

HW3 報告

1. 使用不同線性模型進行迴歸

● Method

Step1：讀入數據集，並將 80%劃分為訓練數據集，20%劃分為測試數據集。

Step2：分別使用 MLP，Lasso, Bayesian ARD, Bayesian Ridge 等線性迴歸模型以及 Random forest regression 對訓練數據集進行擬合(如 Fig 1 所示)。

Step3：對測試數據集進行預測(如 Fig 2 所示)。

● Result

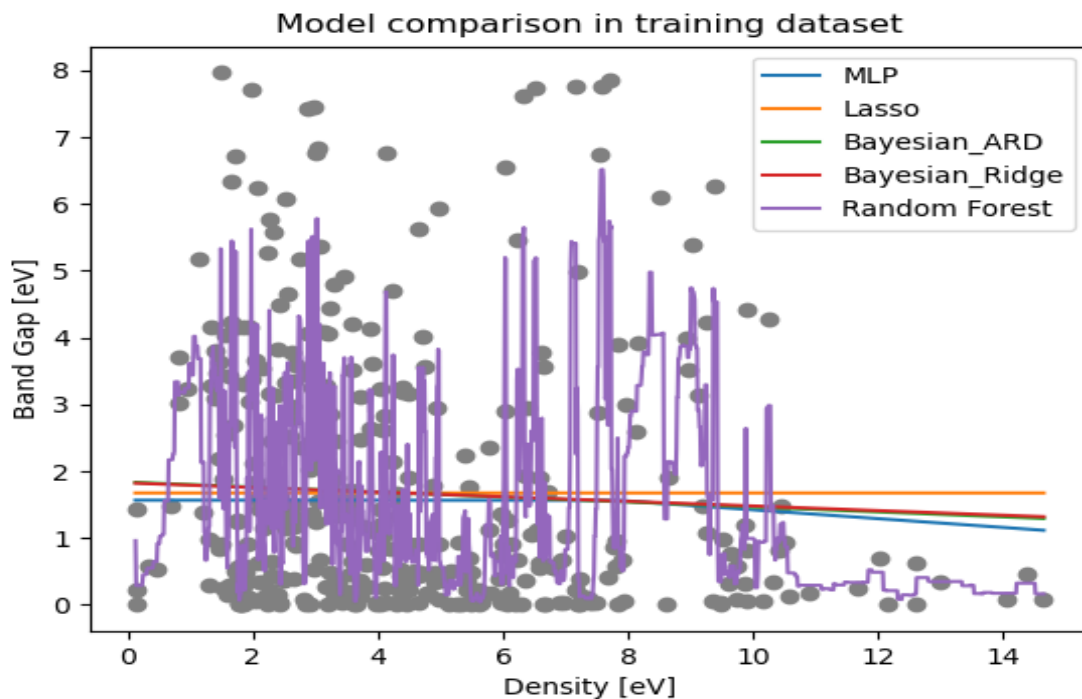


Fig 1. 各種模型對訓練數據集的擬合結果。

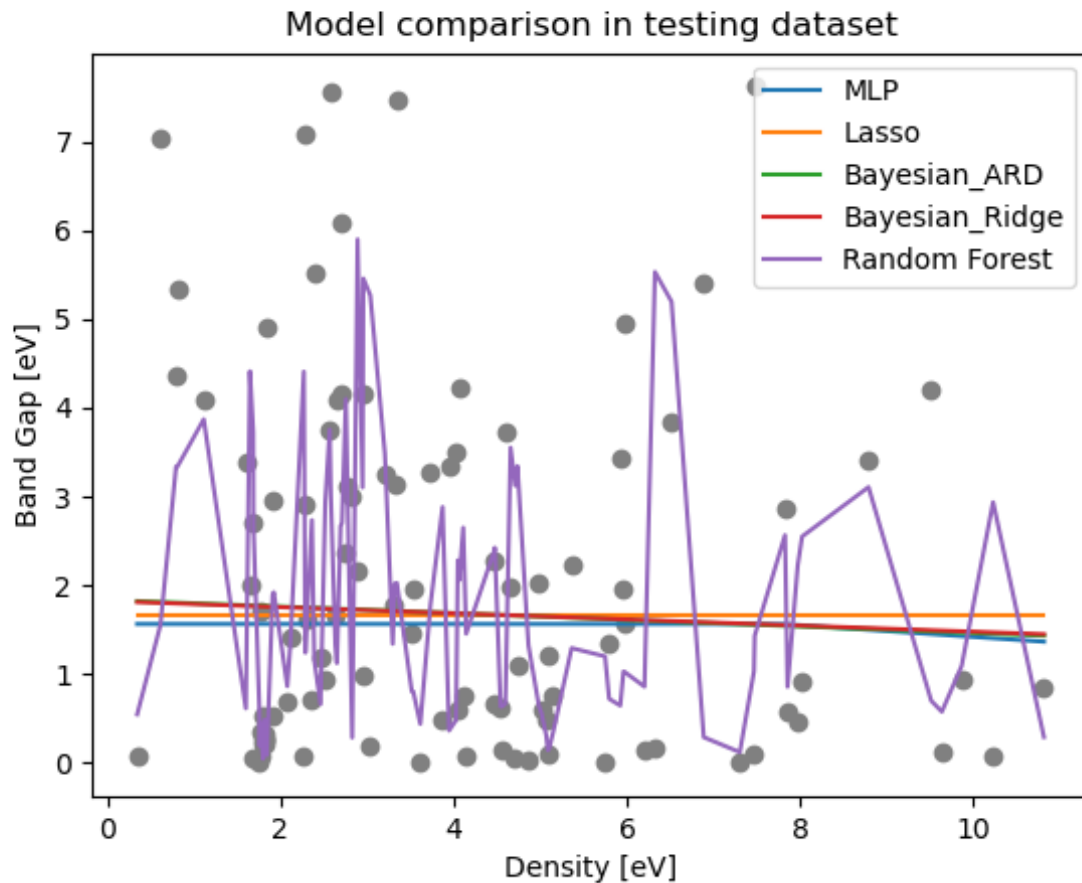


Fig 2. 各種模型對測試數據集的擬合結果。

● Discussion

從數據點的分佈就能很明顯地觀察出，Density 與 Bandgap 之間並非線性關係，所以 MLP, Lasso, Bayesian regression 等線性回歸模型的表現都相當差勁，無法單純地從單一特徵中，擬合出正確的答案。而 Random forest 雖然可以預測出非線性的趨勢，但是因為數據點不夠多(數量與特徵維度都不夠)，所以還是無法完全的擬合出實際的趨勢。

2. 針對 MLP 模型，使用不同的超參數

● Model description

- Step1：讀入數據集，並將 80%劃分為訓練數據集，20%劃分為測試數據集。
- Step2：使用 MLP，並分別使用不同的激活函數()、學習率設置、結構、Optimizer 對訓練數據集進行擬合，並使用測試數據集進行預測(如 Fig 3~Fig 6 所示)。

