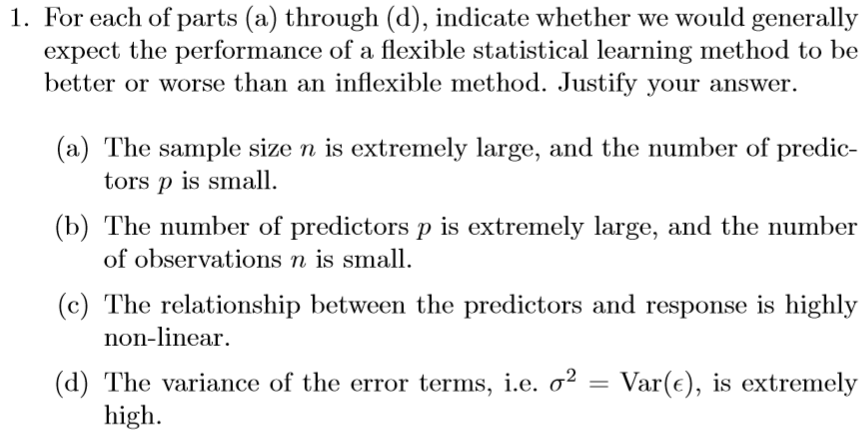
HW2 應數碩一統計組 M072040019 梅瀚中



(a)

Better,當n夠大時,靈活性大的模型比靈活性小的模型表現較好,因為n夠大,可以降低過度擬合的狀況。

(b)

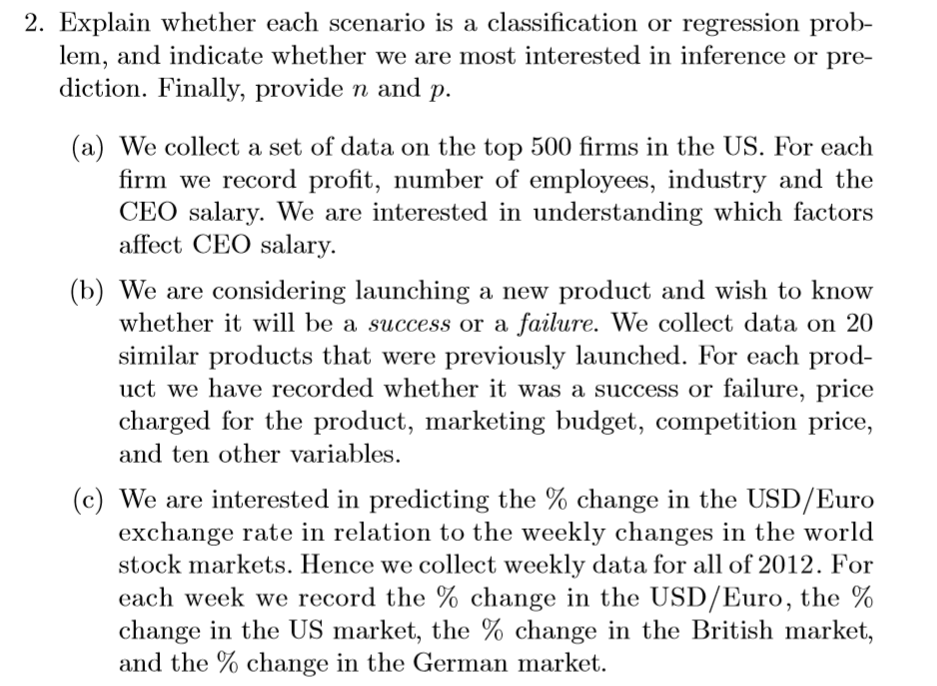
Worse,當n夠小時,如果在使用靈活性大的模型,很容易產生過度擬合的情況,所以使用靈活性較小的模型較適合。

(c)

Better,在高度非線性的狀況下,靈活性較大的模型比較適合,(如果使用靈活性較小的模型容易underfitting)

(d)

Worse,使用靈活性大的模型在此情況容易overfitting,造成模型去配適不能解釋的噪音。



1. regression, inference, n=500,p=3

這是一個regression問題,因為CEO的薪水是一個連續的變數。我們感興趣的是inference,我們想了解自變數(profit,number of employees,industry)是如何影響反應變數(CEO salary)的 。n=500(500firms) p=3(profit,number of employees,industry)

