學號:0411509　　姓名:許家維

採用數據: 教授上課時對於汽車研究的 auto.csv

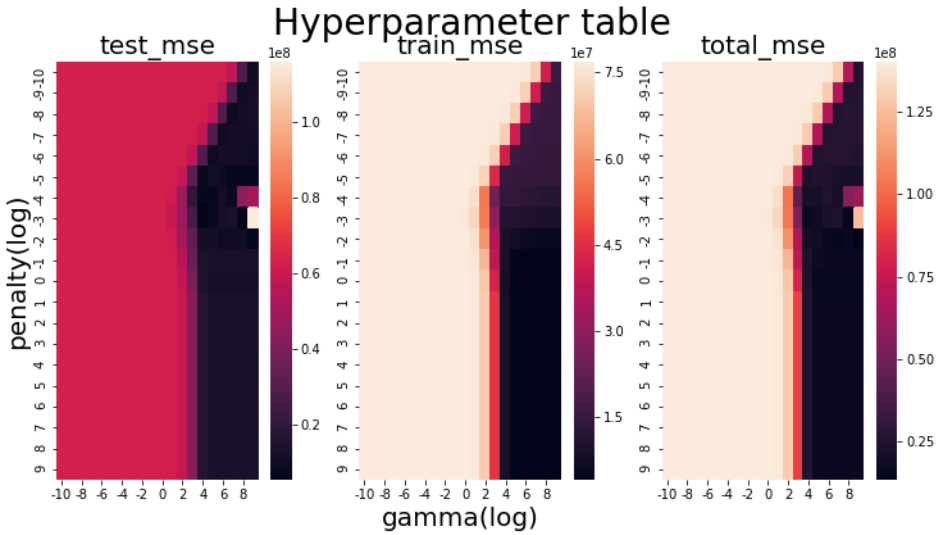
此次作業將回歸分析，引擎大小和價格之間的關係，是否越高檔的車輛須留給引擎較大空間做設計，或是反過來因為是較高檔的車格外注重引擎，進而需要更大空間。

1. linear regression

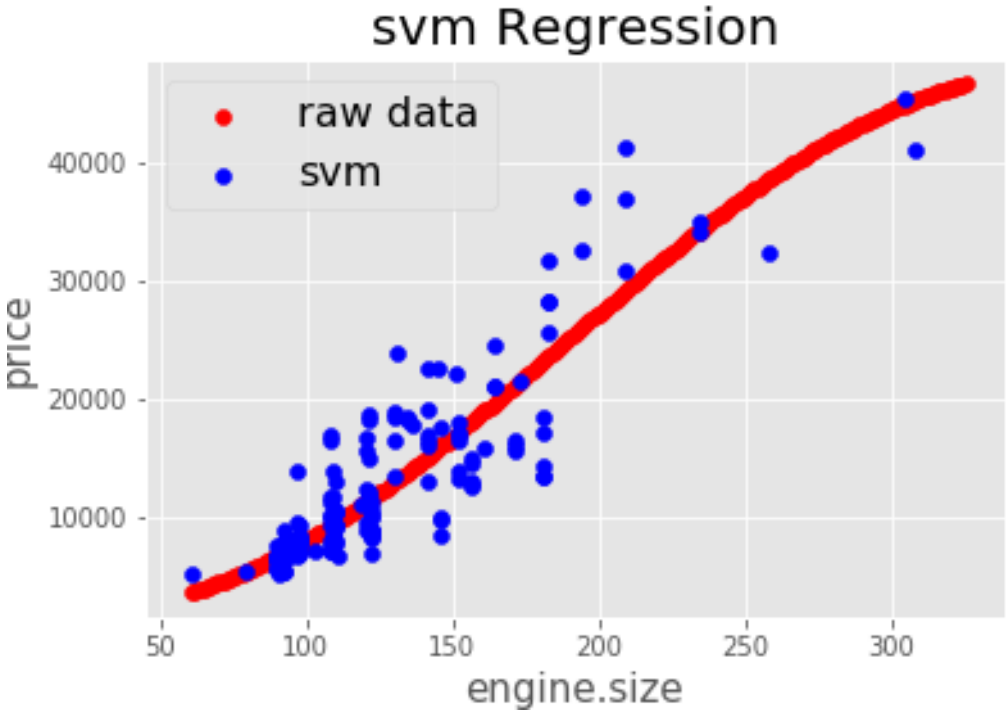
我對此數據使用線性回歸，發現其大致符合 2 著之間的趨勢，不過仍有大量數據並未通過或靠近直線，說明此模型有較大的誤差。