Result

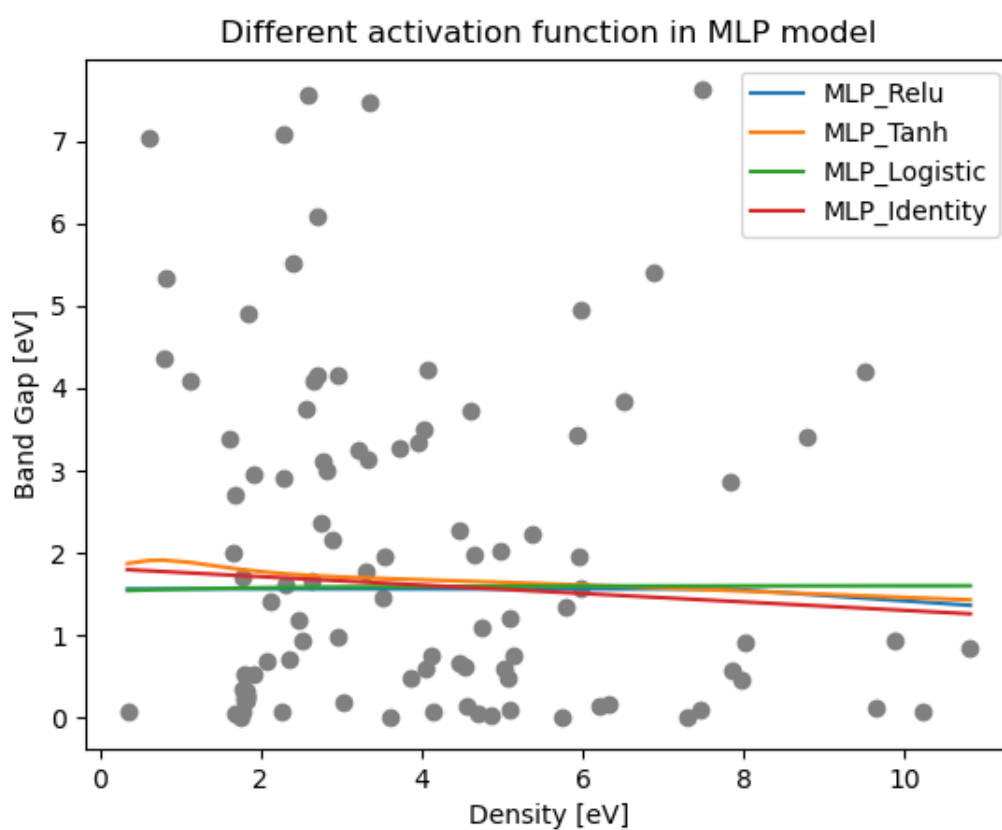


Fig 3. 不同 activation function 在 MLP 模型中的表現

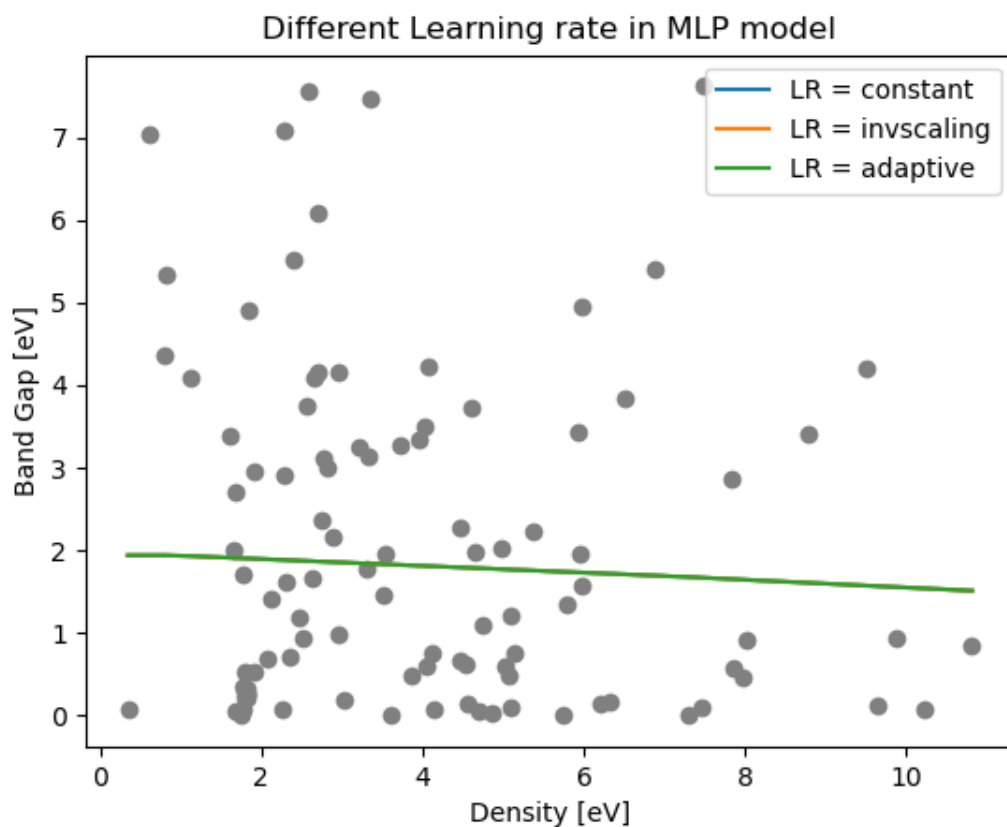


Fig 4. 不同學習率設定在 MLP 模型中的表現

