MODELS FOR CLASSIFICATION

LECTURE 4
SECTION 1
JUNE 8TH

INSTITUTE FOR APPLIED
COMPUTATIONAL SCIENCE
AT HARVARD UNIVERSITY

UNIVERSITY of
RWANDA

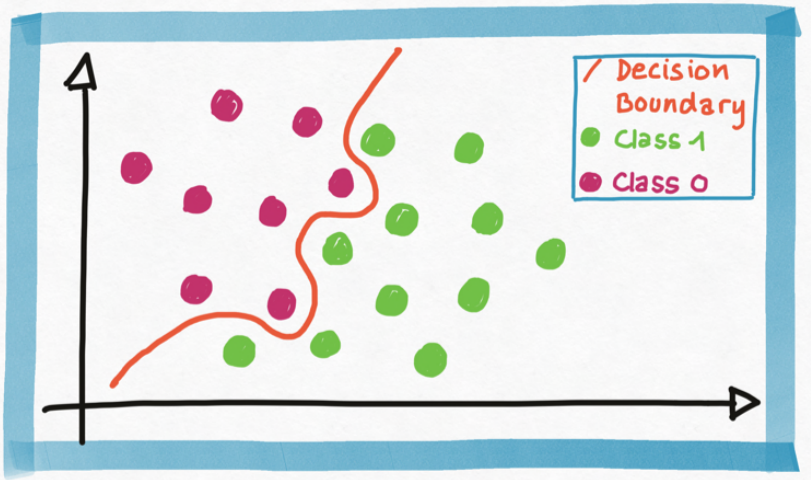
INTRODUCTION TO CLASSIFICATION

CLASSIFICATION AND DECISION BOUNDARIES:

When the label y in our training set is a category, then our prediction task is called classification. Our model is called a classifier.

Classification is the task of finding a boundary in the input space that separates out the different categories or classes in the data. This boundary is called the decision boundary.

A classifer models the decision boundary as a mathematical function.

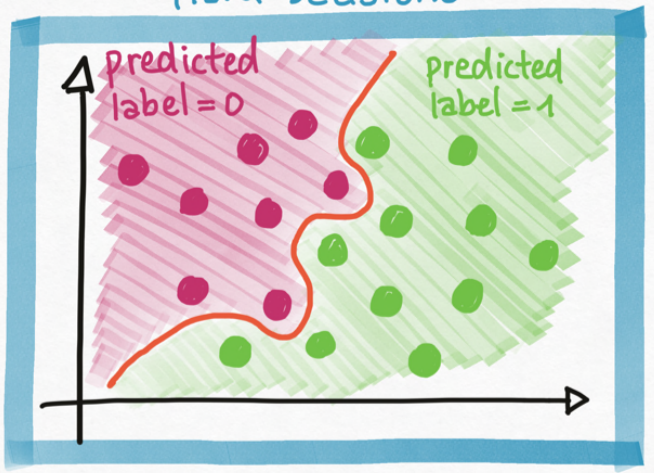


HARD DECISIONS VS SOFT DECISIONS:

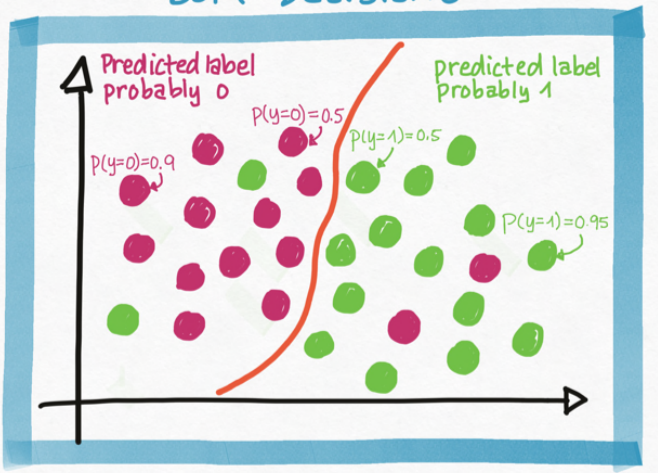
If the classes are clearly separated, then we should classify one side of the decision boundary as one class and everything on the other side as the other class. This decision has no uncertainty, it is a hard decision.

When the classes overlap and cannot be clearly separated, then we should classify points on each side of the decision boundary with some probability. This decision is a soft decision.

Hard Decisions



Soft Decisions



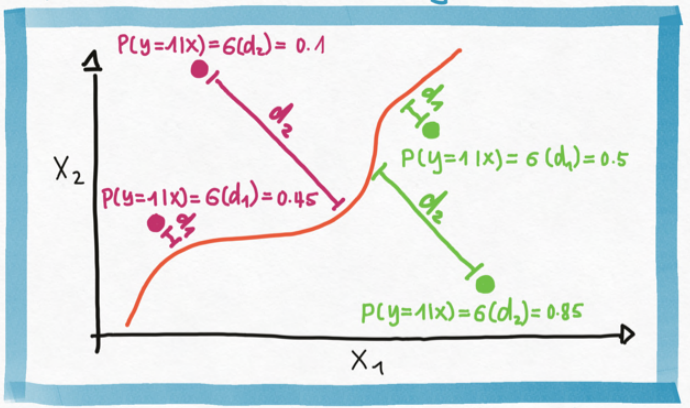
FORMALIZING CLASSIFICATION PROBABILITIES

GENERATING CLASSIFICATION PROBABILITIES:

But how do we make soft decisions? We model the probability of a point being labeled class as a function of its distance to the decision boundary.

