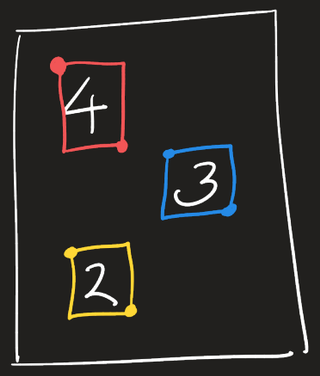
Handwritten Digit object Detection



$t_{4\,top}$
$t_{4\,bottom}$
$t_{3\,top}$
$t_{3\,bottom}$
$t_{2\,top}$
$t_{2\,bottom}$
4
3
2

pixel locations for bounding boxes for each object.

Labels for all objects in the image.

Trigger Word Classification

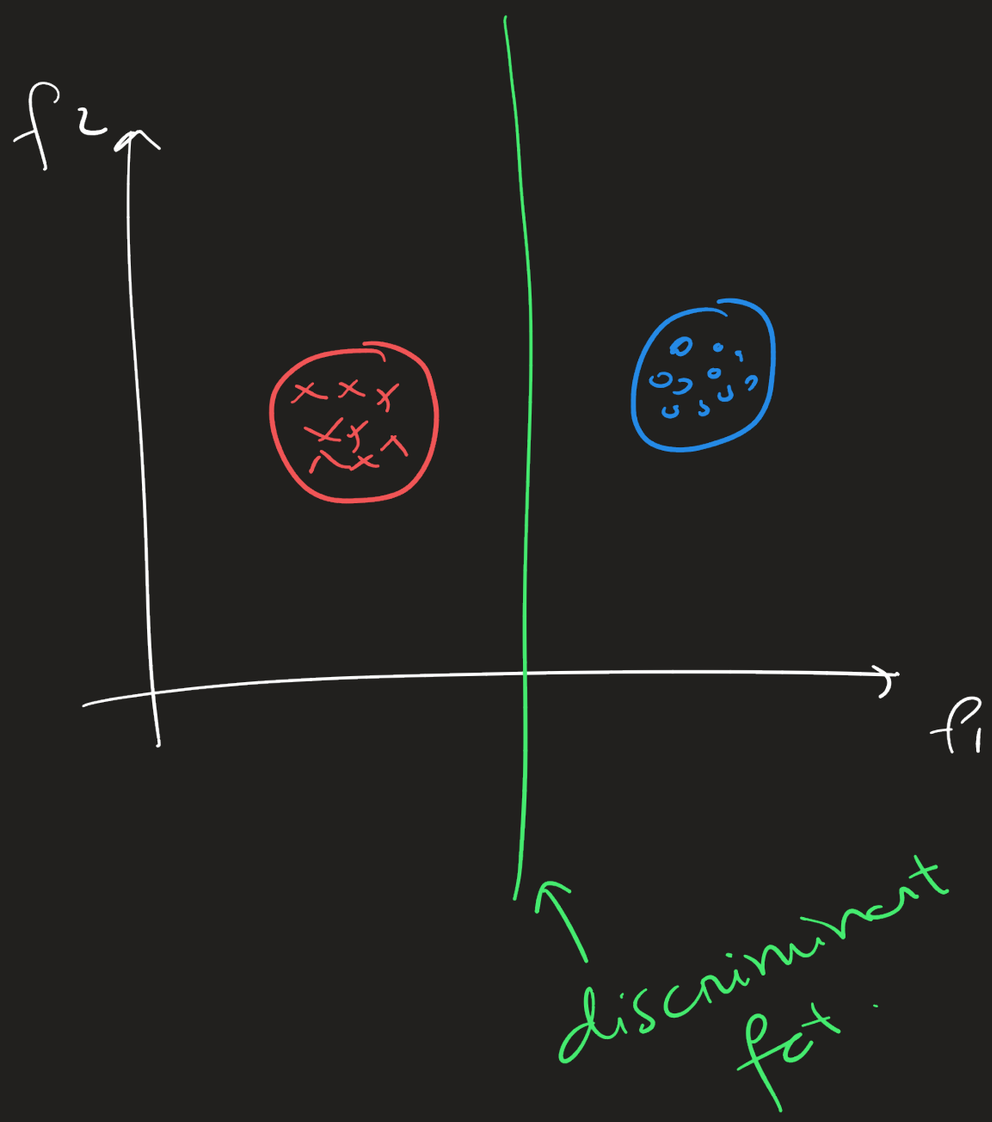
Input: Audio file (2-5 seconds)

