

Practica 2 ApC

Ivan Peñarando Martínez
Joel Marco Quiroga Poma
Ferran Martínez Reyes



Apartado B



La base de datos

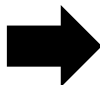
- Mobile Price Classification
- 21 Atributos:
 - Continuos (Battery Power)
 - Binarios (BT, Wi-Fi, ...)
 - Categóricos (Price_range)
- Target escogido: PRICE_RANGE = **0** (Low), **1** (Medium), **2** (High), **3** (Very High)

Único atributo categórico, interesante para clasificación → Clasificar gamma de precio del teléfono dado sus componentes.

- ¿Etiquetas balanceadas?

La media tendría que ser 2

```
print(dataset.mean())
```



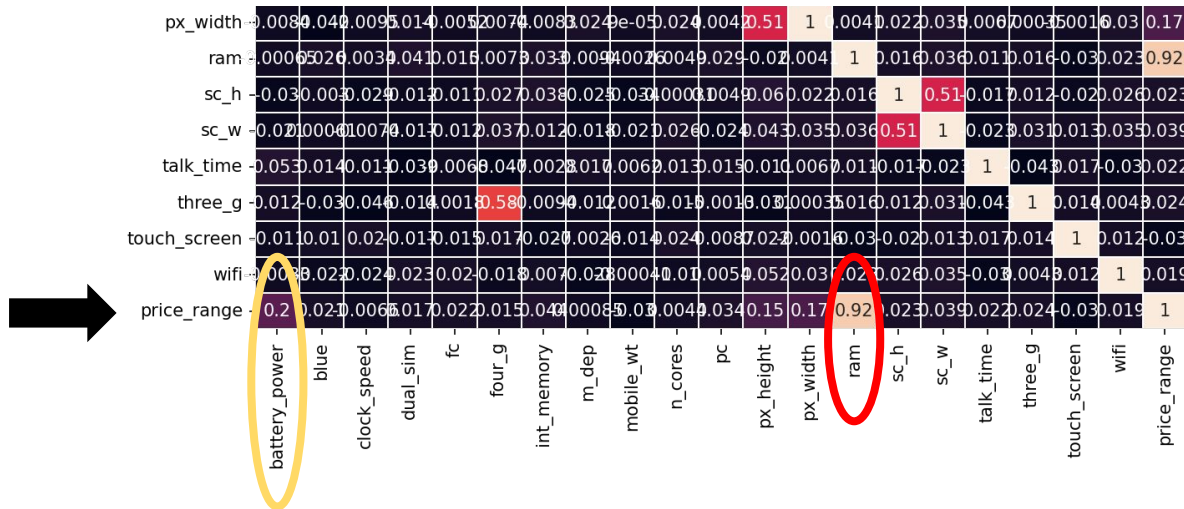
```
price_range    1.50000
```



Correlación entre atributos

Atributo a predecir: Price_range

...

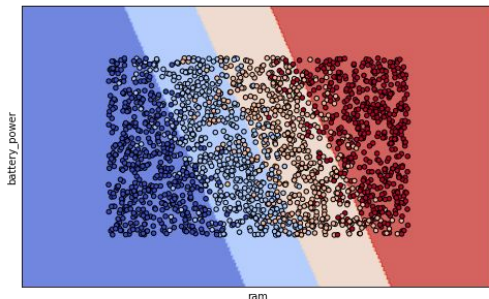


Posibles predictores: RAM, (Battery_power?)



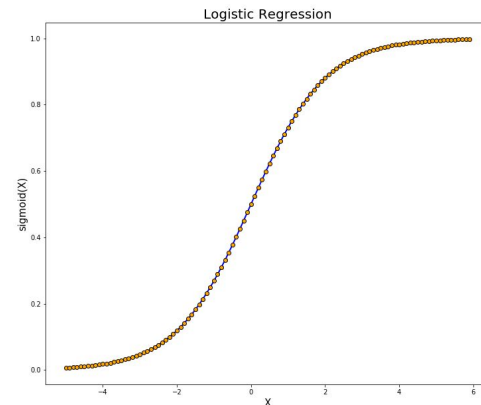
Preprocessing

- ¿Datos Normalizados? ❌ (Varianzas/Rangos de los valores son diferentes)
 - ¿Necesario normalizar? ✅ MVS asume que datos están normalizados en rango [0, 1] (Normalización vía escalado a rango)
- Datos sin errores (NaN, dataset.mean, ...) ✅
- Codificación atributos categóricos → Relación Ordinal → OrdinalEncoder
- No es necesario aplicar PCA → ≤ 2 Predictores → Visualizable:

