# PREDICTIVE ANALYSIS PSET 5

ABHILASHA DAS

2026-02-23

## Problem to demonstrate the utility of K nearest neighbour regression over least squares regression.

*QUESTION:*

*Consider a setting with n = 1000 observations. Generate*

(i) $x_{1i}$ from $N(0, 2^2)$ and $x_{2i}$ from $Poisson(\lambda = 1.5)$.

(ii) $\epsilon$ from $N(0,1)$.

(iii) $yi = -2 + 1.4x_{1i} - 2.6x_{2i} + \epsilon_i$.

Split the data into train and test sets. Keep the first 800 observations as training data and the remaining as test data. Work out the following:

*1. Fit a multiple linear regression equation of y on x1 and x2. Calculate test MSE.*

*2. Fit a KNN model with k = 1, 2, 5, 9, 15. Calculate test MSE for each choice of k.*

```
rm(list=ls())
set.seed(123)
n=1000
x1i=rnorm(n, mean=0,sd=2)
x2i=rpois(n, lambda=1.5)
error=rnorm(n, mean=0,sd=1)
yi=-2+1.4*x1i-2.6*x2i+error
data=data.frame(x1i,x2i,yi)
train_data=data[1:800,]
test_data=data[801:1000,]
# (a) fitting the multiple linear regression model and calculating Test MSE
model1=lm(yi~x1i+x2i,data=train_data);summary(model1)
```

```
##
## Call:
## lm(formula = yi ~ x1i + x2i, data = train_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0727 -0.6573 -0.0125  0.6921  3.2412
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.07300    0.05382  -38.52   <2e-16 ***
## x1i          1.38207    0.01767   78.21   <2e-16 ***
## x2i         -2.55584    0.02768  -92.34   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.98 on 797 degrees of freedom
## Multiple R-squared:  0.9492, Adjusted R-squared:  0.9491
## F-statistic:  7445 on 2 and 797 DF,  p-value: < 2.2e-16
```

```
pred_model1=predict(model1,newdata=test_data);summary(pred_model1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -17.062  -8.170  -6.146  -5.817  -2.822   4.283
```

```
mse.model1=mean((pred_model1-test_data$yi)^2);mse.model1
```

```
## [1] 0.998901
```

```
# (b) Fit a KNN model with k = 1, 2, 5, 9, 15. Calculate test MSE for each choice of k.
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.3.3
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
## Warning: package 'lattice' was built under R version 4.3.3
```

```
set.seed(123)
knnreg(train_data[,c("x1i","x2i")],train_data$yi,k=1)
```

```
## 1-nearest neighbor regression model
```

```
tr_kr_1=knnreg((yi)~(x1i)+(x2i),data=train_data,k=1);tr_kr_1
```

```
## 1-nearest neighbor regression model
```

```
y_knn_1=predict(tr_kr_1,test_data);y_knn_1
```

```
##   [1]  -1.0725745  -5.1325797  -2.5042980   2.7725510  -7.4898445  -8.1285531
##   [7]  -5.3825552  -3.8358912  -7.4898445 -12.0885802 -16.9638395  -7.3854260
##  [13]  -7.2011077 -10.2158240  -7.9027672   0.7147352  -1.9267410 -11.6494040
##  [19]  -7.1538253 -10.8628888  -8.5788138  -1.8820118  -7.2011077  -6.8790069
##  [25]  -3.2097680  -5.1918238  -9.6640328  -7.6078560 -10.9279951  -8.2150344
##  [31]  -3.7392931  -7.4896530  -5.9685300  -5.1865593  -4.8320630  -2.6730267
##  [37] -13.9132236  -7.1651206  -7.4711129  -8.3252663  -7.5893987   0.8009464
##  [43] -12.8247863  -9.4253106  -5.7275124  -6.8887513 -10.3007472  -1.1966388
##  [49]  -0.2951467  -2.8382763 -11.6833936  -5.4816812  -3.9998264  -5.3083824
##  [55]   1.5856767  -6.2240204  -7.3206924 -10.0586200  -6.1604504  -1.5791691
##  [61]  -4.2156130  -1.6801278  -5.4240292  -5.7624011  -7.9293459   1.5856767
##  [67] -12.8189454  -7.0242011  -2.8740306   0.6721274 -11.9375115  -1.6801278
##  [73]  -3.6876892   0.6817057  -2.5031784  -5.2595513  -4.5569394   0.8346024
##  [79]  -1.4003417 -12.0885802  -2.8010883 -11.5749117  -9.4276758  -5.5190725
##  [85]  -2.9380761  -5.9737620  -8.8902894  -3.7582621  -9.0173765 -10.0563284
```

