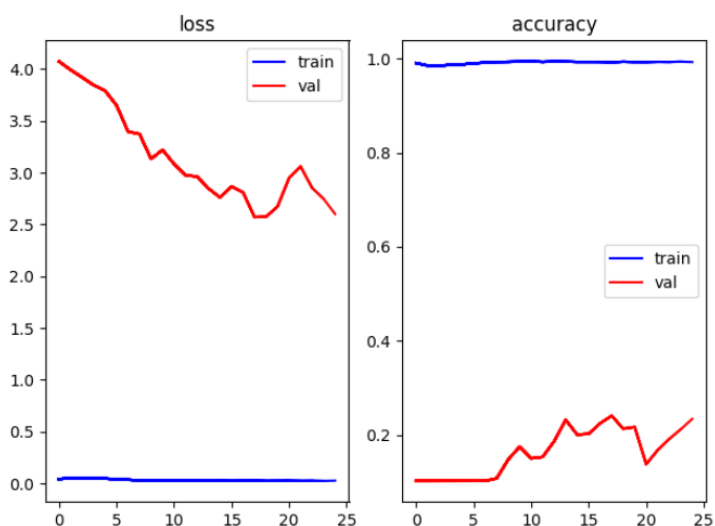


# Progressió Pràctica 2 APC

## Resultats amb el model donat

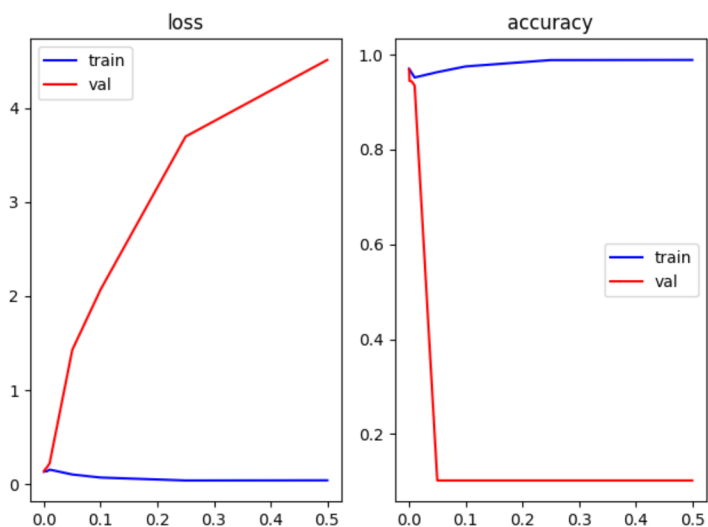
Inicialment tenim un model amb un accuracy molt elevat en el train i molt baix en la validació. Tot i això millora una mica (molt poc a mesura que avancen els epochs)



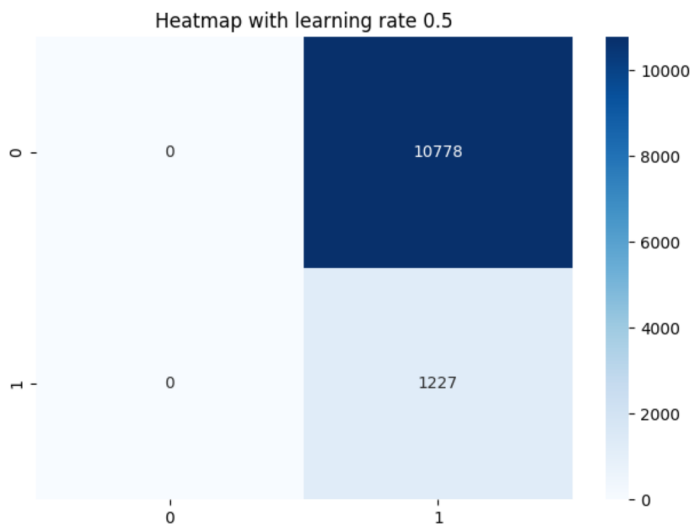
Donat el nostre criteri, considerem que el learning rate té un paper molt important en el rendiment del model. Tot i que la resta de paràmetres són també molt influents, decidim començar l'estudi dels paràmetres per aquest en concret.

