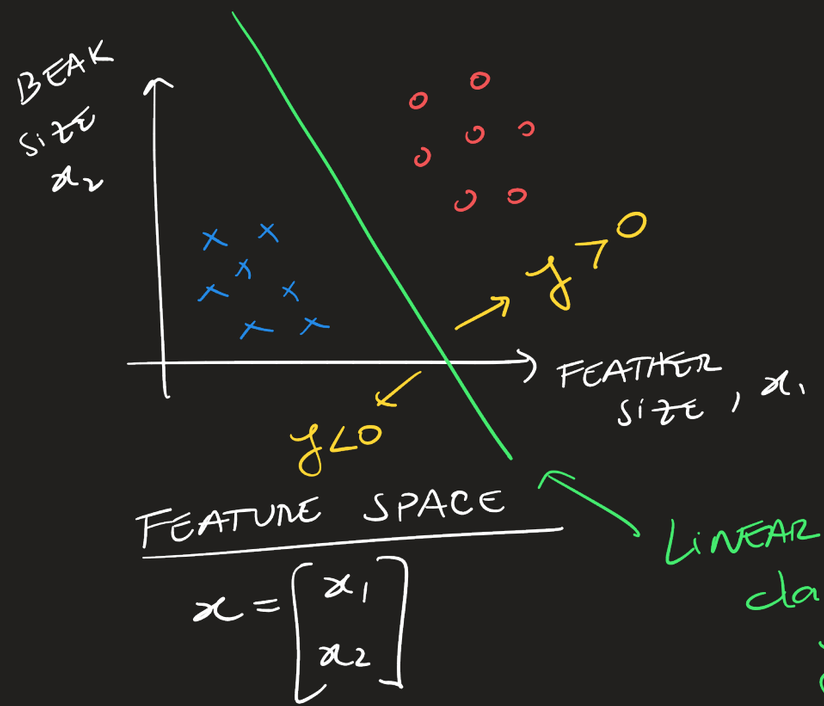


classes or target labels

O : MACAW (LABEL 1)

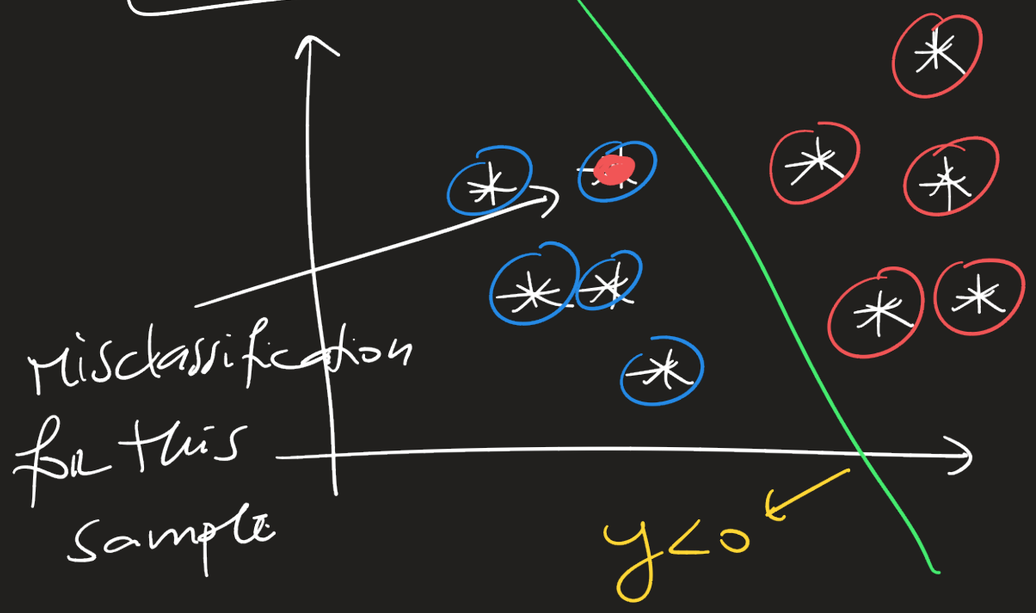
X : CONURE (LABEL 0)

