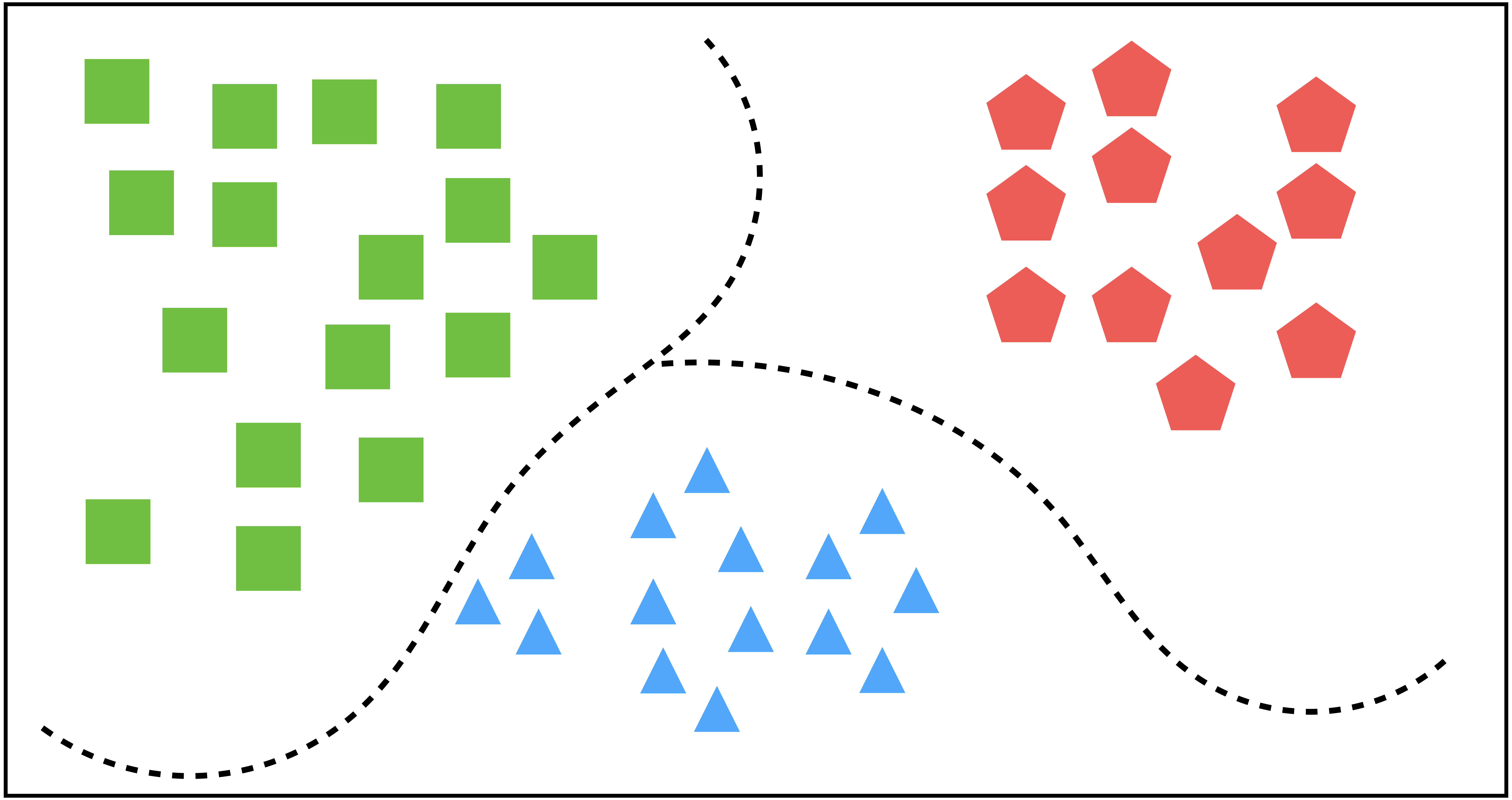
COMPUTATIONAL ASTROPHYSICS

Observatorio
Astronómico
Nacional

Computational Astrophysics

08. Classification

Eduard Larrañaga
Observatorio Astronómico Nacional
Universidad Nacional de Colombia



Classification in Astrophysics

Classification in Astronomy

Classification is a fundamental problem in (nearly) all subfields of astronomy.

