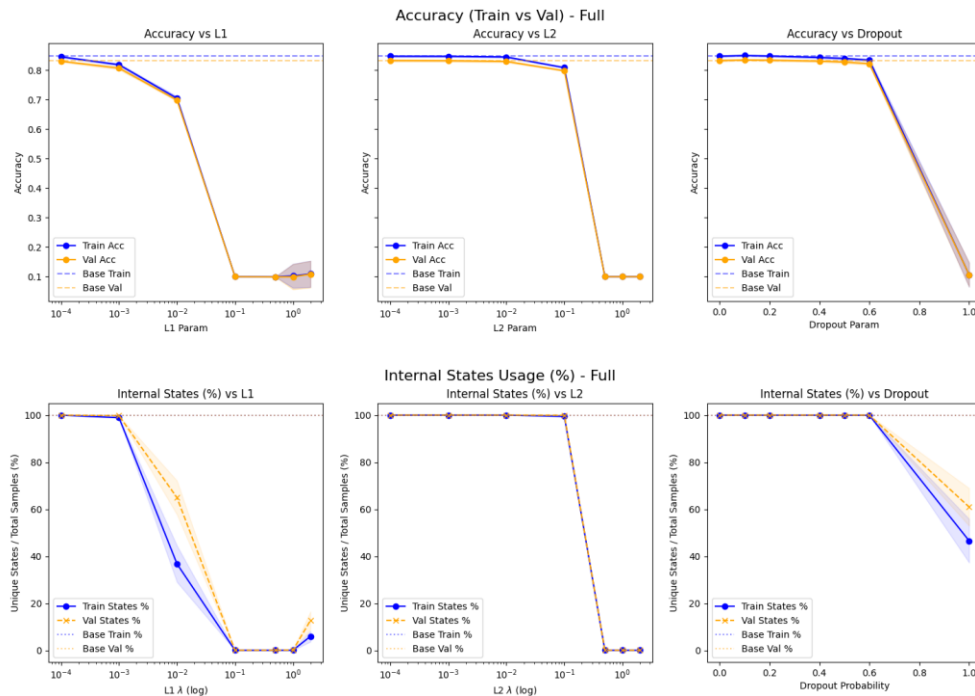
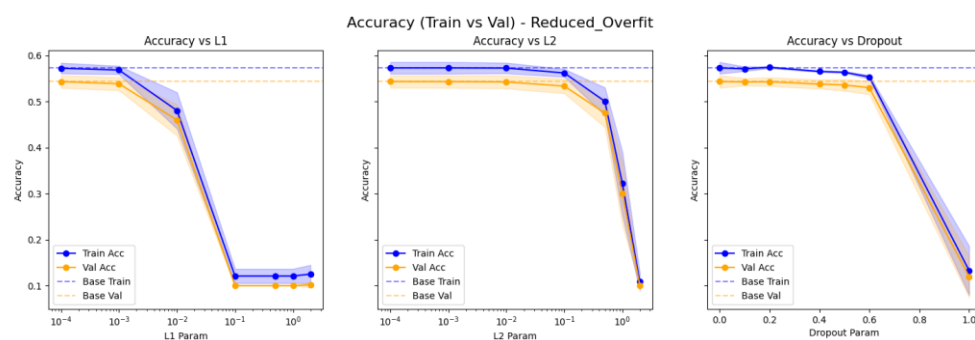


Tras poner que los bin_edges se calcularan simplemente a partir de la red de baseline (la que asumimos que tendrá el histograma más ‘ancho’) se replican bastante bien los resultados de tu notebook.

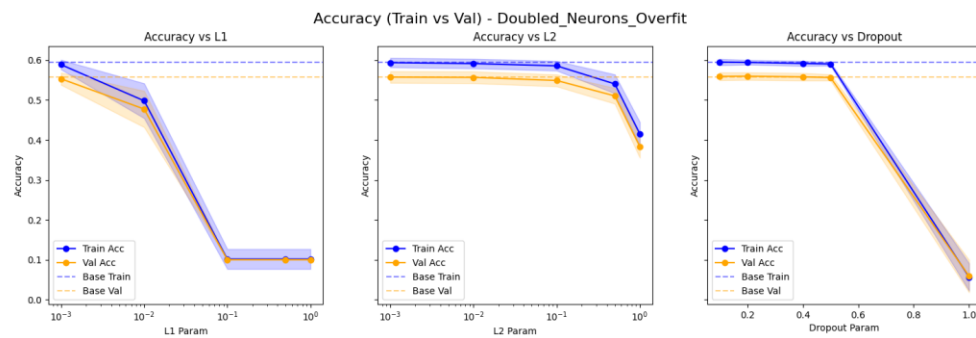


Vemos que la reg L2 es muchísimo más agresiva en la bajada de accuracy y de estados, y que ambas variables parecen estar muy correladas (por ahora más estados => más accuracy).