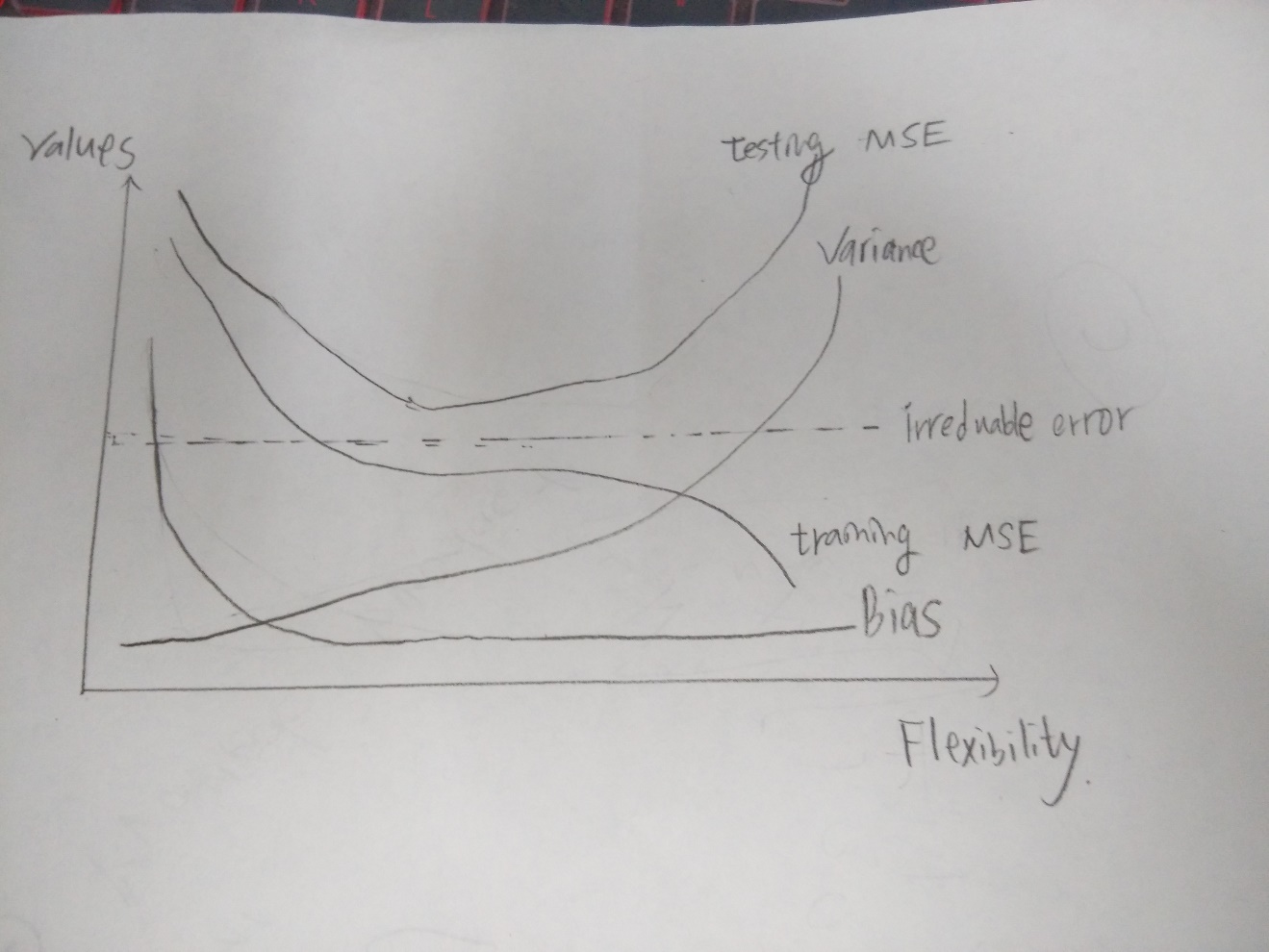
1. classification,prediction ,n=20,p=13

這是一個classification問題,因為我們想把結果分為success和failure。我們感興趣的是prediction,我們已經蒐集過以前的20筆資料,藉由這20筆資料,我們得到的資訊,去預測新的一筆資料是屬於成功或失敗。n=20 (20 similar products previously launched) ,p=13(price charged for the product,marketing budget,competition price,and ten other variables)

1. regression,prediction ,n=52,p=3

這是一個regression問題,因為% change in the USD/Euro dollars是一個連續型的變數。我們感興趣的是**predicting the % change in the US dollar。n=52(365/7=52.14取整數) ,p=3(**% change in US, % change in British, % change in German).

