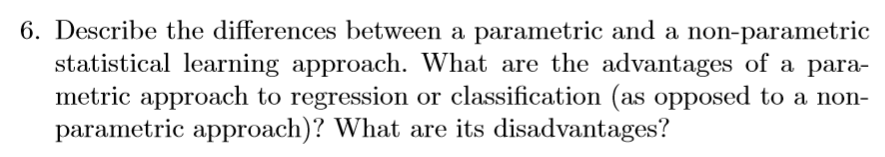
應數一 M072040019 梅瀚中

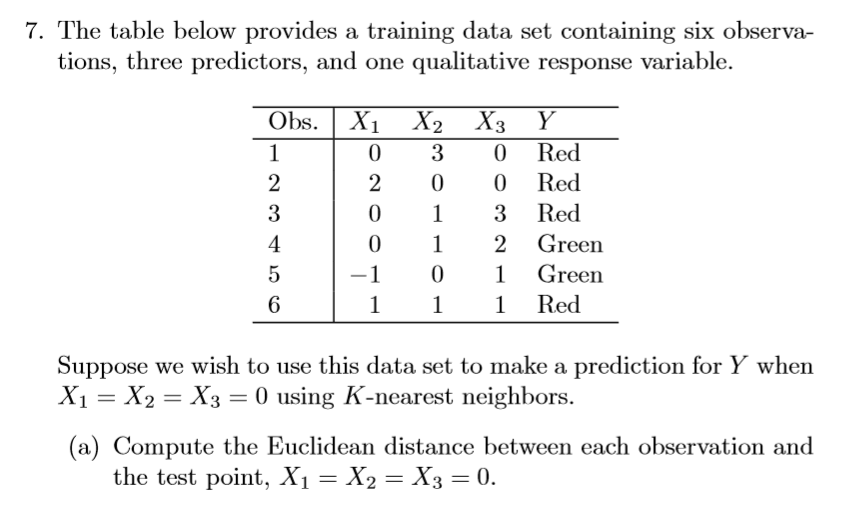


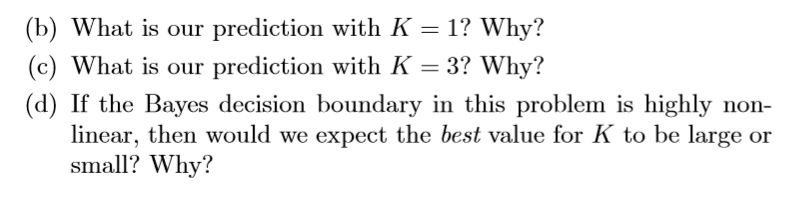
**有母數統計方法:假設母體為特定分配進行檢定。**

**無母數統計方法:所使用的統計量的抽樣分配通常與母體分配無關。**

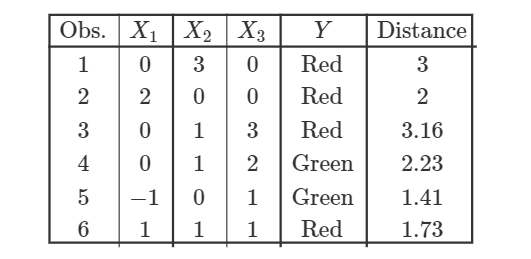
有母數統計方法大大簡化了估計f(x)的問題。拿回歸跟分類問題來說,我們估計β0,β1,...,βp,而不用估計整個f(x)。而缺點是,假如我們選擇的模型如果與f真實的分配不符合,那麼我們的估計就會很差。

無母數統計方法沒有對f(x)作任何的假設,根據data的形式,盡可能去接近數據點(例如:knn), 通過避免f的特定函數形式的假設，它有可能準確地為f提供更寬範圍的可能形狀。無母數方法存在一個主要缺點：由於它們不能減少將f估計為少量參數的問題，因此需要進行大量的觀測（遠遠超過參數方法通常需要的觀測值）。