## Estudi Learning Rate

S'entrena i es valida el model amb les mateixes dades però canviant el valor del learning\_rate. S'obtenen els següents resultats:

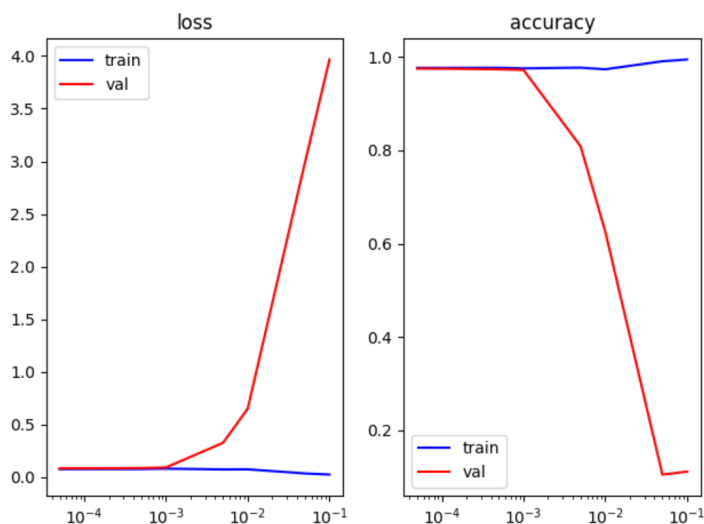


Donat a que els valors a provar eren:  $[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.25, 0.5]$ , no es pot apreciar bé les diferències en els primers valors donada l'escala de l'eix. Tot i això veiem com abans de 0.1 s'equilibra força però amb uns resultats molt baixos. A més hem creat una matriu de confusió per cada learning rate.



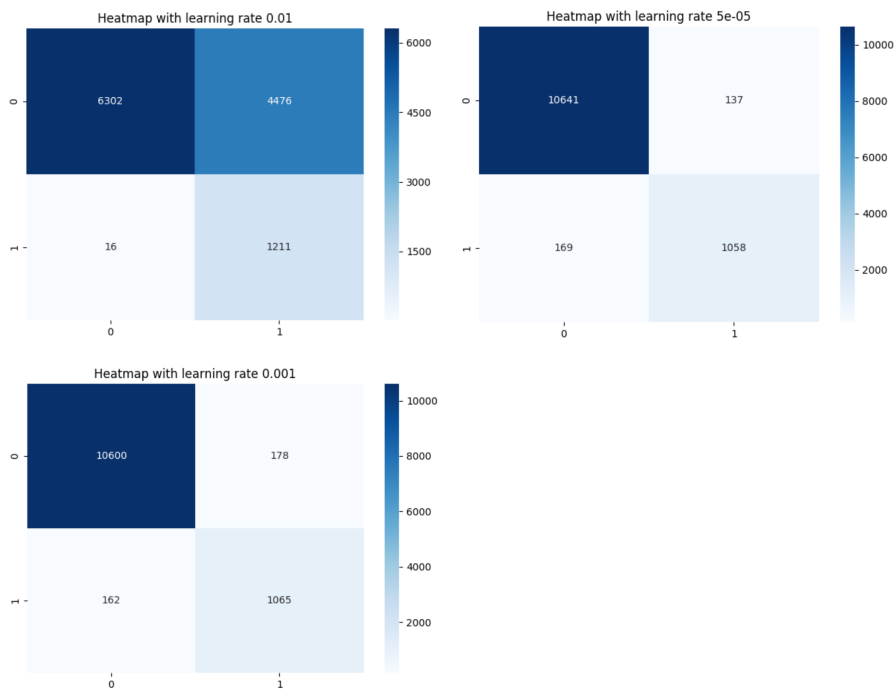
Podem veure com de dolent és el resultat amb valors elevats de  $lr$ .

