

Tasca 8.1:

Interpretació de resultats d'un model de regressió lineal.

Un cop entrenat el model, els coeficients obtinguts son [61.4, 39.2, 24.7] que corresponen respectivament a les característiques ['Length of Membership', 'Time on App', 'Avg. Session Length']. Els coeficients ens diuen **en quina quantitat canvia la variable objectiu per cada variació d'una unitat de cada característica**, i també ens diu en quin sentit (signe positiu implica increment, signe negatiu correspon a decrement).

El model ens diu que **la característica més important es 'Length of Membership'**, ja que presenta el coeficient més alt. Per cada unitat de increment en 'Length of Membership', la despesa anual s'incrementa en 61,4 unitats (signe positiu). La importància relativa dels altres dos coeficients es menor, i suposen un increment de la despesa anual de 39,2 i 24,7 respectivament per cada increment d'una unitat respectivament (també increments ja que els signes son positius).

El error MSE (*Mean Squared Error*) es de 162,5 i ens dona una **mesura del error de predicció del model, penalitzant més els errors mes grans** ja està elevat al quadrat. Donat que el dataset original la característica objectiu 'Yearly Amount Spent' té un valor mig de 499,3 i una desviació estàndard de 79,3, aquest MSE es molt baix i per tant el model fa bones prediccions. De fet, l'arrel quadrada del MSE (12,7) es molt mes petita que la desviació estàndard (79,3).

El coeficient R2 es de 0,975 i ens diu la capacitat del model per captar la variabilitat de les dades del dataset. Un valor prop de 1 vol dir que el model recull molt bé aquesta variabilitat de les dades i un valor prop de 0 seria el contrari. En aquest cas, podem dir que **el model recull molt bé la variabilitat de les dades d'entrenament**. Un valor prop de 1 per el coeficient R2 també podria indicar que el sistema s'ha sobre-entrenat (*overfitting*) i el model ha après del soroll en comptes dels patrons. En el nostre cas aquesta possibilitat es descarta donat que s'han fet servir tècniques per evitar-ho (Split Train/Test de dades i validació creuada amb paquets del 10%).

Finalment, les conclusions de cara al negoci de la botiga de roba son:

- El model funciona molt bé, amb errors petits i prediccions molt fiables en tot el rang de dades disponibles (baix MSE i R2 prop de la unitat). Podem dir que **el desplegament del model en producció millorarà la eficiència i els ingressos** de la botiga.
- La característica més important es 'Length of Membership' i per tant s'han de **dirigir els esforços en mantenir el socis el temps mes gran possible** ja quan mes temps mes gran es la despesa anual que genera. Com suggeriments es podrien crear programes de fidelització a llarg termini amb ofertes exclusives, sistemes de insígnies progressives o privilegis exclusius.