(3)(a)



(b)

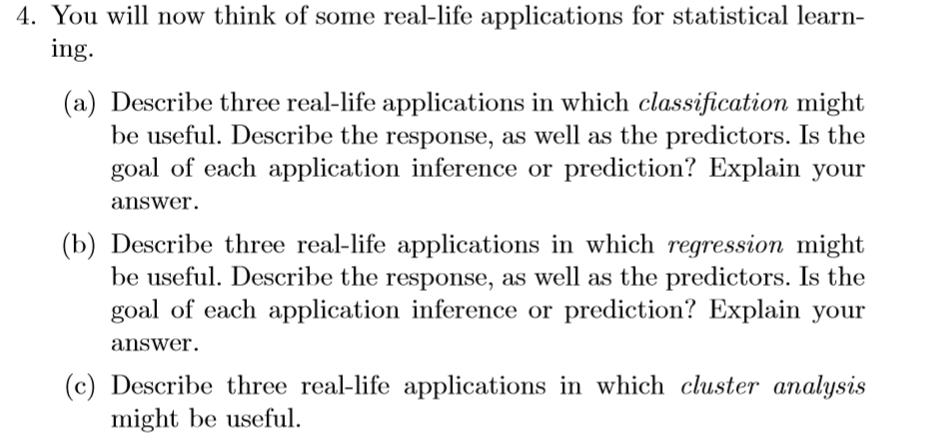
training MSE : training MSE會隨著模型靈活度越高,呈下降趨勢。但過高靈活度的模型,容易造成overfitting的情況,所以單看一個training MSE是一個很小的值,並不代表這是一個好的model

testing MSE: testing MSE 會根據模型靈活度越高,一開始呈現下降趨勢,後隨著模型靈活度過高,造成overfitting狀況,也就是training MSE 過低而testing MSE 過高的狀況。

Bias:隨著模型靈活度越高,代表模型越去描繪data的狀況,造成bias會越低

Variance:隨著模型靈活度越高,variance會越高。

Irreduable error:這是由noise造成的誤差,模型的變化對它不會有任何影響。



(a)

(1)成績及格的分類(>=60)或(<60)。predictor:唸書時間,睡眠時間。Prediction問題

(2)有沒有得肺癌 。 predictor:有無抽菸,有無家人抽菸等。Prediction問題

(3)找工作去面試有沒有上榜。Predictor:為人談吐,專業知識的多寡等。Prediction問題

(b)

(1) 房屋的價錢 ,predictor:坪數,地段等 。inference問題

(2) 一台車可以跑的里程數,predictor:汽缸數,燃料類型等

Inference問題

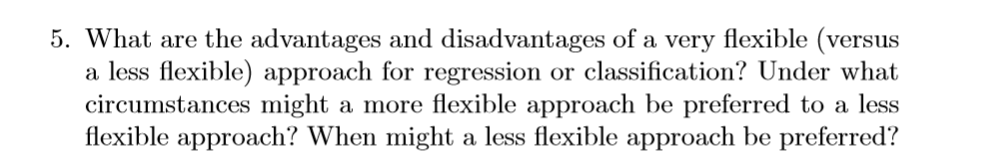
(3)考試成績 ,predictor:唸書時間,睡眠時間,專心程度等 。Inference問題

(c)

(1)將國家分為發達國家，發展中國家和第三世界。

(2)依薪水分為高收入,普通收入,和低收入戶。

(3)把一些相同基因表現的蛋白分為同一群。



Flexible approach:

優點: low bias 。

缺點: (1)high variance ,(2)比較難解釋,(3)可能overfitting。

使用時機: (1)當原本模型underfitting時,可以採用更flexible的方法來調整。(2)當資料呈現高度非線性相關時。

Less flexible approach:

優點: (1)low variance ,(2)解釋性佳。

缺點: (1)high bias ,(2)可能underfitting

使用時機:(1)當原本模型overfitting時,可以採用less flexible 的方法來調整,(2)資料呈現線性關係,(3)需要一個容易解釋的模型的時候。