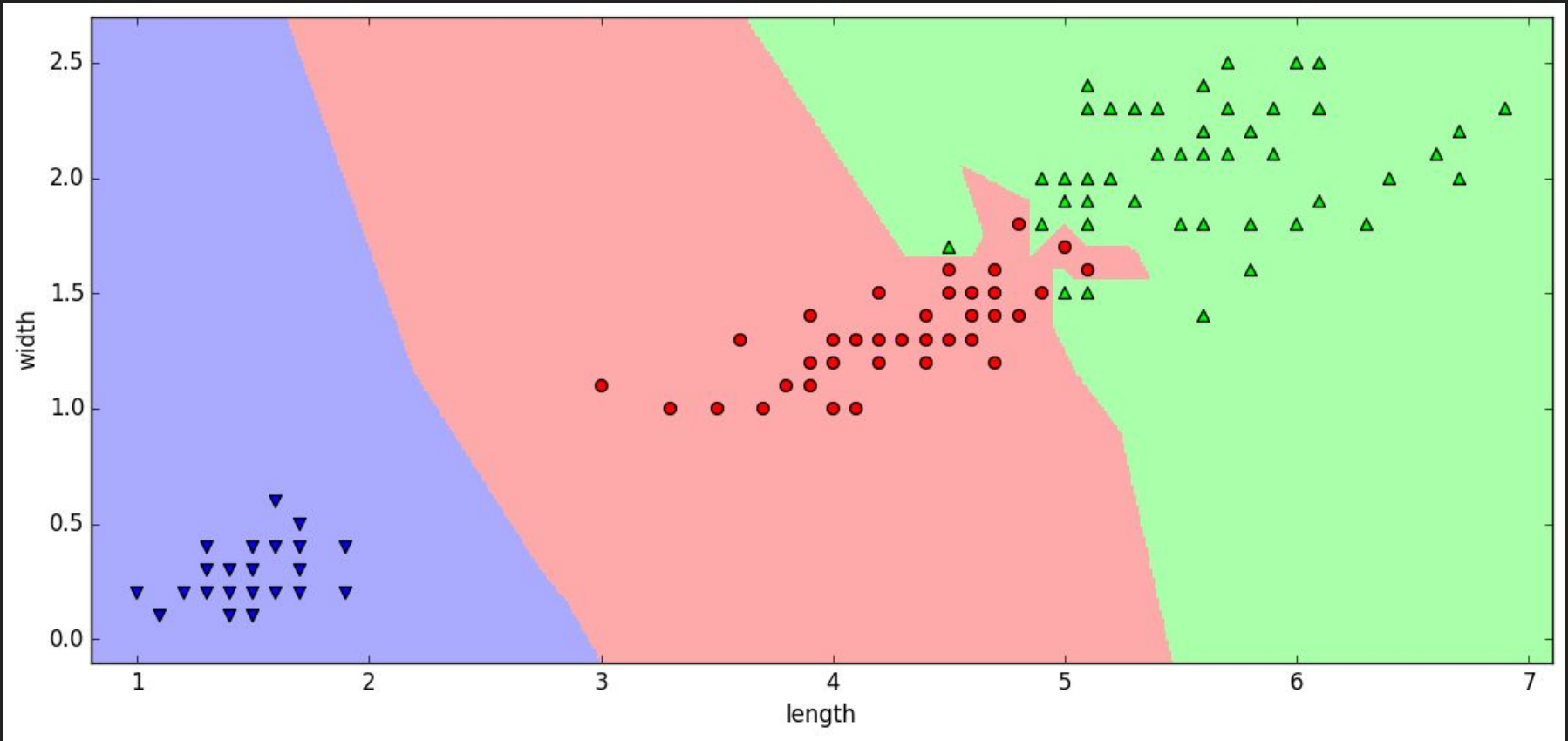
Let’s try a different test point



Here is it’s neighbour

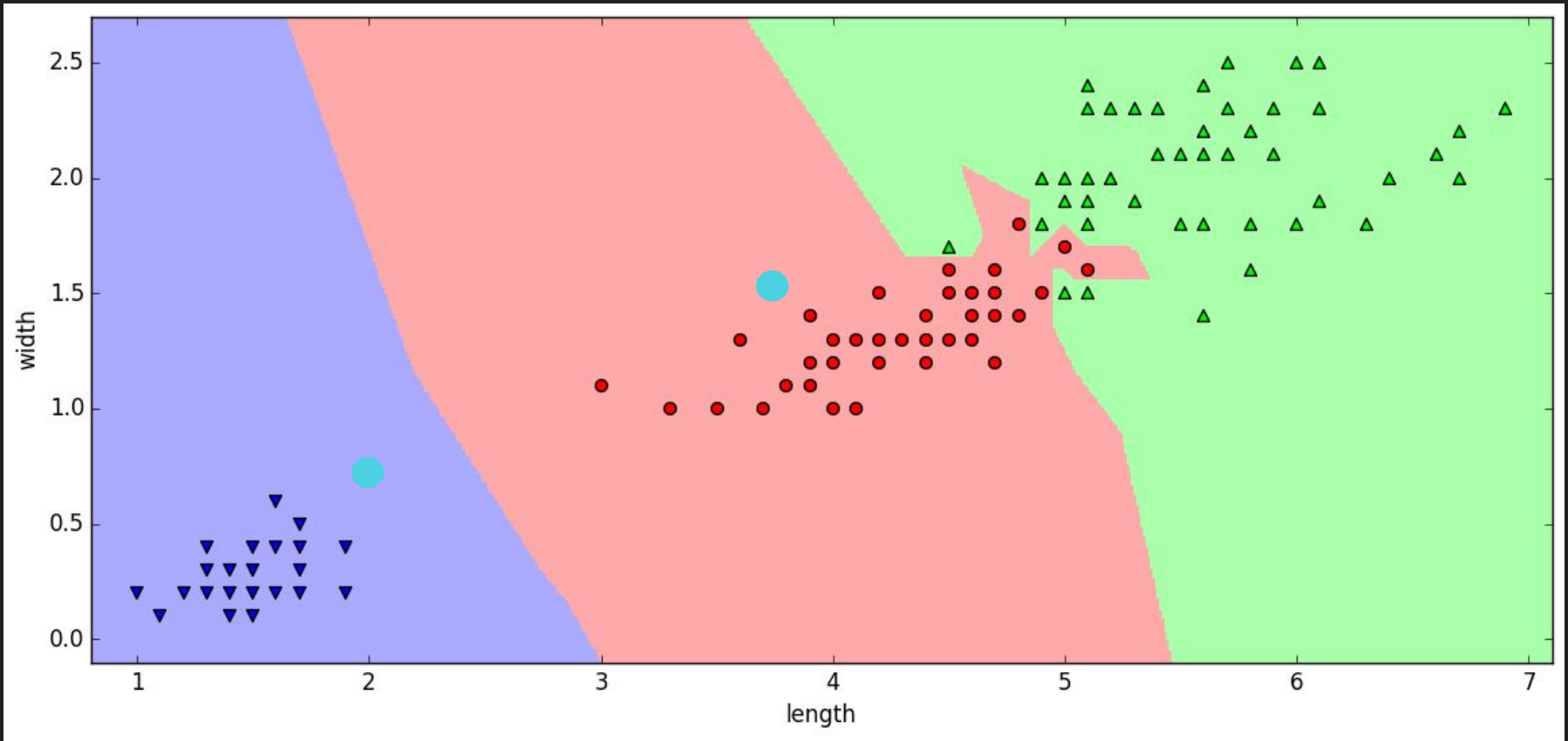


Resulting Nearest Neighbour classification



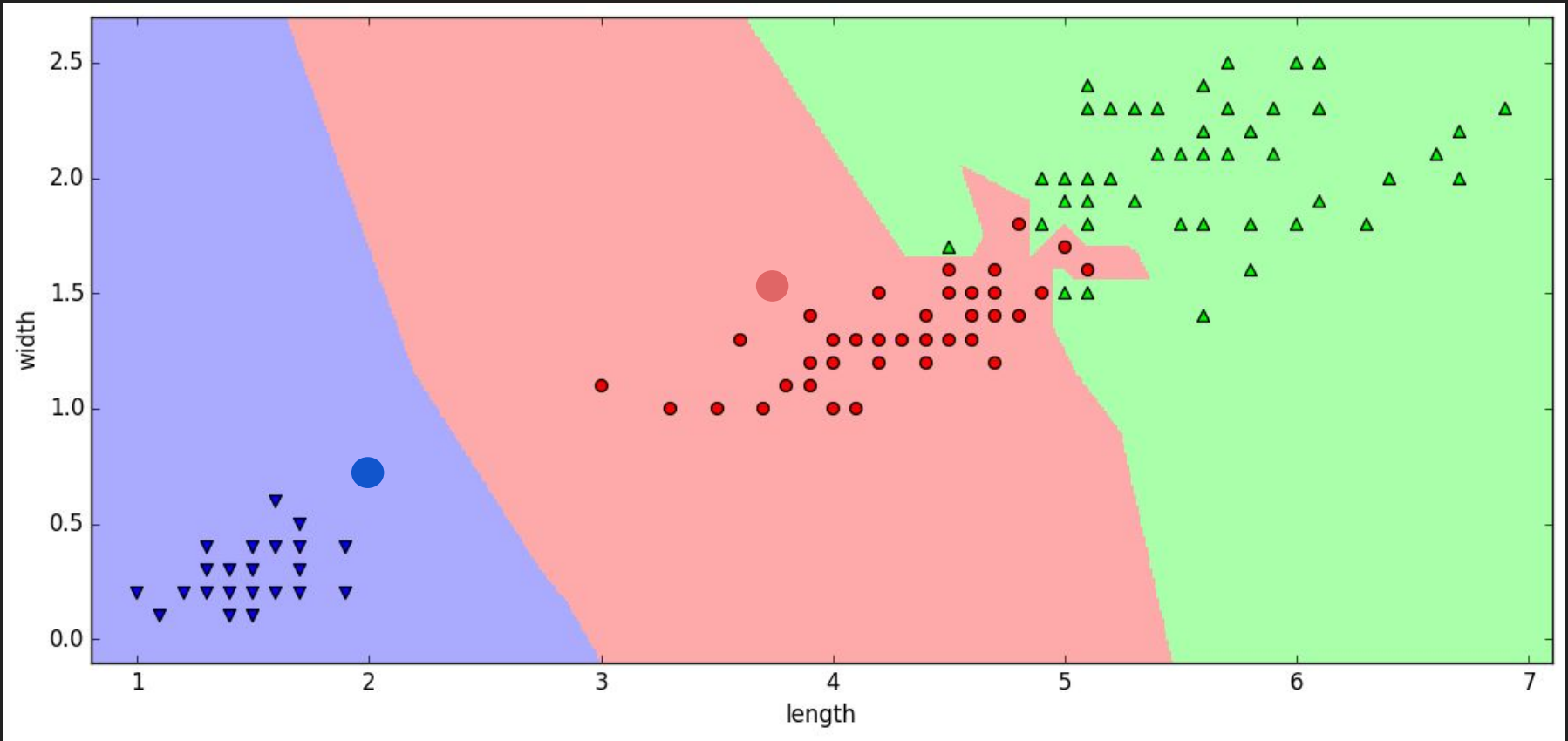
For every point in the space we colour it with the class of the datapoint it is closest to.

Resulting Nearest Neighbour classification



For every point in the space we colour it with the class of the datapoint it is closest to.

Resulting Nearest Neighbour classification



For every point in the space we colour it with the class of the datapoint it is closest to.

Practical example

2_nearest_neighbour.R