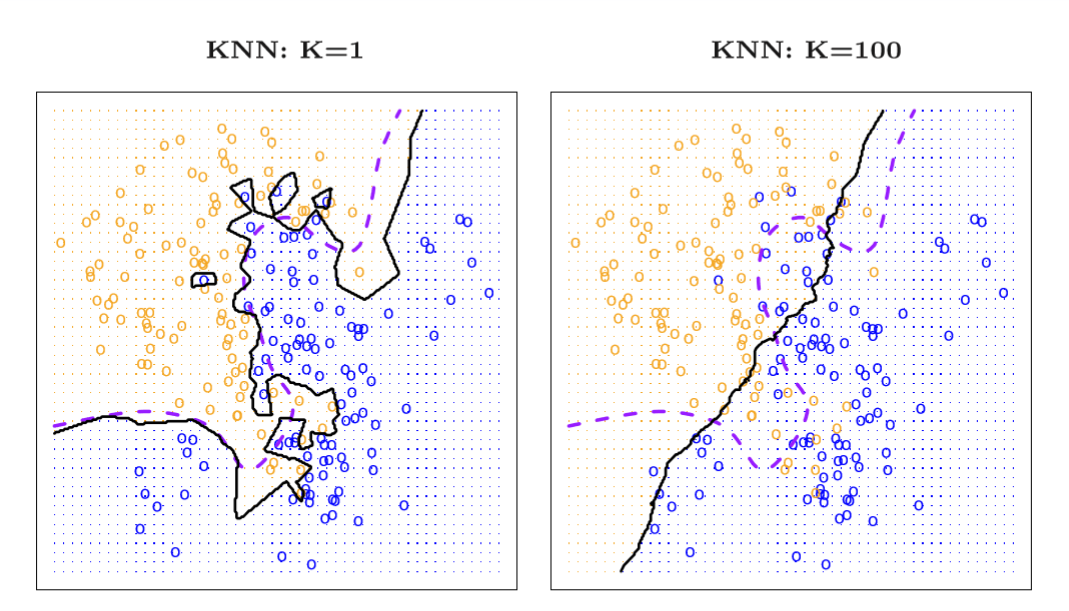
(a)



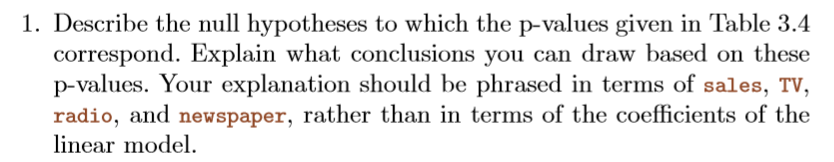
(b) k=1時,我們的Y是Green,因為第5個觀察值離(0,0,0)最近,它是green

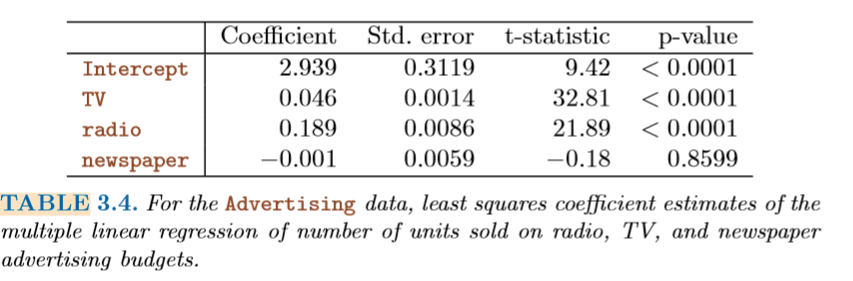
(C)k=3時,我們的Y是Red,因為第5,第6,第2個觀察值離(0,0,0)最近,他們分別是

Green,red,red,然後取顏色多的,所以它是y是red

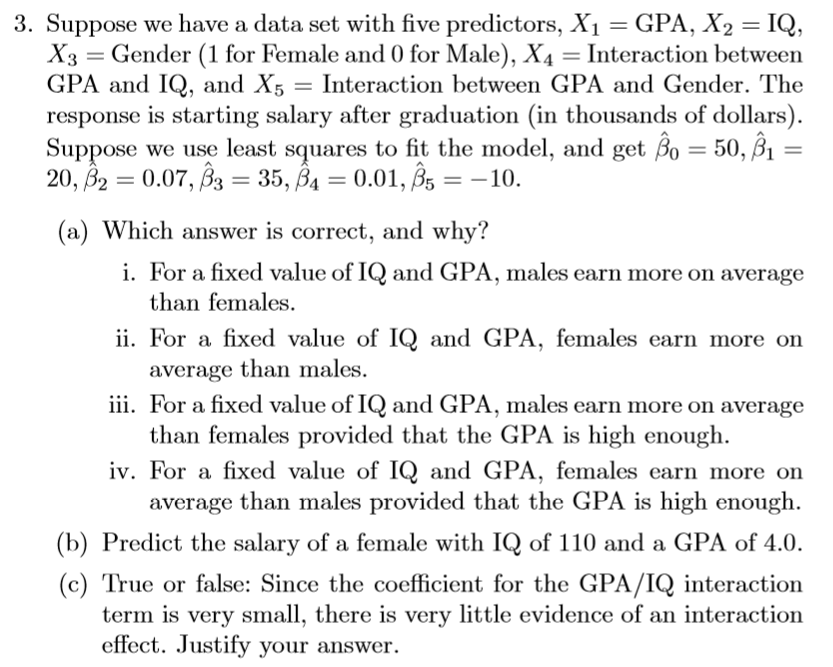
(d)