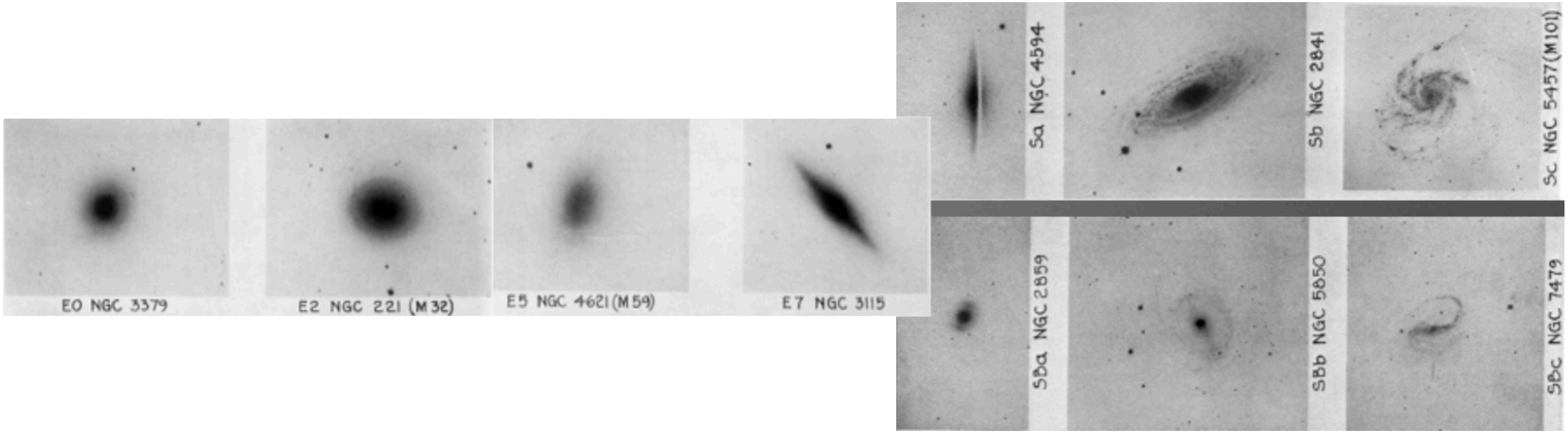
Classification schemes in astronomy are usually well-argued but:

- Sometimes subjective class boundaries are drawn.

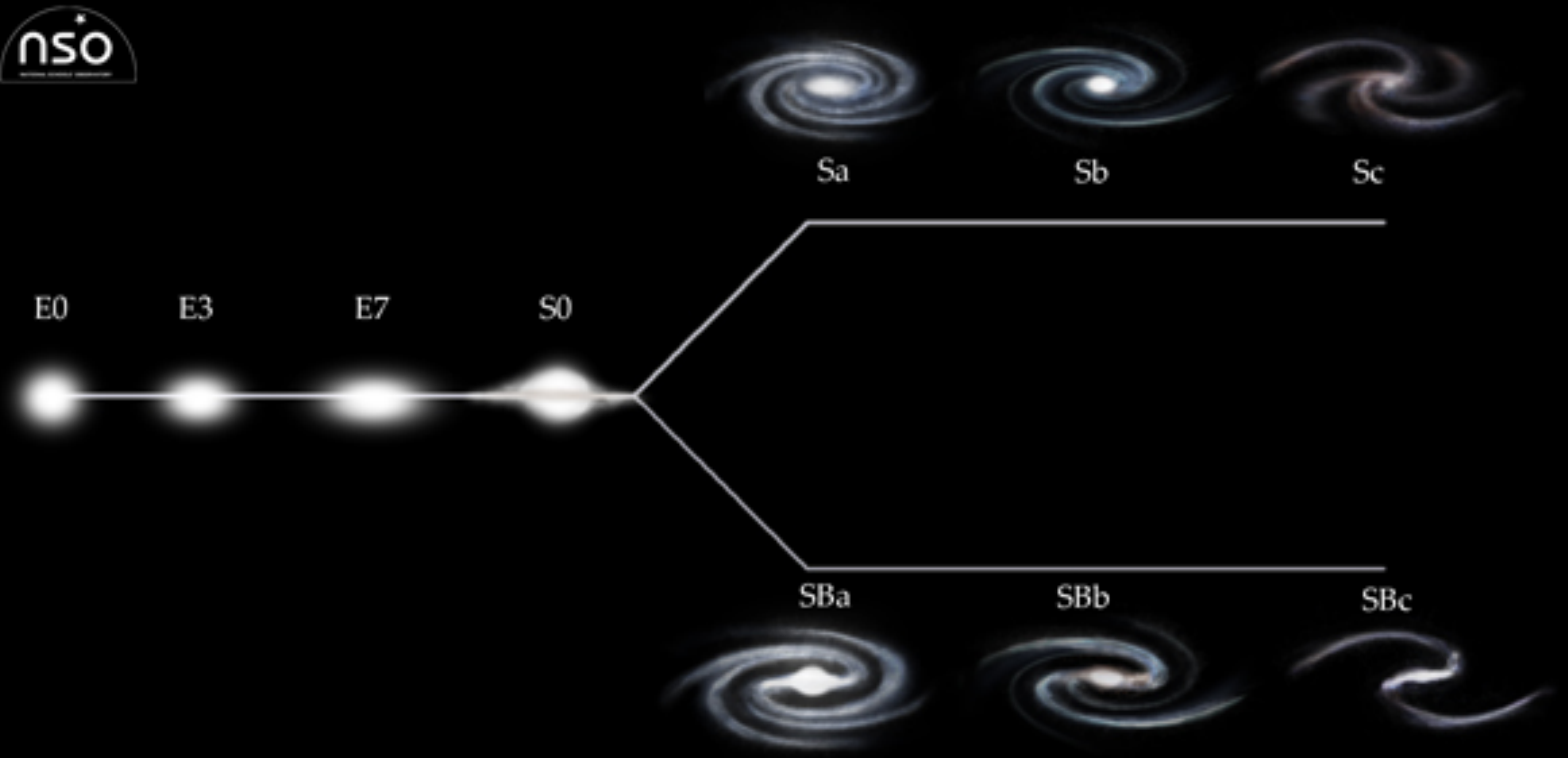
- Is usual that small samples are used for the scheme and then, the results are propagated to larger sets.