```
##  [91]  -4.1799859 -11.6833936  -3.6876892   3.5395996  -0.2802275  -5.4337991
##  [97]  -3.9404071  -8.8473458  -1.9626871  -1.7833641 -15.5055245  -9.0173765
## [103]  -6.3808292   0.6721274  -2.9517094  -9.2039954  -9.4654765  -4.8140494
## [109]  -2.2491534  -6.5125027   1.5856767  -7.7405564  -7.3085192  -9.1001663
## [115]  -8.8067304  -2.9517094  -3.0687264 -12.1298802  -5.8564892  -4.1457314
## [121]  -1.3697276  -8.9042098  -7.5355261  -1.9626871  -6.2059846   1.6059628
## [127]  -6.7111441 -11.9356976   1.5856767  -4.2438793  -3.9115969  -8.2989025
## [133]  -3.4584170  -3.1184766  -7.9245577  -7.1978205 -10.7114154  -7.6846075
## [139]  -3.1184766  -0.5803873  -1.1966388 -10.0586200  -5.6805352  -4.3761955
## [145]  -3.6537526 -11.1692652  -5.1325797  -1.1021024  -2.6319909  -8.7754942
## [151]  -2.8990563 -11.2912782  -7.3336448  -1.7833641  -5.1503762 -16.3800959
## [157]  -1.7833641  -6.3808292  -8.2150344  -0.3484242  -6.8790069   5.1984181
## [163]  -8.1285531  -4.5113204  -8.1865906  -3.9668678 -13.9132236  -5.6805352
## [169]  -2.2417933  -7.2517184  -7.4711129 -10.7616407  -9.3250600  -3.6876892
## [175]  -2.3763939  -9.6589816 -15.6910193  -3.9668678 -10.0586200  -1.6801278
## [181]   1.5856767  -7.8965176  -4.9086049  -2.6910624   0.9473857 -12.1251513
## [187] -15.7645141  -4.5569394  -5.4422760  -1.5791691  -5.6510933  -8.9943675
## [193]  -5.9737620  -4.8066443  -9.4400574  -2.2417933  -7.9162980  -8.2989025
## [199]  -5.7551080  -6.2988375
```

```
mse_knn_1=mean((test_data$yi-y_knn_1)^2);mse_knn_1
```

```
## [1] 2.219793
```

```
tr_kr_2=knnreg((yi)~(x1i)+(x2i),data=train_data,k=2);tr_kr_2
```

```
## 2-nearest neighbor regression model
```

```
y_knn_2=predict(tr_kr_2,test_data);y_knn_2
```

```
##   [1]  -1.34780400  -5.77126666  -1.95141755   2.05804179  -7.00117360
##   [6]  -8.01641722  -5.16419028  -3.91320939  -7.00117360 -11.83766296
##  [11] -16.11786708  -8.70435419  -6.53486850 -10.17214622  -8.11401675
##  [16]   0.87690426  -2.32734135 -10.48591978  -6.74685198  -9.95089685
##  [21]  -8.19482569  -1.57990627  -6.53486850  -6.74544049  -2.92087947
##  [26]  -6.22177110  -8.60003697  -7.68304760 -10.89544191  -8.59381036
##  [31]  -4.30294206  -8.12020719  -4.64515132  -5.21640287  -4.88295458
##  [36]  -3.95498528 -13.89319204  -6.49348374  -8.44821175  -8.88266181
##  [41]  -7.22877675   1.14096388 -11.51890599  -8.74629924  -5.96674850
##  [46]  -7.21213866  -9.79463906  -2.59823261   0.02093518  -2.59595610
##  [51] -11.49158749  -5.80466424  -2.59823261  -5.48421027   0.92670825
##  [56]  -5.70792207  -8.14663935 -10.43346042  -4.69397548  -2.86245022
##  [61]  -4.58240897  -1.09899110  -5.89899983  -5.83064239  -7.16141104
##  [66]   0.92670825 -12.32750457  -7.26679964  -3.46573640   0.18849037
##  [71] -11.43275328  -1.09899110  -4.60616567   0.42518133  -3.27477052
##  [76]  -5.32376763  -3.88996277   0.79172894  -1.20803185 -11.83766296
##  [81]  -2.87456061 -10.66016492  -9.62633214  -4.28843379  -3.33161070
##  [86]  -8.01504519 -10.09078380  -4.32651265  -8.69285464  -9.73989160
##  [91]  -4.45822813 -11.49158749  -4.60616567   2.54815845  -0.76387468
##  [96]  -5.76532576  -5.85822796  -7.50588785  -3.52978848  -2.40427072
```