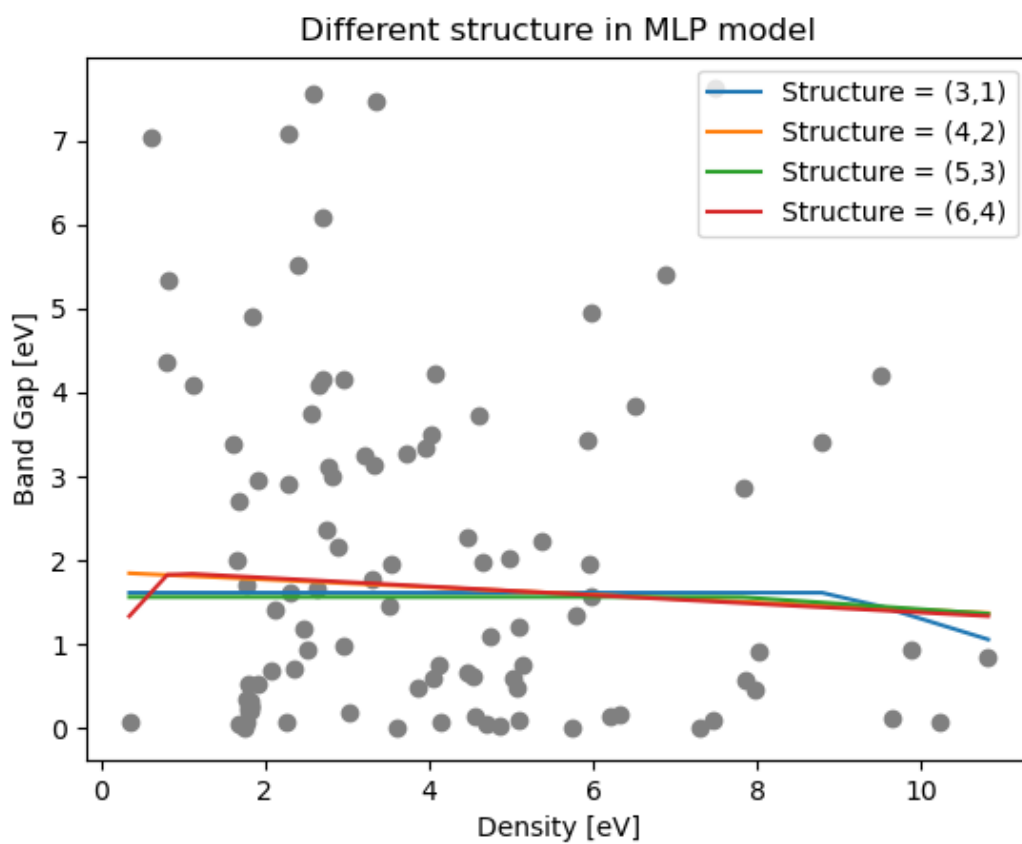


Fig 5. 不同網路架構設定在 MLP 模型中的表現

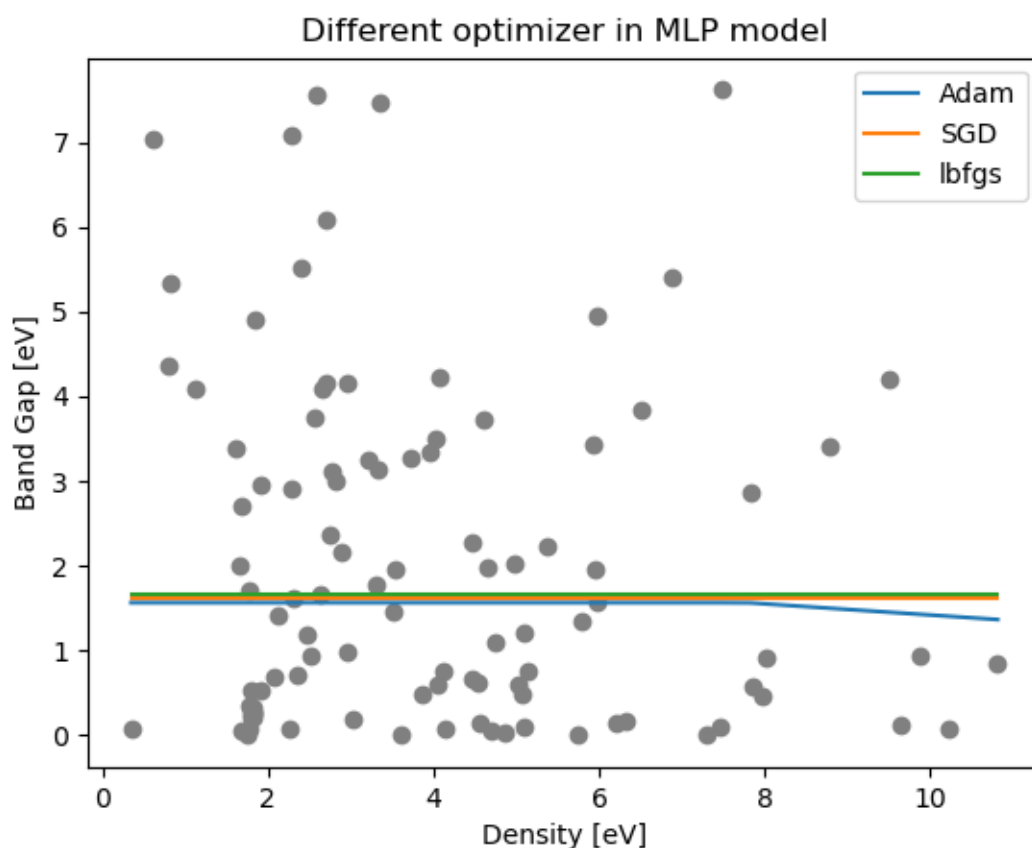


Fig 6. 不同 Optimizer 設定在 MLP 模型中的表現

● Discussion

如同第一部份中討論提到的一樣，這個問題並非一個線性的關係，因此即便使用了不同的超參數對模型進行了調整，模型的擬合能力依然不佳，因此針對題目的要求，合理的選擇模型很重要。除此之外，數據集應該要有更多相關的資訊量在裡面，如果數據只有一維，則模型很難有理想的表現(除非問題是線性的)。