

心得報告

本次作業內容是預測 Titanic 的生存率，使用了 keras 這個 API，給定兩筆資料，一筆作為測試資料，一筆作為訓練資料。

最開始依照講義提供的方法對資料進行預處理，然後使用 keras 建構 DNN 模型，在建構過程中測試了不同的啟動函數，測試過發現如果使用錯誤的啟動函數會顯著降低準確率，甚至輸出錯誤的結果，relu 和 sigmoid，或是 hard_sigmoid 較為適合。不過對於這邊的相關修改對於提高準確率來說幫助不大。測試發現改變層數修改 neuron 數量也無法顯著提升準確率，測試結果一直在 60%到 74%浮動。修改 loss function 到 logcosh 會降低準確率。優化器因為 adam 是比較常用的，也是相對更好的，就沒有換其他的做實驗。

分析下來，感覺還是在資料預處理上存在問題，測試集有十五個條目，講義中引入了 9 個，像是性別的話，女性生存率大於男性，這些都是顯而易見的，但是仔細觀察後發現，在 boat 欄目下適合生存率是非常相關的，如果上了救生艇就不會死，沒有上的大部分都死了，如果能將此條目也考慮在內的話，準確率應該會顯著上升。

測試後發現只是簡單將 boat 這個也加入預處理中，可以使得準確率上升。

A10515003

鄧鵬宇