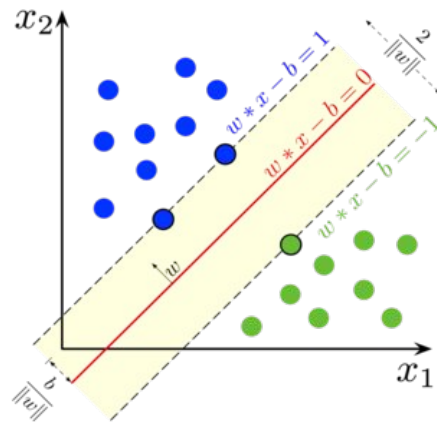
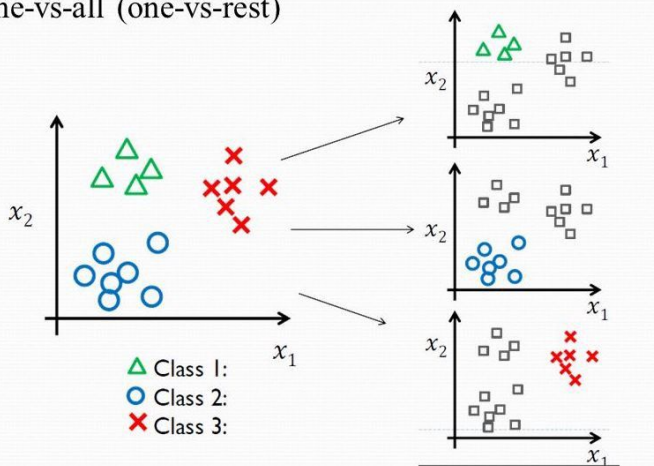


Model Selection

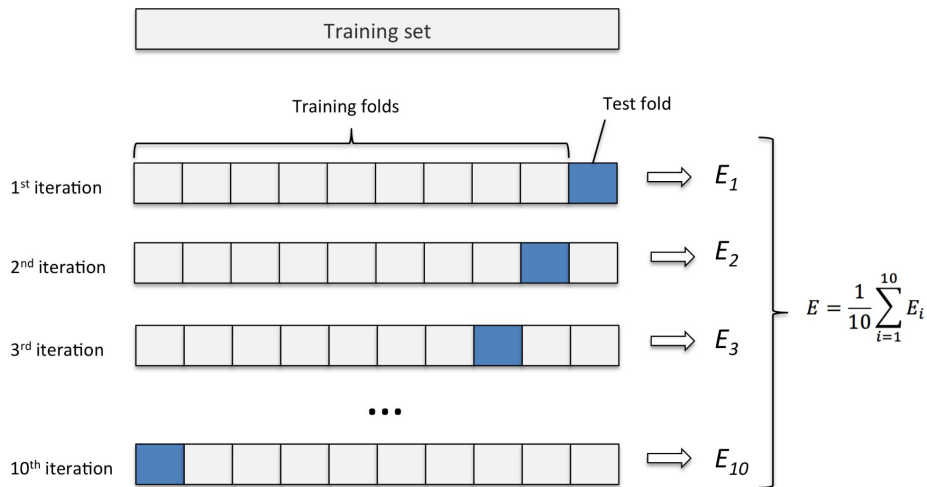
- Regresión logística
- Máquinas de vectores de soporte (SVM)
- “price_range” categórico → Clasificación Multiclase OvR



One-vs-all (one-vs-rest)