If $y>0$: predict 1

If $y<0$: predict 0

TRAINING PHASE

TEST PHASE



DATASET $X = \{(x_i, t_i)\}_{i=1}^{N}$, $x_i \in \mathbb{R}^2$, $t_i \in \{0, 1\}$

100%

70%

80%

30%

20%

TRAINING SET

- used to choose system configurations
- used to tune hyperparameters of model

TEST SET

- used to evaluate performance of final trained model
- it must NOT be used during training to avoid data leakage.

TRAIN FOLD

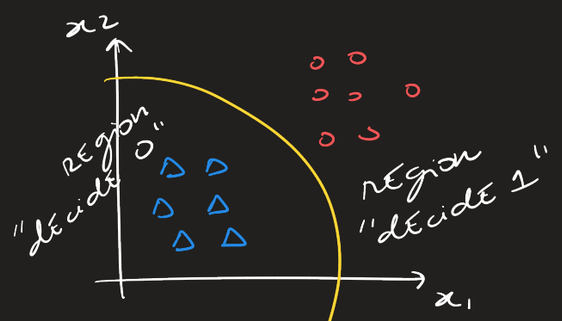
- used to train the model's parameters

VALIDATION FOLD

- used during training to check the model's generalization ability.

classifiers

DISCRIMINATIVE



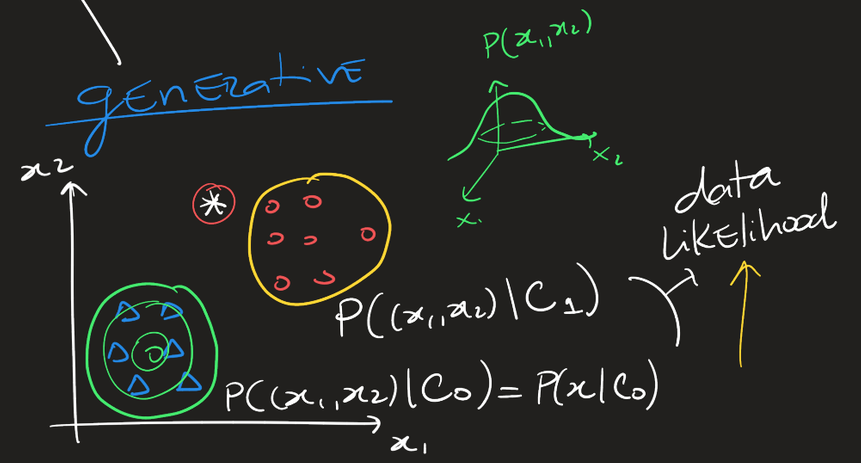
Discriminate / separate feature space into class regions

/

probabilistic discriminative classifier

e.g. Logistic Regression

GENERATIVE



Models a probability distribution for each class.

$$P(C_0|*) + P(C_1|*) = 1$$

data likelihood

(class) prior prob.

$$0.8 = P(C_0|*) = \frac{P(*|C_0)\,P(C_0)}{P(*)}$$

posterior probability

Bayes' theorem

Evidence prob.