```
## [101] -15.26500565  -9.26975277  -6.93533685   0.18849037  -3.34550124
## [106]  -9.75237134  -8.17224171  -4.14436326  -2.27234683  -7.00117360
## [111]   0.92670825  -7.87830862  -7.10922178  -9.75006590  -8.98686443
## [116]  -3.51139860  -2.78595239 -12.40896785  -5.26195917  -2.86245022
## [121]  -1.71967803  -9.52919757  -7.21213866  -3.55653166  -5.92853891
## [126]   2.19310424  -7.65223969 -11.65287369   0.92670825  -5.05018422
## [131]  -3.59214743  -7.99994301  -3.87609777  -3.03511541  -7.40665446
## [136]  -6.57286416 -10.07542518  -7.78832493  -3.42705262  -0.38564898
## [141]  -1.49477574 -10.43346042  -6.26389634  -3.04759109  -3.37444295
## [146] -11.26965749  -5.77126666  -1.93020965  -2.80870418  -8.29316590
## [151]  -2.55048632 -10.51834820  -7.92873352  -2.40427072  -3.55653166
## [156] -15.28279960  -2.40427072  -6.18476475  -8.22017685  -0.27570952
## [161]  -6.74544049   5.03812145  -6.99165637  -4.36241852  -7.74755489
## [166]  -3.94660370 -13.89319204  -6.26389634  -2.67521327  -7.22476947
## [171]  -7.69260738 -10.72098887  -9.94970356  -4.60616567  -1.67639735
## [176]  -7.92817179 -16.27124113  -3.65851858 -10.28886046  -2.72113667
## [181]   0.92670825  -7.94146003  -4.84042946  -3.09423127   1.49887028
## [186] -12.34362454 -15.54546291  -3.88996277  -5.18725696  -1.76797347
## [191]  -4.96692131  -7.59586472  -8.01504519  -5.56665351  -8.88266181
## [196]  -2.67521327  -7.49194698  -7.99994301  -6.09040431  -7.48458770
```

```
mse_knn_2=mean((test_data$yi-y_knn_2)^2);mse_knn_2
```

```
## [1] 1.729587
```

```
tr_kr_5=knnreg((yi)~(x1i)+(x2i),data=train_data,k=5);tr_kr_5
```

```
## 5-nearest neighbor regression model
```

```
y_knn_5=predict(tr_kr_5,test_data);y_knn_5
```

```
##   [1]  -1.32924318  -5.83988196  -2.03671350   1.26813575  -6.41992647
##   [6]  -7.09914548  -4.57080406  -3.79686396  -6.41992647 -12.53612132
##  [11] -15.09988855  -9.22414997  -6.95353851  -9.31921766  -8.64739436
##  [16]   0.28602103  -2.76235910  -9.55516739  -6.76969227 -10.24109668
##  [21]  -8.28009736  -2.63669810  -6.95353851  -7.82032819  -2.73556168
##  [26]  -6.36265818  -7.78579463  -7.25570278 -10.42817044  -8.66159590
##  [31]  -4.41077002  -8.74085247  -5.27987827  -6.13285718  -5.04031410
##  [36]  -3.13577529 -14.45237335  -6.49118642  -7.91382448  -8.64739436
##  [41]  -6.87824998   2.76760039 -10.80667221  -8.09176423  -5.51959832
##  [46]  -6.98175469  -9.39867568  -1.89844723   0.32594588  -2.07345506
##  [51] -10.81575060  -5.00448297  -1.97364906  -5.12460558  -0.47482825
##  [56]  -6.39832161  -8.88154404 -10.37249398  -3.95741937  -2.39972909
##  [61]  -4.45852614  -2.20966275  -6.46755721  -6.12039467  -8.39491051
##  [66]   0.75364116 -12.38622852  -7.05379841  -3.63004118   0.32594588
##  [71] -10.33151793  -2.20966275  -4.72452088   0.18536347  -3.56935632
##  [76]  -4.72435294  -3.05039296   0.24159299  -0.60624026 -12.53612132
##  [81]  -3.02417170  -9.91576158 -10.01707189  -3.59784915  -3.24135399
##  [86]  -8.68545481  -9.61286178  -3.86330167  -9.35837740  -8.64377376
##  [91]  -4.51393715 -10.81575060  -4.97904553   2.89849802  -0.84931507
```