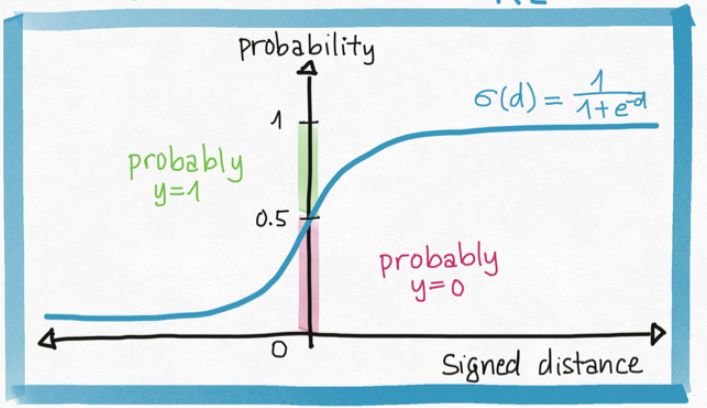
The farther from the boundary the more certain we are of the label.
The closer to the boundary the more uncertain we are of the label.

We measure the signed distance d from a point to the decision boundary, this will be an arbitrary real number.



points on one side of the boundary have positive d.
points on the other side have negatively signed d.

We transform d into a probability, a number between 0 and 1, using the sigmoid function: $\sigma(d) = \frac{1}{1+e^{-d}}$



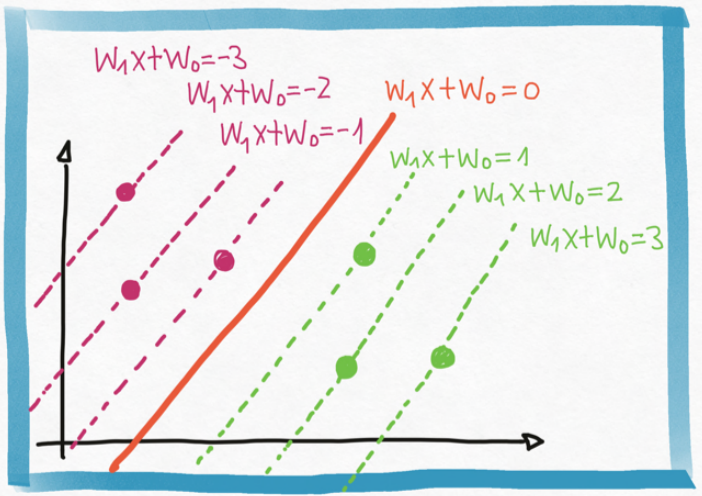
The more negative d is, the closer to zero σ(d) will be, i.e. y=0.
The more postive d is, the closer to one σ(d) will be, i.e. y=1.

MODELING DISTANCE TO THE DECISION BOUNDARY:

We model the decision boundary as a surface, i.e. the points obeying the equation: $f_w(x)=0$.

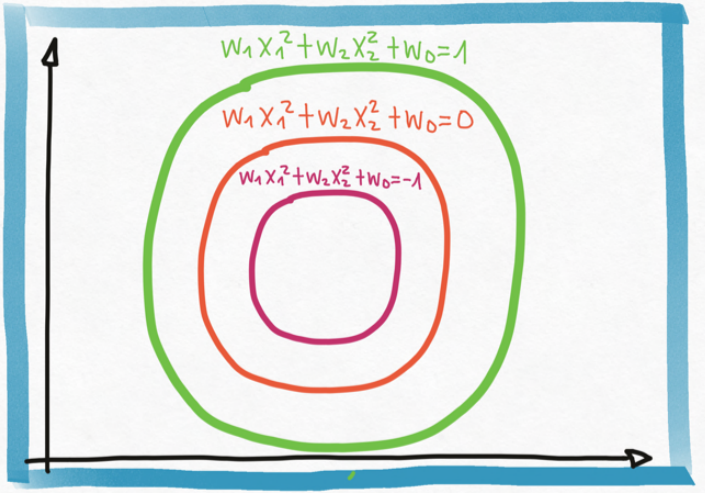
Then the signed distance from x to the decision boundary is given by $d=f_w(x)$.

Linear Decision Boundary



$$f_w(x)=\sum_{d=1}^{D} w_d x_d = w^T x$$

Quadratic Decision Boundary



$$f_w(x)=\sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij} x_i x_j$$

PROBABILISTIC MODEL FOR CLASSIFICATION

THE LOGISTIC REGRESSION MODEL:

Given training data $D = \{(x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)})\}$, where $y^{(n)} = 0$ or 1, we assume the following probabilistic model:

$$y^{(n)} \sim \text{Bernoulli}\left(\sigma\left(f_w(x^{(n)})\right)\right)$$

where σ is the sigmoid function, f_w is a function with unknown parameters w.

In other words: $p(y^{(n)} = 1 \mid x^{(n)}, w) = \sigma\left(f_w(x^{(n)})\right)$

This model is called the Logistic Regression Model.

The likelihood of the training set D is given by:

$$\mathcal{L}(w) = \prod_{n=1}^{N} p(y^{(n)} \mid x^{(n)}, w) = \prod_{n=1}^{N} \underbrace{\sigma(f_w(x^{(n)}))^{y^{(n)}} (1 - \sigma(f_w(x^{(n)})))^{(1-y^{(n)})}}_{\text{Bernoulli}\left(y^{(n)};\, \sigma(f_w(x^{(n)}))\right)}$$