- Is usual that the model is developed in low-dimensional spaces.

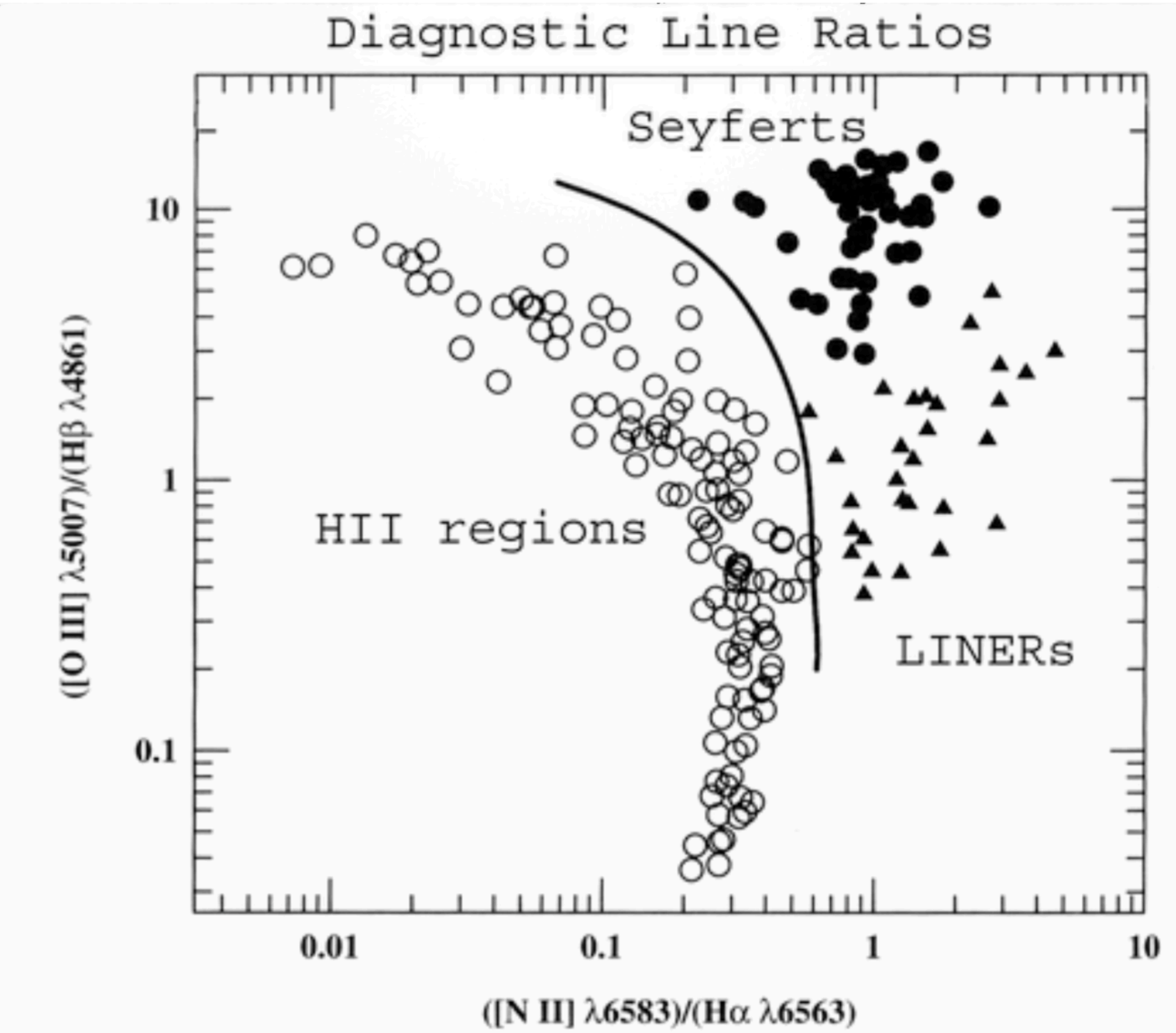