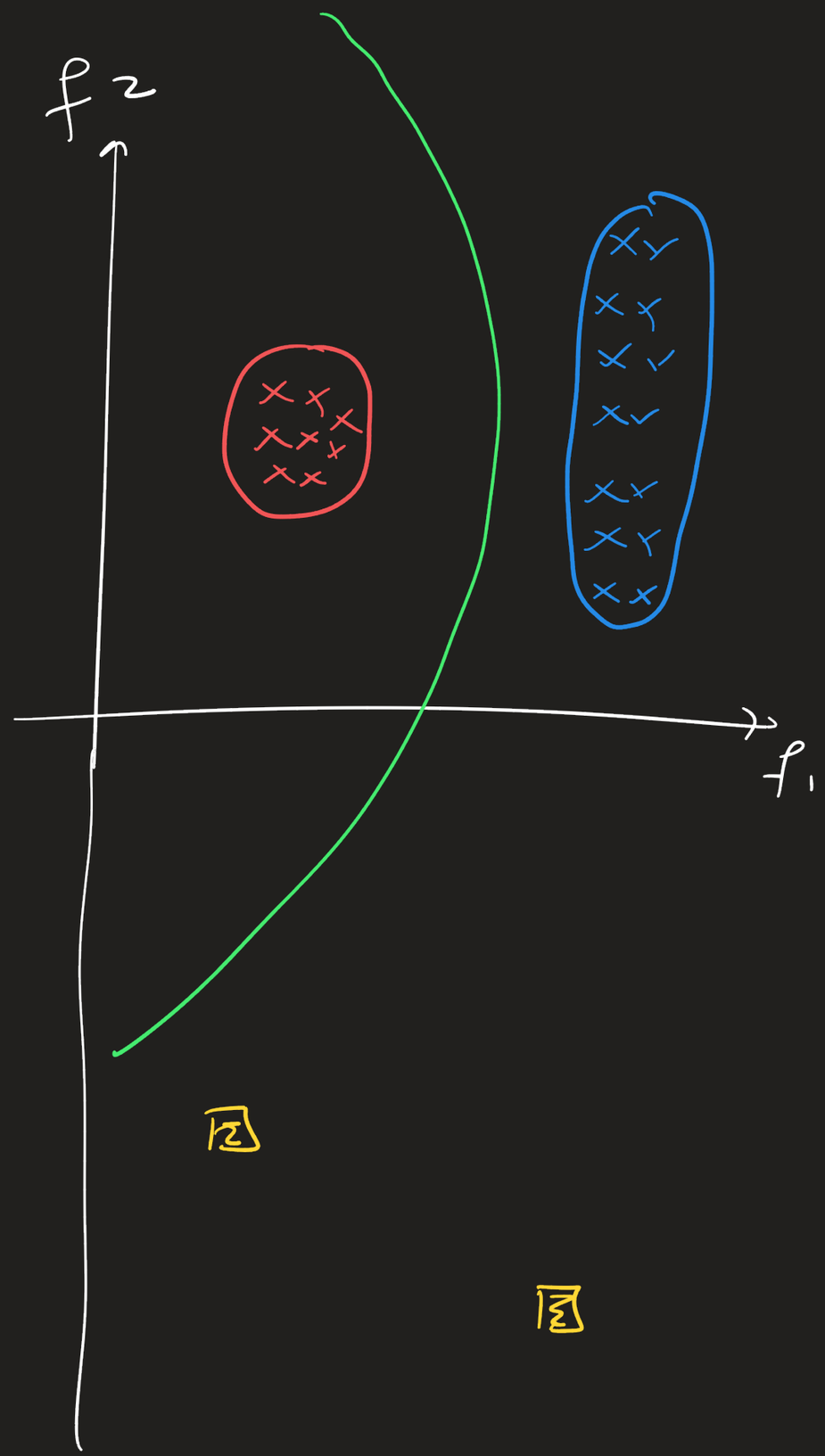
Record reading of sentence Containing trigger word : "Let's go to the grocery store" ⟶ go (0)

"Stop at traffic light" ⟶ stop (1)

⋮

LABEL


f2
f1
discriminant fct.


f2
f1