$$0.2 = P(C_1|*) = \frac{P(*|C_1)\,P(C_1)}{P(*)}$$

If $P(C_0|*) > P(C_1|*)$:
decide $* \in C_0$

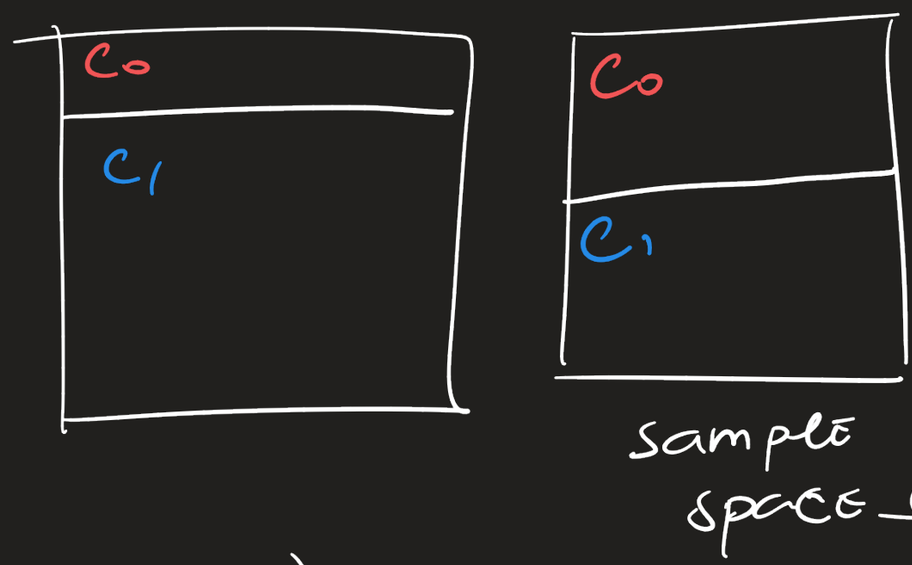
Otherwise:
decide $* \in C_1$

Evidence prob.

$$P(*) = \underbrace{P(*|C_0)}\cdot P(C_0) + \underbrace{P(*|C_1)}\cdot P(C_1)$$

↓

Law of Total probability.