Galaxy Morphology Classification



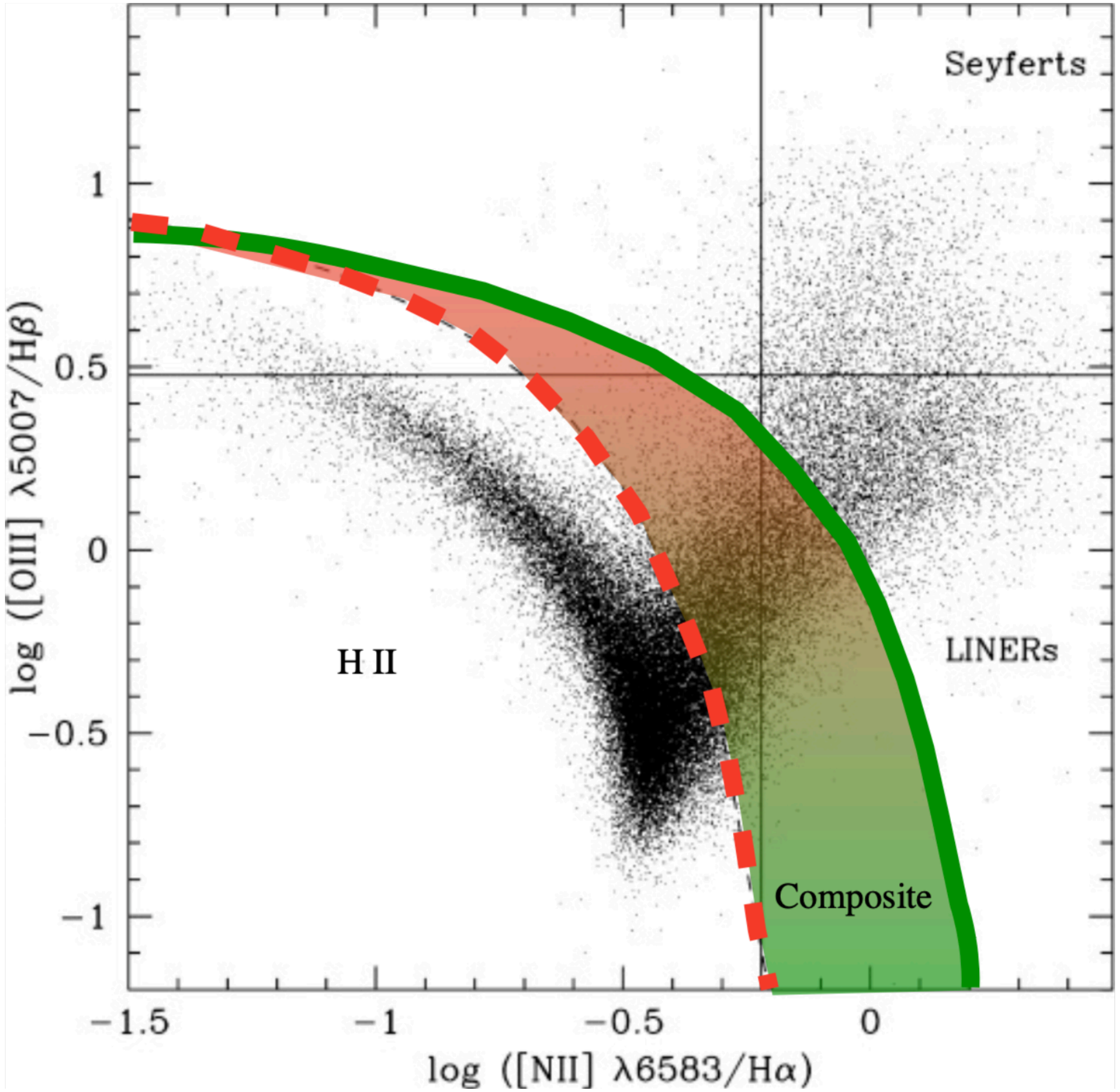
Hubble's Galaxy Classification (c. 1939)