K值取小的比較好,k值取小的bias會比較小,variance會比較大,適合在非線性的資料。





由圖表可看出:TV,radio,的p-value顯著拒絕H0:βi=0,可認為TV跟radio對於sales是有影響。而newspaper的p-value 不顯著拒絕H0: βi=0,所以我們沒有顯著證據證明newspaper對於sales有影響。





(a)iii.是正確答案

(b) 

(c) false,我們應該去檢定 H0:β4=0 去看P-value才能決定。

1. This question involves the use of simple linear regression on the Auto data set.
2. Use the lm() function to perform a simple linear regression with mpg as the response and horsepower as the predictor. Use the summary() function to print the results. Comment on the output. For example:
   1. Is there a relationship between the predictor and the response?

library(ISLR)

## Warning: package 'ISLR' was built under R version 3.4.4

data(Auto)  
lm.fit=lm(mpg ~ horsepower,data=Auto)  
summary(lm.fit)

##   
## Call:  
## lm(formula = mpg ~ horsepower, data = Auto)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -13.5710 -3.2592 -0.3435 2.7630 16.9240   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 39.935861 0.717499 55.66 <2e-16 \*\*\*  
## horsepower -0.157845 0.006446 -24.49 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.906 on 390 degrees of freedom  
## Multiple R-squared: 0.6059, Adjusted R-squared: 0.6049   
## F-statistic: 599.7 on 1 and 390 DF, p-value: < 2.2e-16

## p-value : <2e-16 ,我們可以顯著認為horsepower跟mpg有關係

1. How strong is the relationship between the predictor and the response?

## R-squared : 0.6059 代表用這個回歸模型mpg可以被horsepower解釋的變異有60.59%

1. Is the relationship between the predictor and the response positive or negative?

## horsepower的係數是-0.157845,所以prefictor和response的關係是負的

1. What is the predicted mpg associated with a horsepower of 98? What are the associated 95% conﬁdence and prediction intervals?

predict(lm.fit, data.frame(horsepower = 98), interval = "prediction")

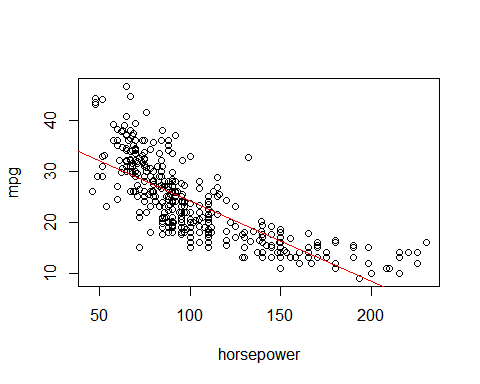
## fit lwr upr  
## 1 24.46708 14.8094 34.12476

predict(lm.fit, data.frame(horsepower = 98), interval = "confidence")

## fit lwr upr  
## 1 24.46708 23.97308 24.96108

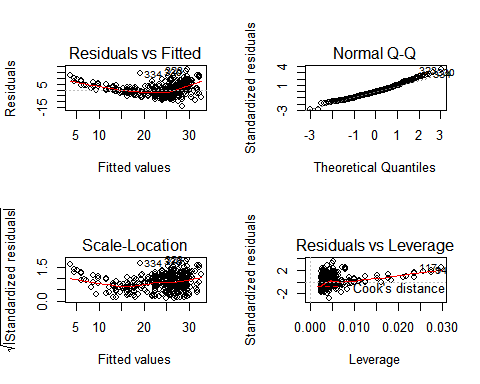
1. Plot the response and the predictor. Use the abline() function to display the least squares regression line.

attach(Auto)  
plot(horsepower,mpg)  
abline(lm.fit,col="red")



1. Use the plot() function to produce diagnostic plots of the least squares regression ﬁt. Comment on any problems you see with the ﬁt

par(mfrow=c(2,2))  
plot(lm.fit)



(1)從Residuals vs Fitted 圖中可以看出predictors和response有一點非線性的趨勢 (2)從Normal Q-Q 圖中得知殘差大致符合標準常態分佈 (3)從Scale-Location圖中大致得知取線周圍的點應該隨機分布,而圖中有一些點有微outliers的趨勢