2. support vector regression

我推估可能此模型不能用線性去解釋，所以採用建立超平面去預測母體空間數據狀態，不過因為其有 2 個超參數需手動調整，為逞罰係數、Gamma。所以我先透過建立這 2 個超參數棋譜，並觀察如何設定其 mse(mean square error) 會最小。

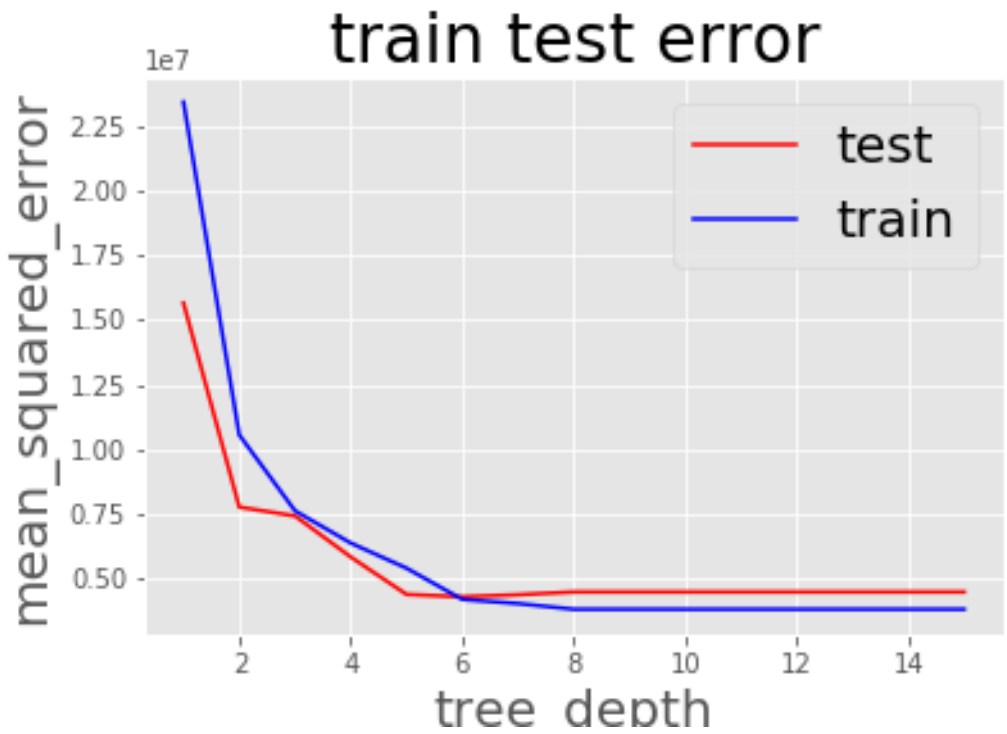


以 10 次方做一次棋盤點，並分別計算 test 和 train 的 error 以確保模型不會 overfitting。以下是 SVR 所預測的曲線。

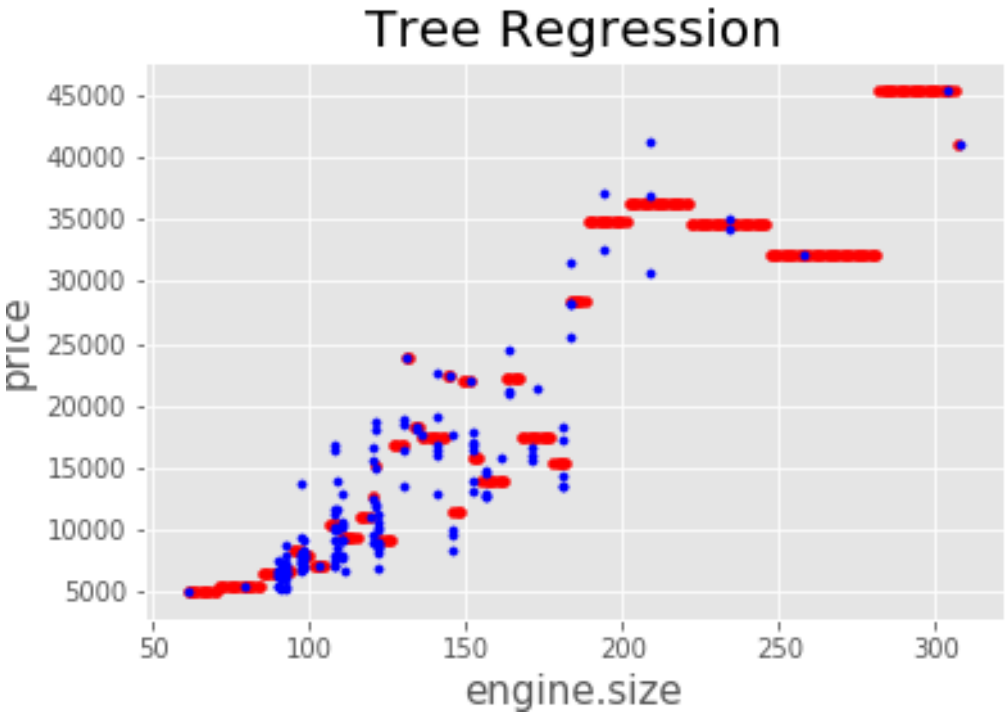


3. tree regression

也同樣面臨著超參數問題，但此方法超參數為樹的深度，很直觀的越深越容易 overfitting，以下是對深度和誤差做的圖。