Baldwin, Phillips & Terlevich (BPT) Diagram

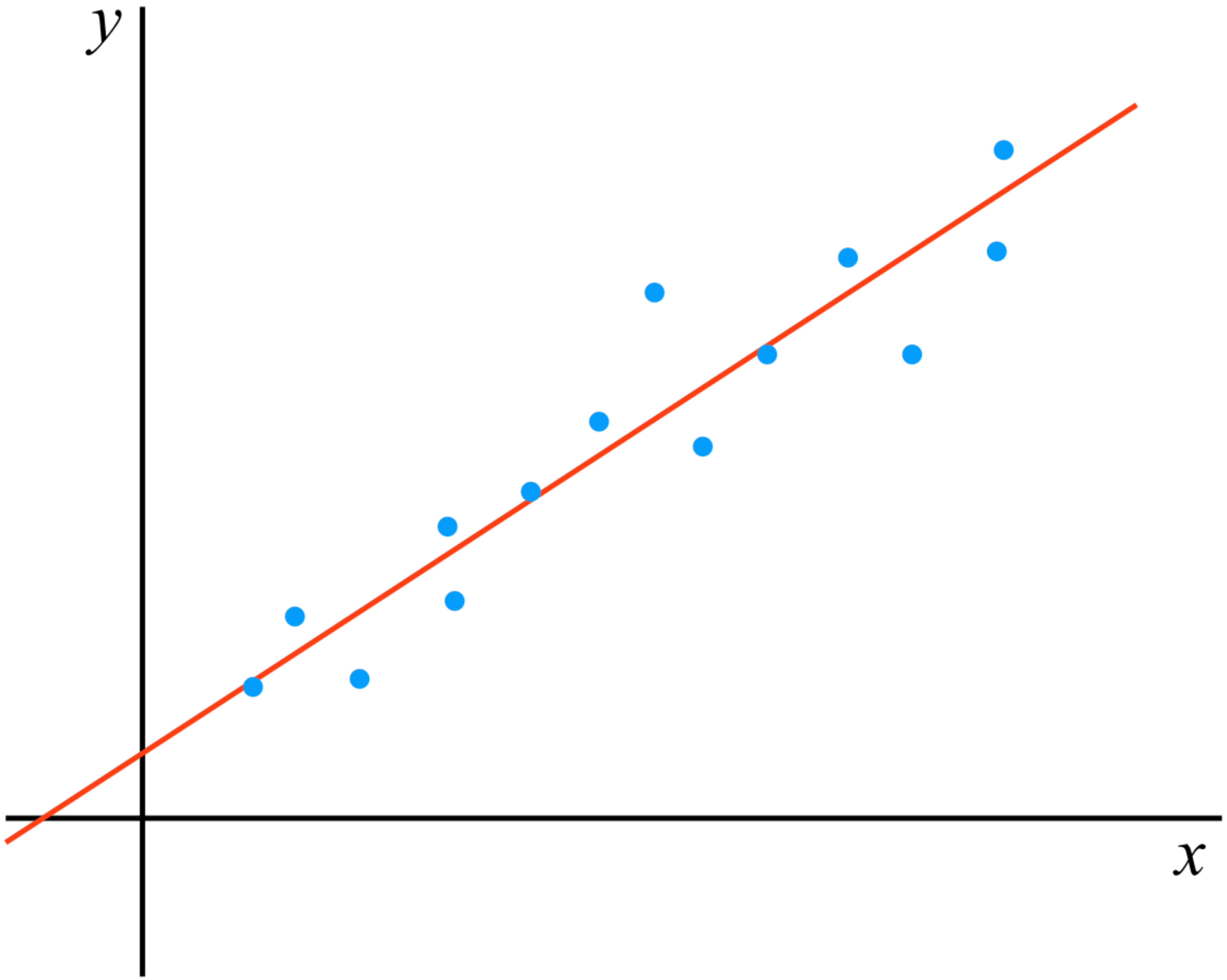


Baldwin, Phillips & Terlevich (BPT) Diagram



Logistic Regression (Classification) Algorithm