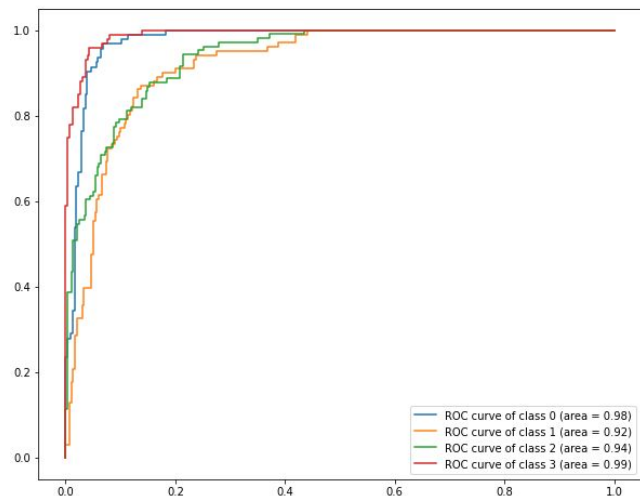


Crossvalidation



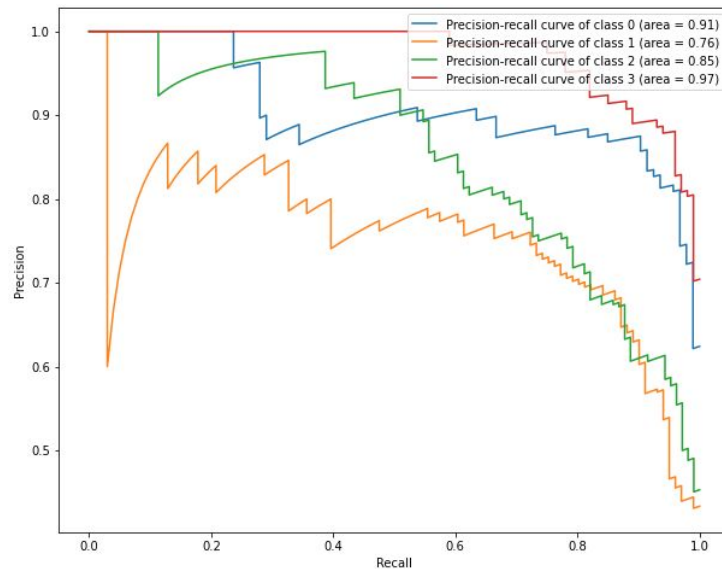
- Importante para comprobar si nuestro modelo está haciendo un buen trabajo
- Garantiza que los resultados sean independientes de la partición entre train y test

Metric analysis



Curva de ROC

Valores muy buenos...



Curva precision-recall

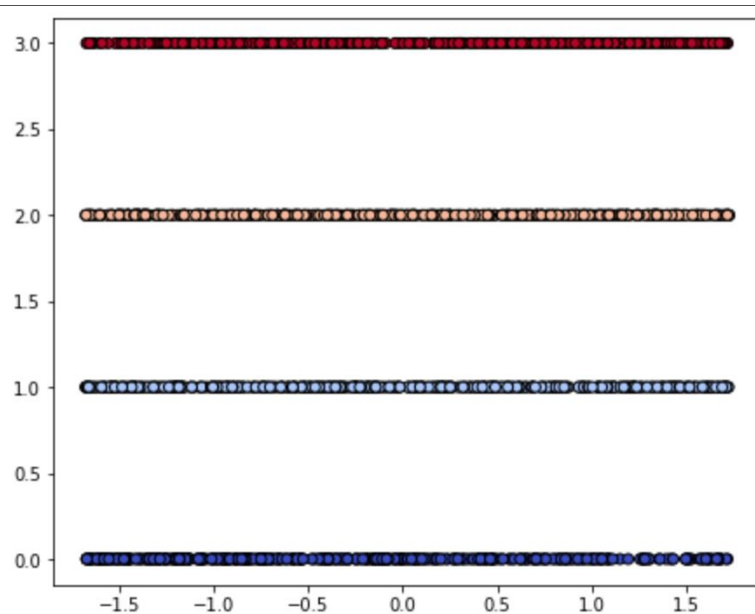
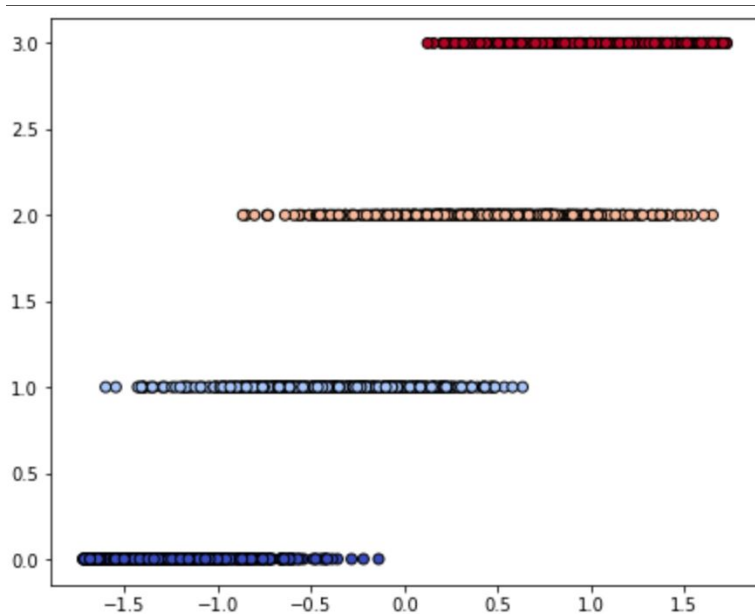
Valores no tan buenos