Redefinim una nova llista eliminant el rang recent estudiat i obtenim els següents resultats.



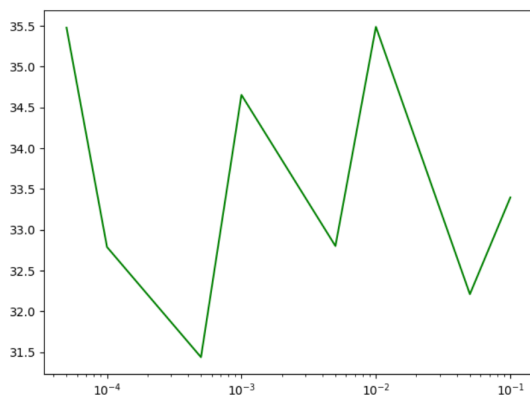
Com podem observar els resultats són molt millors per a valors petits de learning rate (menors de  $10^{-3}$ ). Observem el heatmap de 3 casos:

- El pitjor de l'estudi
- El millor
- Un intermig amb valors molt semblants al millor



És evident que el primer genera molt d'error al classificar si la mostra és de la classe del target o no (classifica un  $\approx 40\%$  de les vegades com que sí que ho és, quan realment no)

Per si de cas, vam voler comprovar el temps d'execució necessitat per cada learning rate, en cas de que per a lr molt petits es necessités més temps, podríem considerar el 3r cas dels que acabem de mostrar com a solució òptima.



Tot i semblar que és força aleatori, realment la diferència no és exageradament gran (varien en un rang de 5s) però en cas de que realment tinguí un motiu, llavors prioritzarem als valors intermitjos de cada ordre ( $5 \times 10^{-k}$ ).

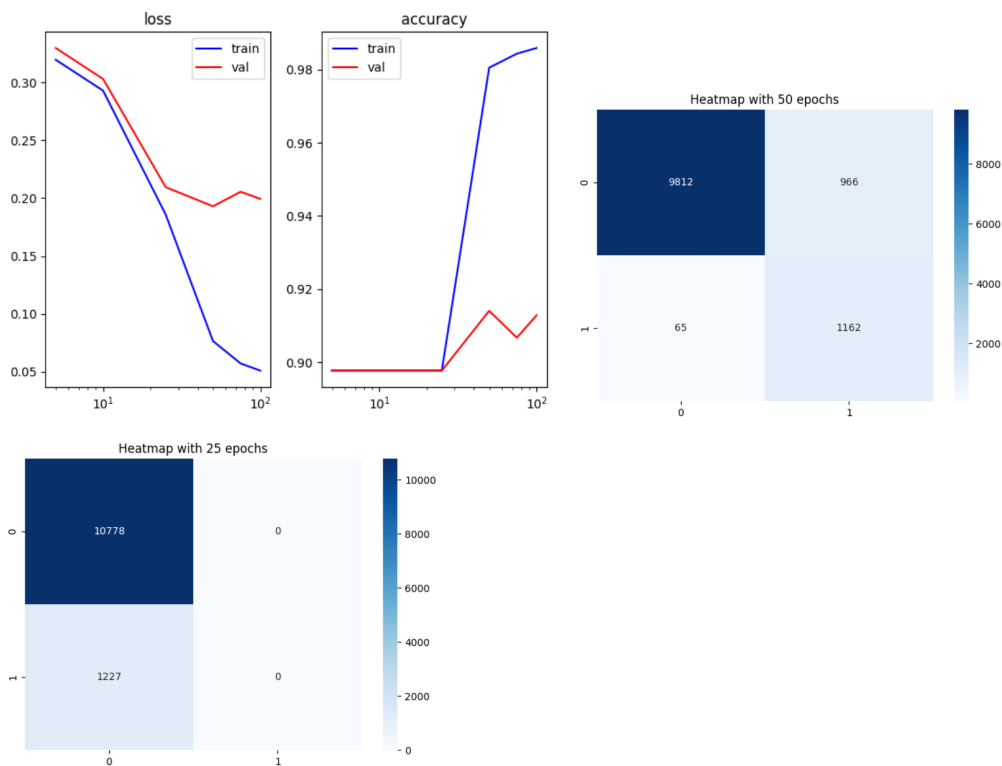
## Anotació

Donades les matrius de confusió podem observar com el model no tendeix al que teníem pensat: classificar-ho tot com a no target donat a que hi ha moltes més mostres negatives que no pas positives -> accuracy fàcilment adaptable a això.