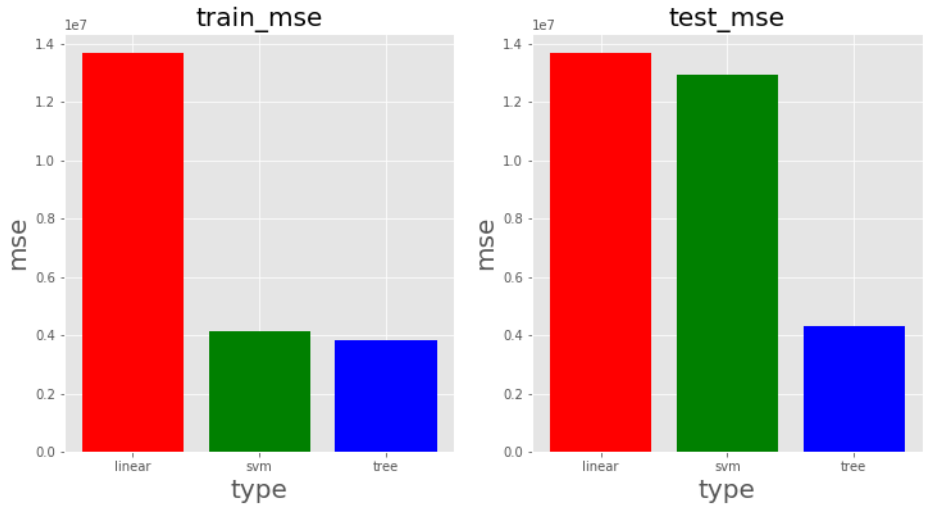


可以看到模型隨深度誤差越少，但到了一定範圍 test 的誤差反而上升，此為過擬和的特徵，所以將樹深訂為 7，進行預測。



總結

對以上 3 種回歸方法進行評估，可以發現論預測能力 svm=tree>linear，但是在預測 test data 時 svr 預測就不太理想，推測可能是數據過於集中在前半部，對大引擎的數據十分稀少，導致其較難預測大引擎車輛的價格。



我觀察到在中階價格的車子，會隨著引擎的大小價格成長幅度較高階和低階大，可以推測因為中階為主流廠商主打車輛性能，造成車價隨引擎大小而有劇烈變動，或是低階引擎到高階引擎中間的過度十分劇烈，造成中間有此段落差，高低階引擎有大量的設計改變進而影響引擎大小。