Apartado A

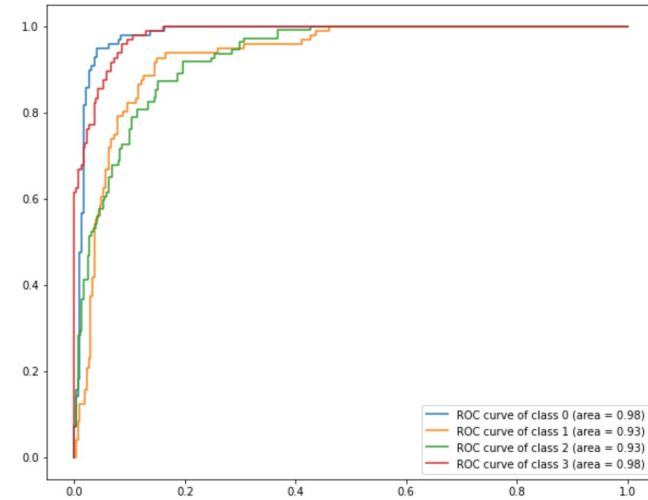
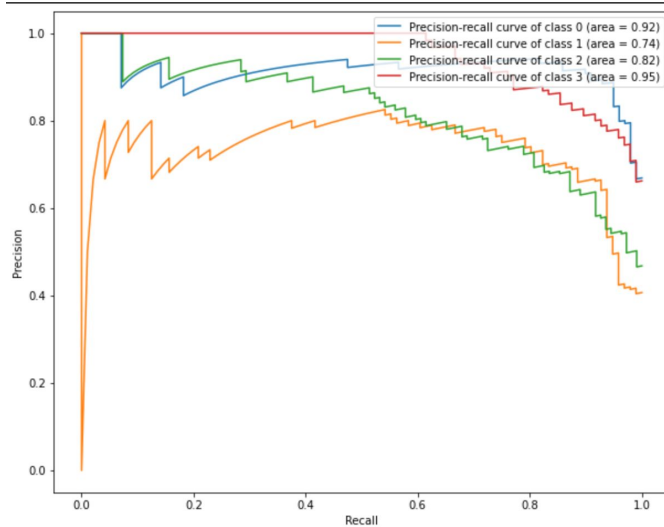


Atributos para los modelos

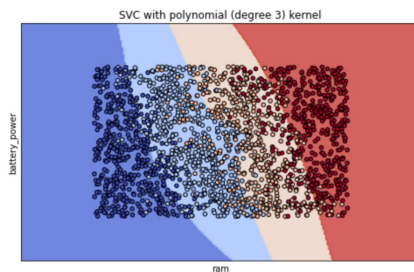
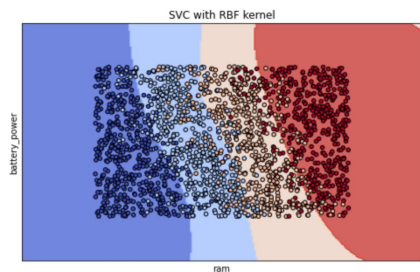
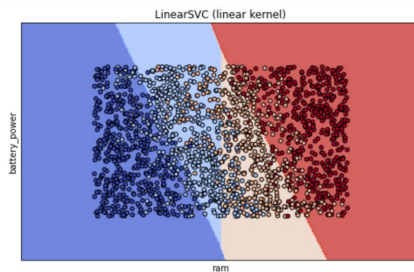
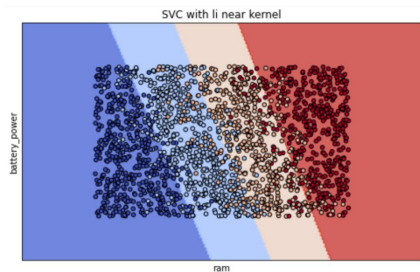
Correlación con el target de los 2 mejores atributos: ram y battery_power:



k-fold con el modelo SVC



Resultados de diferentes modelos



gráficos de clusterización de los modelos:

-SVC con kernel linear.

-SVC con kernel rbf.

-SVC con kernel poly.

-LinearSVC.

Muchas Gracias

Ivan Peñarando Martínez
Joel Marco Quiroga Poma
Ferran Martínez Reyes

