

# *Aprendizaje automático*

Departamento de Ingeniería en Informática  
ITBA

## Trabajo Práctico 5

- Árboles de decisión

```
T = classregtree(X,Y)
```

Crea el árbol de decisión  $T$  para la respuesta  $Y$  como función de los predictores  $x$ , que es una matriz de  $n \times m$ .

$T$  es un árbol binario donde cada nodo no terminal se divide basado en los valores de las columnas de  $x$

```
T = classregtree(X,Y,'param1',val1, 'param2',val2, ...)
```

Algunos de los parámetros que se pueden dar como input son:

- o 'names': un arreglo con los nombres de las variables predictoras, en el orden en que aparecen en la matriz  $x$  a partir del cual será creado el árbol
- o 'prune':
  - 'on': (default) obtiene el árbol completo y la secuencia óptima de sub-árboles podados
  - 'off': obtiene el árbol completo sin podas
- o 'minparent': Un número  $k$  tal que los nodos impuros pueden tener  $k$  o más observaciones que podrían (default:10)
- o 'splitcriterion': Criterio para elegir la división, algunos son:
  - 'gdi': (default) índice de diversidad de Gini

```
view(T)
```

Muestra el árbol de decisión  $T$  obtenido con `classregtree`.

Cada rama del árbol se rotula con su regla de decisión, y cada nodo terminal con el valor predicho para ese nodo.

Al hacer “click” en cada nodo se puede obtener más información sobre el mismo, por ej. el número de nodo y su clase (si es terminal).

```
sfit = eval(T,X)
```

Devuelve qué clase asignó.

```
mean(strcmp(sfit,Y))
```

Calcula la proporción correctamente clasificada.