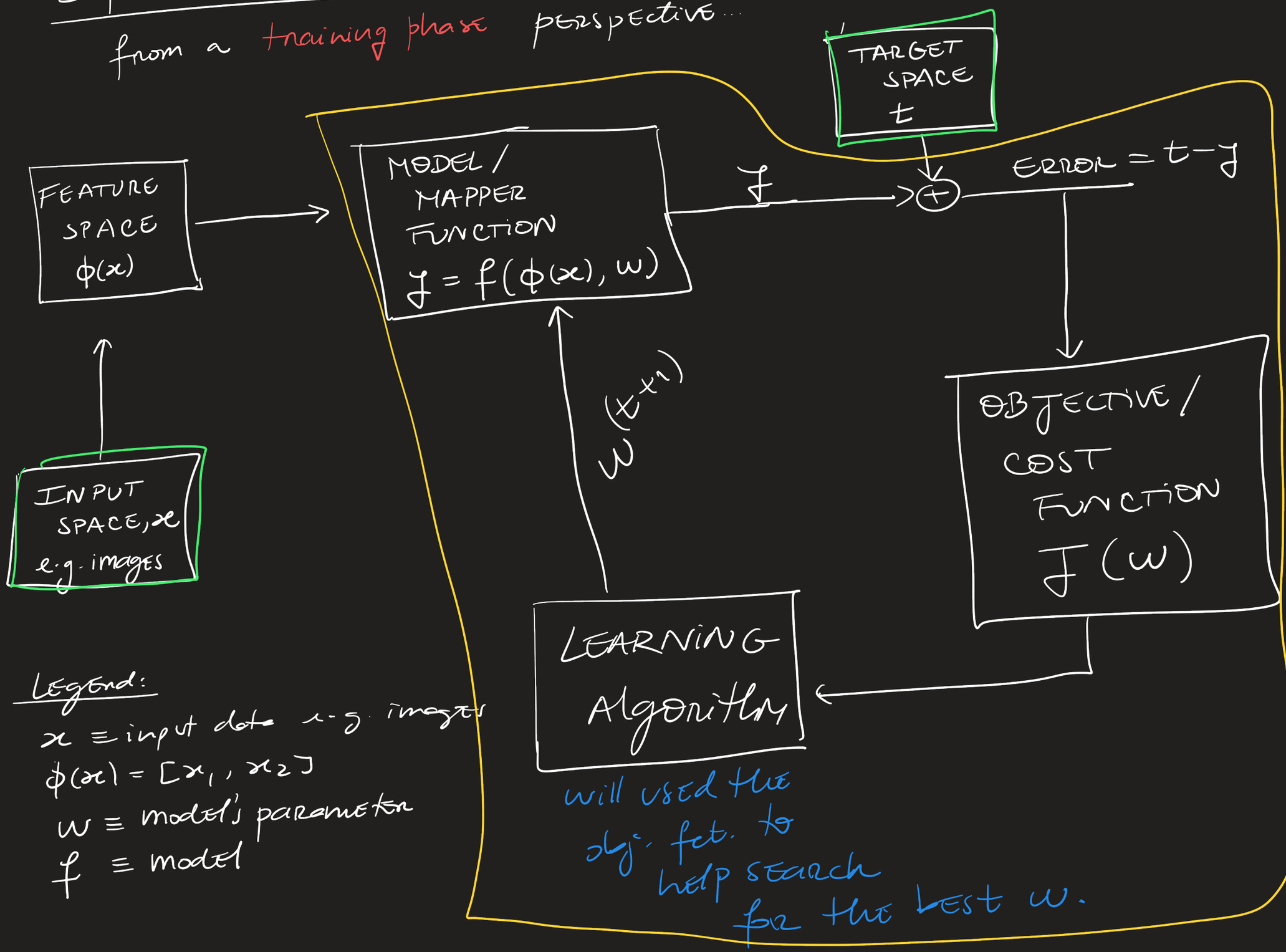


$$= G(*|\mu_0, \Sigma_0)\cdot P(C_0) + G(*|\mu_1, \Sigma_1)\cdot P(C_1)$$

↙ for example

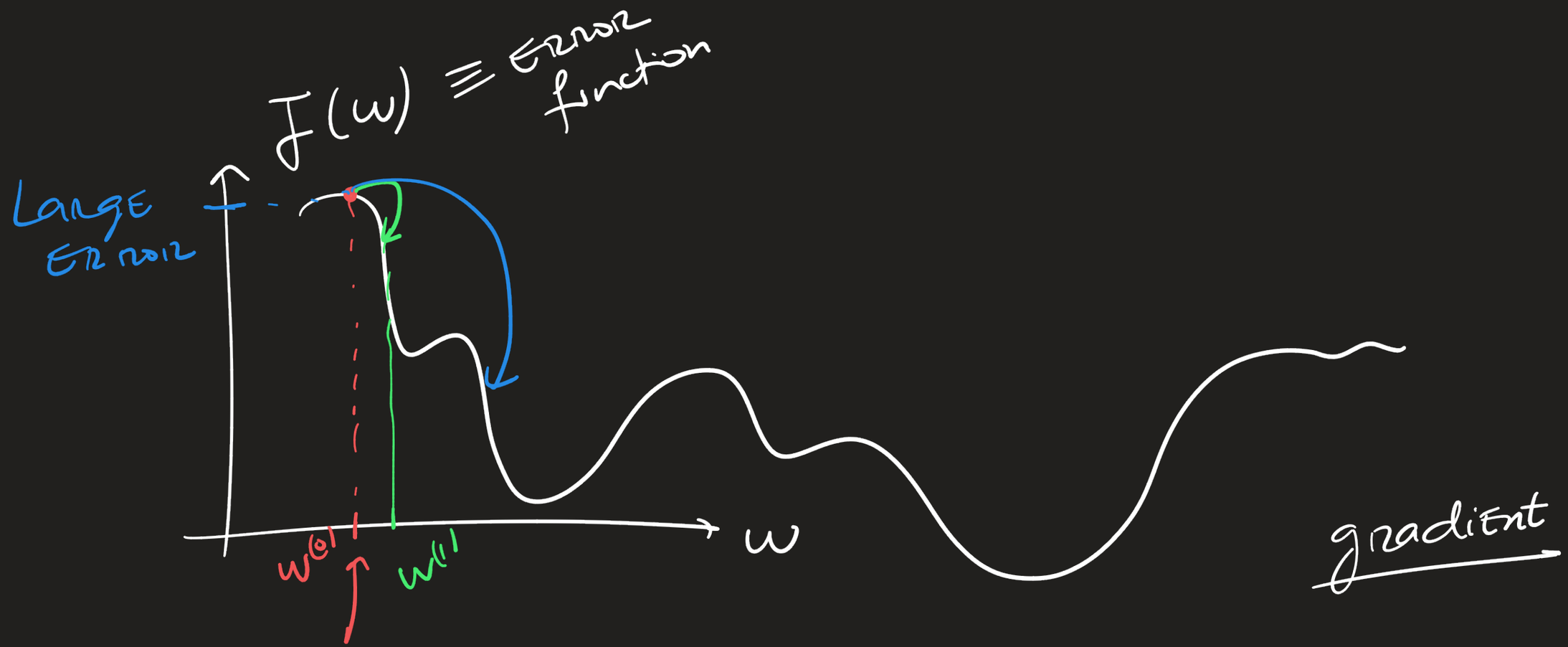
Supervised Learning Flowchart

from a training phase perspective...



Legend:

- $x \equiv$ input data e.g. images