```
##  [96]  -6.86253571  -6.25698128  -7.82270161  -3.32968143  -2.02111583
## [101] -15.17863197  -8.71797433  -6.24629405   0.32594588  -4.05080403
## [106]  -8.99507144  -7.72531466  -4.73604280  -3.35864646  -6.54095995
## [111]   0.75364116  -8.18435256  -7.24665321  -9.03783234  -8.54057541
## [116]  -4.05080403  -3.56935632 -12.38622852  -5.36880761  -2.84624387
## [121]  -1.87044212  -8.39138329  -7.08371439  -3.85626162  -5.38015329
## [126]   1.53448158  -6.95353851 -12.26706328   0.37303141  -4.94458174
## [131]  -3.94972658  -6.46755721  -4.05845595  -3.12494078  -7.39084920
## [136]  -6.55332285  -9.09735177  -6.51457455  -2.96736174  -0.25478079
## [141]  -1.89844723 -10.37249398  -6.45389466  -3.92724901  -2.88886586
## [146] -12.08810683  -5.83988196  -0.81730145  -3.13007776  -8.18435256
## [151]  -2.63354816  -9.61286178  -8.55544490  -2.02111583  -4.21310575
## [156] -14.36303720  -2.22202281  -6.34137175  -8.66159590  -0.08112942
## [161]  -7.59085007   4.82517155  -6.80440758  -4.41690553  -7.97952653
## [166]  -3.30012669 -14.45237335  -5.97730333  -2.53097969  -6.55332285
## [171]  -7.91382448 -10.07326447  -8.88154404  -4.72452088  -1.59299091
## [176]  -8.14672622 -15.58279669  -2.98755761 -10.31301681  -2.20966275
## [181]  -0.47482825  -8.14672622  -4.22415831  -3.03378285   1.10536665
## [186] -11.75173254 -15.20859989  -3.24255033  -4.72452088  -2.50112440
## [191]  -5.66615761  -8.13987322  -8.68545481  -5.35304602  -8.79362322
## [196]  -2.75924191  -7.52429297  -6.46755721  -6.16672811  -9.00356752
```

```
mse_knn_5=mean((test_data$yi-y_knn_5)^2);mse_knn_5
```

```
## [1] 1.303978
```

```
tr_kr_9=knnreg((yi)~(x1i)+(x2i),data=train_data,k=9);tr_kr_9
```

```
## 9-nearest neighbor regression model
```

```
y_knn_9=predict(tr_kr_9,test_data);y_knn_9
```

```
##   [1]  -1.523680580  -6.467273024  -2.319843659   0.793152357  -6.222652661
##   [6]  -6.814430510  -4.862561178  -4.394287720  -6.222652661 -12.598878951
##  [11] -14.305828203  -9.111194051  -7.089759633  -8.781592998  -9.115598013
##  [16]   0.000103839  -2.870338232 -10.839495485  -6.620655437 -10.437993408
##  [21]  -8.296908519  -2.958669529  -7.089759633  -7.587287184  -3.349285271
##  [26]  -6.332828591  -7.982563261  -6.864373011 -10.437993408  -8.632341197
##  [31]  -4.527800928  -8.718687261  -5.136922699  -5.844332913  -4.950552641
##  [36]  -3.459586649 -14.069477461  -6.449610481  -7.526122285  -9.229005267
##  [41]  -6.699155468   2.800064764 -10.838179293  -7.526122285  -5.474589713
##  [46]  -7.134026662  -9.241712030  -1.723838279   0.206577010  -2.344196229
##  [51] -11.317571026  -4.906038095  -1.687516601  -5.082716465  -0.747284498
##  [56]  -6.332828591  -8.526228023 -10.600323584  -4.475410900  -2.911852468
##  [61]  -4.408454350  -2.675006066  -5.950362620  -6.401213427  -7.872886631
##  [66]  -0.232080321 -12.410192621  -6.839439153  -3.643439258   0.374937585
##  [71] -10.202397547  -2.675006066  -4.834164579   0.236578537  -3.333317941
##  [76]  -4.762814300  -2.851585231   0.400685763  -0.879275064 -12.503673251
##  [81]  -3.114789586  -9.948425299 -10.324061721  -3.950999845  -3.098528542
##  [86]  -8.667416013 -10.316683410  -4.091369067  -9.143479901  -8.757665044
```