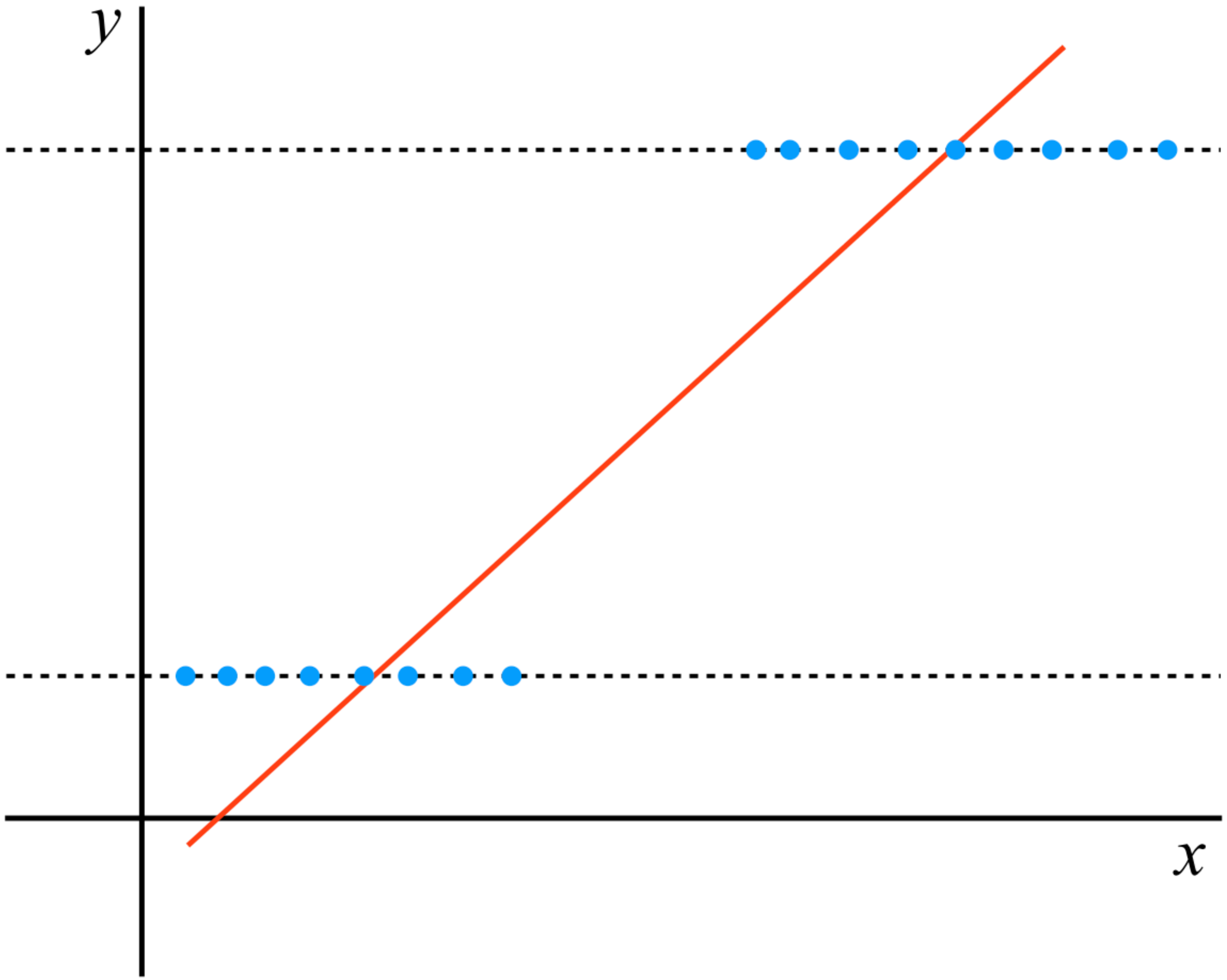
Linear Regression



Supervised Algorithm

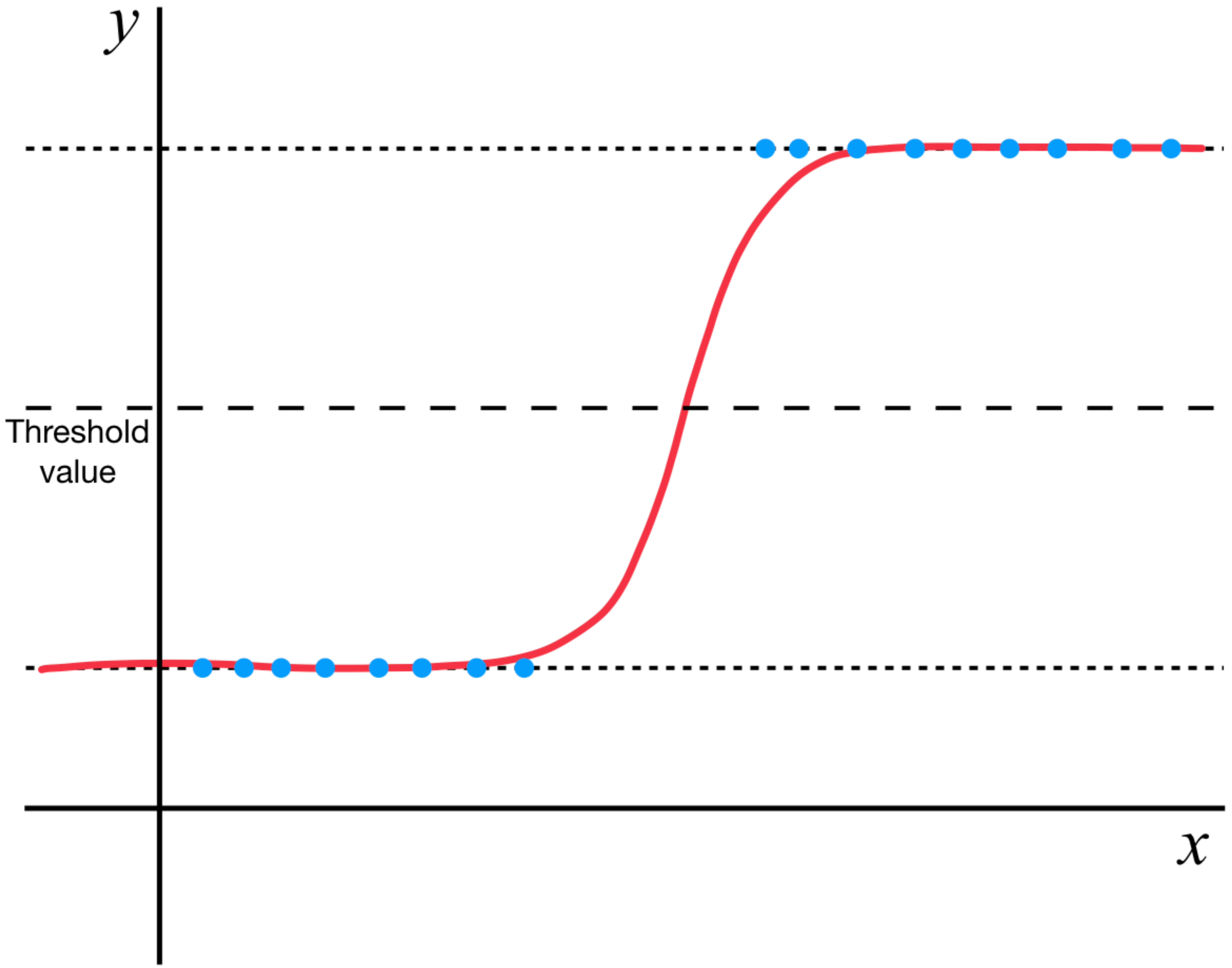
$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n$$

Binary Classification



The target variable (output) can take only two discrete values (say 0 and 1)

Logistic Binary Classification



Including a **sigmoid** activation function, the curve may fit better the data.

