

# Tarea n°2

La meta de esta tarea es explorar un otro clasificador : El clasificador por regresión logística. La regresión logística se trata de regresar una función de probabilidad de pertenencia condicional  $p_i(x) = \mathbb{P}(Y = i|X = x)$  de la forma

$$p_i(x) \approx \sigma(a_i^T X + a_{i,0}),$$

donde  $\sigma$  es un mapeo continuo y creciente de  $\mathbb{R}$  a  $[0, 1]$ . En general se toma  $\sigma(t) = 1/(1 + e^{-t})$ .

1. Suponiendo que tenemos acceso a las regresiones logísticas de los  $p_i$ , describir el clasificador óptimo  $g$  basado en los  $\hat{p}_i$ .
2. Mostrar que las fronteras entre las diferentes clases son sub-conjuntos de hiperplanos (Es un clasificador lineal).
3. Usar el script siguiente para cargar los datos de iris.

```
from sklearn import datasets
iris = datasets.load_iris()
X = iris.data[:, :2]
Y = iris.target
```

4. De la librería `sklearn`, usar las funciones `DecisionBoundaryDisplay` para visualizar las  $R_i$  de la clasificación y `LogisticRegression` para definir el modelo de clasificador por regresión logística.

```
from sklearn.inspection import DecisionBoundaryDisplay
from sklearn.linear_model import LogisticRegression
```

Producir un plot que permite visualizar las regiones  $R_i$  del clasificador. Superponer los datos de entrenamiento.

5. Cuantos datos están mal clasificados ?