```
##  [91]  -4.725852838 -11.317571026  -4.834164579   2.927113909  -0.989637327
##  [96]  -6.259831490  -6.355137843  -8.065411808  -3.405263869  -2.151828537
## [101] -14.439423792  -9.304610191  -6.103979433   0.374937585  -4.527800928
## [106]  -8.672168020  -7.293902818  -4.723693149  -3.520608728  -6.413590018
## [111]  -0.232080321  -8.298104990  -7.235718842  -9.316405706  -7.969308479
## [116]  -4.527800928  -3.472688881 -13.134068952  -5.571809993  -2.911852468
## [121]  -2.441485681  -8.309390660  -7.134026662  -3.725168340  -5.678749506
## [126]   1.564662753  -7.227319270 -11.862929069  -0.232080321  -5.336671577
## [131]  -4.012431604  -6.007856055  -4.117208965  -2.753959716  -7.134026662
## [136]  -6.260067970  -9.494305350  -6.473317594  -2.753959716  -0.494704679
## [141]  -1.723838279 -10.548598095  -6.043455379  -4.158616064  -2.986133322
## [146] -11.939834729  -6.467273024  -1.031283571  -3.349285271  -8.125739345
## [151]  -2.746906337 -10.060970163  -9.264183722  -2.151828537  -3.725168340
## [156] -13.975716519  -2.040783110  -6.066014423  -8.632341197   0.194700001
## [161]  -7.587287184   3.977866294  -6.763495007  -4.202206835  -7.235718842
## [166]  -2.852030980 -14.483891281  -5.906356990  -2.512990498  -6.332828591
## [171]  -7.526122285  -9.764546746  -9.096368489  -4.834164579  -1.739007959
## [176]  -8.067495797 -16.066734809  -2.534890245 -10.357140671  -2.675006066
## [181]  -0.747284498  -8.037655935  -3.812374064  -3.643439258   1.647228890
## [186] -11.623307252 -14.175454340  -2.802289334  -4.725852838  -2.472286528
## [191]  -5.629478302  -8.067495797  -8.667416013  -5.146441962  -9.229005267
## [196]  -2.520833613  -7.123006212  -5.950362620  -6.304190633  -8.636178267
```

```
mse_knn_9=mean((test_data$yi-y_knn_9)^2);mse_knn_9
```

```
## [1] 1.205371
```

```
tr_kr_15=knnreg((yi)~(x1i)+(x2i),data=train_data,k=15);tr_kr_15
```

```
## 15-nearest neighbor regression model
```

```
y_knn_15=predict(tr_kr_15,test_data);y_knn_15
```

```
##   [1]  -1.28974221  -6.35527453  -2.29641065   0.73848232  -6.12888424
##   [6]  -7.04304966  -4.79819428  -3.39699649  -6.12888424 -12.28969899
##  [11] -13.92644101  -9.01637838  -7.65674203  -9.02194105  -9.06103281
##  [16]   0.11300789  -2.70569128  -9.73328663  -6.86544792 -10.46138265
##  [21]  -8.20860180  -2.82851717  -7.65674203  -7.50909676  -3.47595425
##  [26]  -6.23932572  -8.15862715  -6.65834643 -10.52704184  -8.82235446
##  [31]  -4.49363226  -9.03731632  -5.04108872  -6.18426832  -5.32000454
##  [36]  -3.67951156 -13.59621815  -6.51821282  -7.32770558  -8.78746893
##  [41]  -6.36726136   2.09194588 -10.68564775  -7.41035904  -5.66340448
##  [46]  -7.47871731  -8.86255570  -1.52791220   0.01074579  -2.53803550
##  [51] -10.29317708  -5.01460404  -1.50108915  -5.37156547  -1.20693506
##  [56]  -6.20299162  -8.77849663 -10.41590697  -4.48098116  -2.83327312
##  [61]  -4.63679516  -2.64992849  -5.76713930  -6.60480493  -7.83886731
##  [66]  -0.88860203 -12.24824140  -6.99367955  -3.39358496   0.05850468
##  [71] -10.20670019  -2.56390138  -5.37156547   0.21329290  -3.21273054
##  [76]  -3.80279514  -2.57460152   0.45298986  -0.95592564 -12.42725627
##  [81]  -3.25330312 -10.02339855 -10.13148201  -4.11226243  -2.97362563
```