Logistic Binary Classification

Sigmoid function

$$y = S(z) = \frac{1}{1 + e^{-z}}$$

where

$$z = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n.$$

This gives the relation

$$\log\left[\frac{y}{1 - y}\right] = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n$$

Logistic Binary Classification in 1-D

$$y_p(x; W, b) = \sigma(z(x; W, b))$$

where

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

$$z(x; W, b) = Wx + b.$$

and then

$$y_p(x; W, b) = \frac{1}{1 + e^{-(Wx+b)}}$$

Logistic Binary Classification in 1-D

Cost Function

$$\cancel{f_c = \frac{1}{n}\sum_{i=1}^{n}\left(y_p(x_i) - y_i\right)^2}$$

Logistic Binary Classification in 1-D

Cost Function

In order to define the correct cost function, consider the following probabilities:

- Probability to obtain a result of $y_i = 1$ for a given input x_i,

$$P(y_i = 1 \,|\, x_i; W, b) = \frac{1}{1 + \exp\left[-\left(Wx_i + b\right)\right]}$$

- Probability to obtain a result of $y_i = 0$ for a given input x_i

$$P(y_i = 0 \,|\, x_i; W, b) = 1 - P(y_i = 1 \,|\, x_i; W, b)\,.$$

Logistic Binary Classification in 1-D

Maximum Likelihood

In general, given all possible outcomes from a dataset $D = \{(x_i, y_i)\}$ with the binary labels $y_i \in \{0,1\}$, where the data points are drawn independently, it is used the **M**aximum **L**ikelihood **E**stimation (**MLE**) principle.
It states that we need to maximize the probability of seen the observed data and this can be written as the product od the individual probabilities of a specific outcome y_i, i.e.

$$P(D; W, b) = \prod_{i=1}^{n} \left[P(y_i = 1 \,|\, x_i; W, b)\right]^{y_i} \left[P(y_i = 0 \,|\, x_i; W, b)\right]^{1-y_i}$$

$$P(D; W, b) = \prod_{i=1}^{n} \left[P(y_i = 1 \,|\, x_i; W, b)\right]^{y_i} \left[1 - P(y_i = 1 \,|\, x_i; W, b)\right]^{1-y_i}$$

Logistic Binary Classification in 1-D

Maximum Likelihood

Taking the logarithm of this probability, we obtain the (log-likelihood) cost as

$$C(W,b) = \sum_{i=1}^{n} \left\{ y_i \left[\log P(y_i = 1 \,|\, x_i; W, b) \right] + (1 - y_i) \log \left[1 - P(y_i = 1 \,|\, x_i; W, b) \right] \right\}.$$

Then, we will define a cost function as

$$f_c = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log(y_p(x_i)) + (1 - y_i) \log(1 - y_p(x_i)) \right]$$

where we included a minus sign in order to obtain a function that must be minimized.

Logistic Binary Classification in 1-D

Maximum Likelihood

$$f_c = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(y_p(x_i)) + (1-y_i)\log(1-y_p(x_i))\right]$$

- For a sigle sample with target value $y_i = 0$, the cost function reduces to $f_c = -\log(1-y_p)$. Note that a prediction near $y_p \sim 1$ gives a huge cost, $f_c \to \infty$, while a prediction near $y_p \sim y_0 = 0$ gives a low cost, $f_c \to 0$.

- For a sigle sample with target value $y_i = 1$, the cost function reduces to $f_c = -\log(y_p)$. This time, a prediction of $y_p \sim 0$ gives a huge cost, $f_c \to \infty$, while a prediction of $y_p \sim y_0 = 1$ gives a low cost, $f_c \to 0$.

$f_c(W, b)$ is known in statistics as the **cross entropy** for two density functions y and y_p. In a future lecture we will study the entropy and cross entropy in detail.

Logistic Binary Classification in 1-D

Gradient of the Cost Function

$$
\begin{aligned}
\frac{\partial f_c}{\partial W} &= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{y_p}\frac{\partial y_p}{\partial W} - \frac{1-y_i}{1-y_p}\frac{\partial y_p}{\partial W}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{y_p} - \frac{1-y_i}{1-y_p}\right]\left(y_p(1-y_p)x_i\right) \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i(1-y_p) - (1-y_i)y_p\right]x_i \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i - y_p\right]x_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[y_p - y_i\right]x_i\,.
\end{aligned}
$$

Logistic Binary Classification in 1-D

Gradient of the Cost Function

$$
\begin{aligned}
\frac{\partial f_c}{\partial b} &= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{y_p}\frac{\partial y_p}{\partial b} - \frac{1-y_i}{1-y_p}\frac{\partial y_p}{\partial b}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{y_p} - \frac{1-y_i}{1-y_p}\right]\Big(y_p(1-y_p)\Big) \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i(1-y_p) - (1-y_i)y_p\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i - y_p\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[y_p - y_i\right].
\end{aligned}
$$

@astronomiaOAN

AstronomiaOAN

@astronomiaOAN