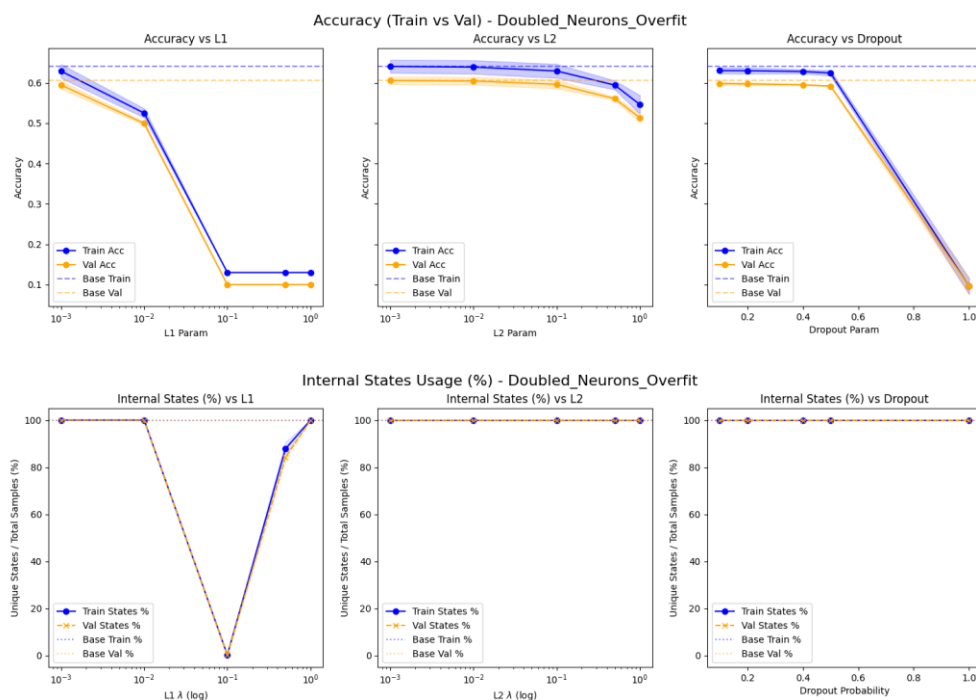
En un intento de overfittear la red, he intentado reducir los datos a 1000 imágenes



Duplicar el número de neuronas (64 => 128) manteniendo los datos a 1000...



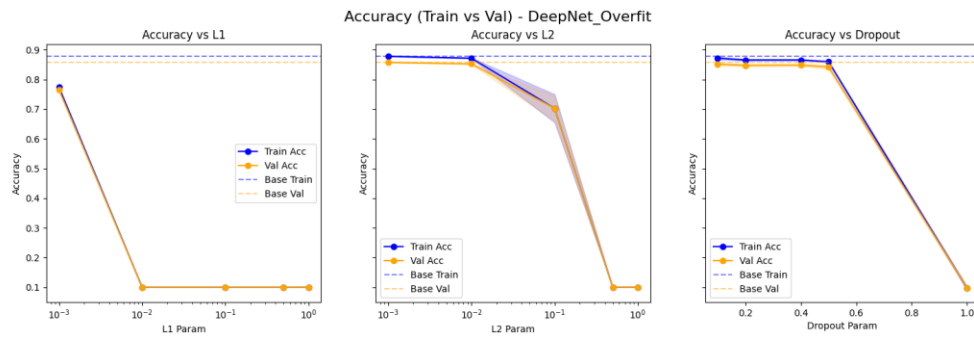
Volver a aumentar las neuronas a 512 con 1000 datos...



(de esta última gráfica de estados concluyo que con el 512 neuronas podría ser lógico ajustar el número de bins en consecuencia porque no es fair comparar “cuánto se agrupan” vectores de 64dims con vectores de 512dims)

Sin mucho éxito sobreajustando, pero sí que se aprecia que la regularización L2 “aguanta más” antes de sufrir mucha pérdida en accuracy. Por último, he introducido 4 capas de 128 neuronas cada una, midiendo las activaciones en la última anterior a la de salida, y de nuevo sin demasiado éxito sobreajustando. Esto lo he probado tanto con 1000 datos como con el dataset completo, pero resulta que tan pocos datos no son suficientes para que una red de ese tamaño aprenda, así que invalida un poco el experimento...

Por otro lado, con estas 4 capas se mejora el rendimiento en baseline hasta casi 90% de accuracy, pero claro, sigo sin ver un caso en el que la regularización mejore la performance...



En cuanto a número de estados, similar