- $\phi(x) = [x_1, x_2]$
- $w \equiv$ model's parameter
- $f \equiv$ model



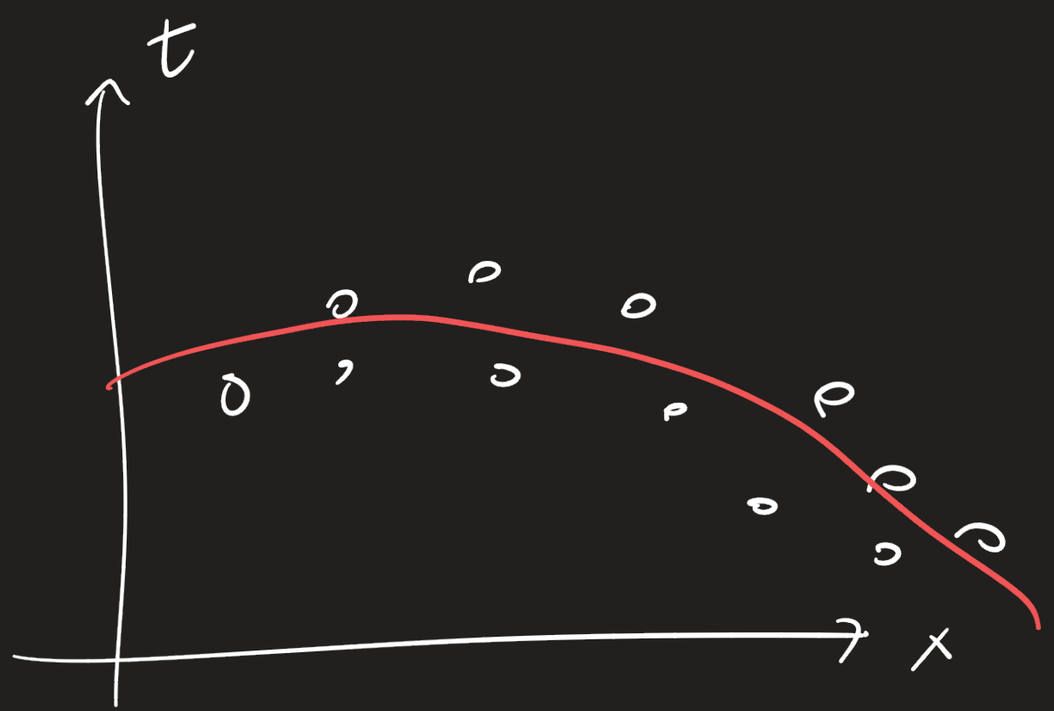
initial parameters

$$w^{(1)} = w^{(0)} - \eta \cdot \nabla J(w^{(0)})$$

learning rate

Gradient Descent

example of a learning algorithm.



Regression task

↳ Supervised Learning.