Tot i això, seria interessant estudiar el comportament en funció del recall, doncs cara l'apartat B ens interessa que tinguí una tasa d'encert el màxim d'alta per a les mostres que són el target realment.

## Estudi Epochs

Per a fer aquest estudi utilitzem un  $lr=0.0005$  seguint els resultats del cas anterior. Generem un model per cada valor d'epochs i avaluem el seu resultat. Obtenim el següent:

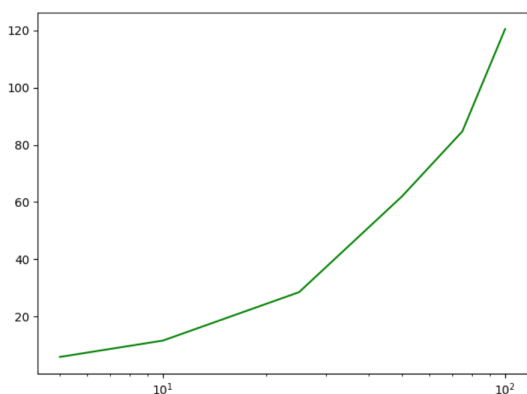


**Dubte:** Com es possible que donat els mateixos valors pels paràmetres que en l'anterior test ens donin resultats tant diferents? (epochs == 25).

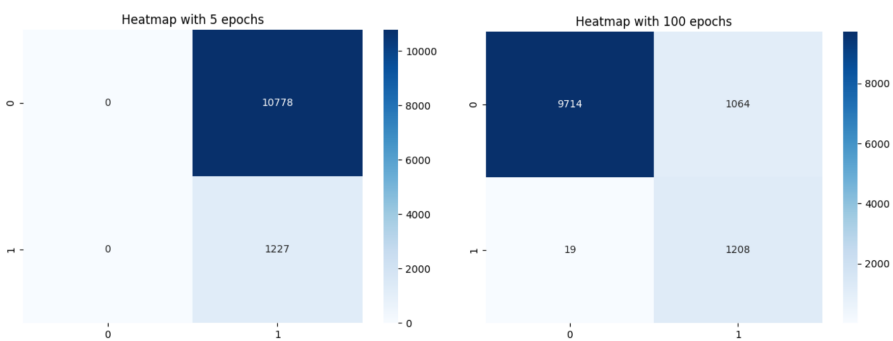
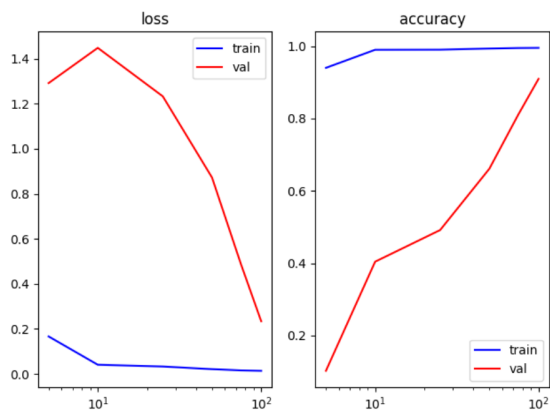
El millor resultat és per epochs = 50 i veiem com per a 25 el model es conforma en classificar-ho tot com a no target (ja ho hem comentat abans els motius i tal).

A partir de 50 epochs sembla que hi ha overfitting tot i que realment tenim bons resultats tant a val com train.

