

Algoritmes de Classificació

Part 1. Coneixer els algoritmes més característics

En aquest punt, aprofitarem que tenim creat un procés on es fa ús d'un arbre de decisió, que és el detallat a la Teoria 01 (Titanic amb cross-validation) per a presentar-vos els algoritmes de classificació més habituals.

Atenció, alterant aquest procés presentat a la sessió anterior, ens allunyem del que significa una tècnica EDA, és a dir, l'objectiu en aquesta pràctica és comparar diferents algoritmes en les mateixes condicions.

Part 2. Procediment

Tornar a muntar el procés descrit a la teoria, és a dir, dataset Titanic, cross-validation, decision tree, apply-model,...

Anar canviant l'algoritme de classificació (arbre de decisió) i observar els resultats.

Omplir la taula següent d'observacions, substitueix els interrogants per les teves eleccions de paràmetre o els teus resultats.

	Parametre 1	Parametre 2	Parametre 3	Accuracy
Decision Tree	maximal_depth=20	Confidence=0.25	Minimal gain=0.1	78.69%
Decision Tree	maximal_depth=15	Confidence=0.25	Minimal gain=0.1	78.69%
Logistic Regression	C=NOT	Max-iter=NOT	Penalty o Solver=NOT	71.81%
Random Forest	n_estimators=100	Criterion = accuracy	Maximal depth = 10	96.33%
Support Vector Machine	C=10	Kernel=radial	Gamma=1.0	NOT
Support Vector Machine	C=10	Kernel diferent al anterior=?	Gamma igual al anterior=?	NOT
KNN	k_neighbors=3	weight=NOT	p=NOT	65.39%
KNN	k_neighbors=5	weight=NOT	p=NOT	65.62%
KNN	k_neighbors=10	weight=NOT	p=NOT	65.47%
Naive Bayes	Laplace_correction=true	var_smoothing	priors	89.69%

Part 3. Entregable

Document PDF amb la taula anterior completada, anotacions sobre l'execució de la comparativa (algoritmes lents, algoritmes impracticables, Accuracy baix, necessitat d'operadors addicionals, com afecta un paràmetre diferent a un mateix algoritme,...) i reflexions sobre l'interpretabilitat de les classificacions obtingudes.

En el árbol de decisión, en precisión no se encuentra una diferencia entre 20 y 15.
 Regresión logística no se ha definido ningún parámetro y su precisión es del 71.81%.
 En Máquina vectorial se han definido los parámetros, pero ha generado errores.
 KNN en las 3 ha dado una precisión similar las 3 siendo la más alta la 5.
 Naive Bayer no se ha definido los nada y su precisión es alta con un 89.69%.