El temps d'execució és evident que augmenta com a més epochs, doncs es fan més cicles.



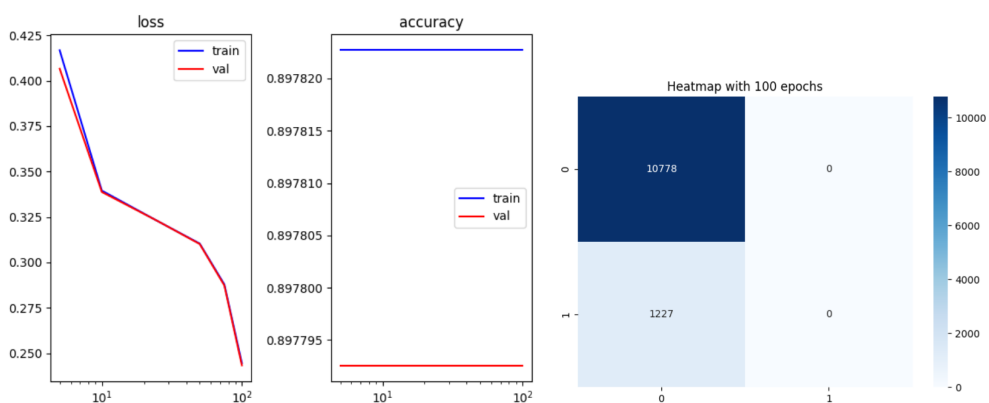
**FEM UN SEGON TEST AMB LEARNING RATE = 0.005**



Inicialment és un model amb molt d'error a diferència del previ (començava directament classificant tot com a no target). A mesura que avança millora en la bona direcció. Tot i això amb 100 epochs no té suficient doncs encara classifica amb un 50% d'error la classe target (que és el que més ens interessa).

### FEM UN TERCER TEST AMB LEARNING RATE = 0.00005

Terrible. Ens trobem amb contradiccions respecte l'estudi anterior:

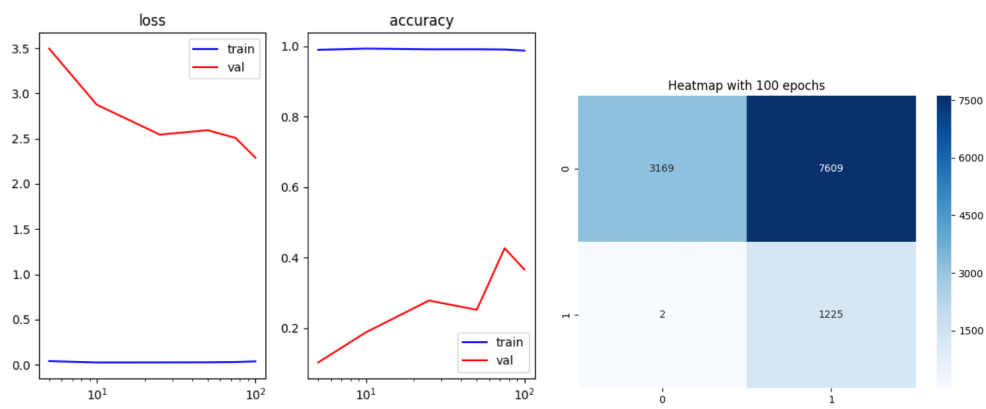


El heatmap és igual per a tots els valor de epoch. Té lògica que sigui d'aquesta manera. Tot i això és estrany que el comportament de lr normal estigui en el cas més alt mentre que té un rendiment molt alt amb el valor de lr intermig.

Caldria confirmar la validesa de l'estudi anterior. En cas que fos valid, analitzar el perquè d'aquesta diferència.

### ULTIM TEST PER SORTIR DE DUBTES AMB LR = 0.05

Genera més confusió doncs ens reafirma que el lr òptim és d'ordre menor. Aquests són els resultats:



Tenim un overfitting de la hostia. Potser s'ha de filar encara més prim amb el learning rate.