```
##  [86]  -8.63807379 -10.01492863  -3.89613223  -9.33088184  -8.85722319
##  [91]  -4.46131586 -10.29317708  -5.37156547   2.39183348  -0.87953587
##  [96]  -6.44409467  -6.49640595  -8.15862715  -3.39190399  -1.84316927
## [101] -13.33554527  -9.47040997  -6.12292411   0.01074579  -4.37913642
## [106]  -8.88289834  -7.39669852  -4.70556499  -3.53069025  -6.13712501
## [111]  -0.88860203  -8.24926275  -7.27237251  -9.26780082  -8.13756815
## [116]  -4.26579641  -3.24712793 -12.24824140  -5.59133546  -3.21273054
## [121]  -2.73692233  -8.24991523  -7.47871731  -3.39190399  -5.79740459
## [126]   1.46158855  -7.65674203 -11.36347153  -1.20693506  -5.53093665
## [131]  -3.94546172  -5.66308100  -4.36174822  -2.97687423  -7.47871731
## [136]  -6.20493577  -9.29583792  -6.39918553  -2.85199438  -0.44002572
## [141]  -1.52791220 -10.54320494  -6.22757589  -4.03469559  -2.81325095
## [146] -11.79610511  -6.35527453  -1.82474740  -3.47595425  -8.20860180
## [151]  -2.71838305 -10.02339855  -9.03706383  -1.89089555  -3.39190399
## [156] -13.24840752  -2.04715117  -5.97418917  -8.60590415   0.25008135
## [161]  -7.50909676   2.99123515  -7.04304966  -4.08030490  -7.27237251
## [166]  -2.68004118 -13.98800496  -6.22757589  -2.61406792  -6.20493577
## [171]  -7.32770558 -10.15629845  -8.83880863  -5.37156547  -1.62037195
## [176]  -8.05113127 -15.73434018  -2.67548654 -10.45471772  -2.56390138
## [181]  -1.20693506  -8.05113127  -3.37606149  -3.39358496   1.85182609
## [186] -11.45605614 -13.48145399  -2.57460152  -4.63737860  -2.71784611
## [191]  -5.75026480  -7.93131160  -8.78251243  -5.32260587  -8.78746893
## [196]  -2.48874837  -6.99367955  -5.66308100  -5.95723882  -8.44843875
```

```
mse_knn_15=mean((test_data$yi-y_knn_15)^2);mse_knn_15
```

```
## [1] 1.23273
```

## RESULTS OF LINEAR MODEL

$$y_i = -2 + 1.4x_{1i} - 2.6x_{2i} + \epsilon_i$$

*Interpretation:*

Since the data generating process is *linear in parameters*, the classical multiple linear regression model is *correctly specified*. Under correct specification, the ordinary least squares (OLS) estimator is *unbiased and efficient* among linear estimators, and therefore is expected to perform well in terms of predictive accuracy.The true relationship between the predictors and the response is linear, the linear regression model captures the structural form of the data generating mechanism. Consequently, the test Mean Squared Error (MSE) for the linear regression model is expected to be low. *K-nearest neighbours (KNN)* regression, being a non-parametric method, does not assume any specific functional form and instead relies on local averaging of neighbouring observations. While KNN can approximate linear relationships, it does not explicitly exploit the known linear structure of the true model. For very small values of $k$, KNN may overfit the training data, leading to high variance and potentially higher test MSE. For very large values of $k$, excessive smoothing may occur, increasing bias. Therefore, for moderate values of $k$, KNN may perform reasonably well, but it is unlikely to outperform the correctly specified linear regression model.

*Hence, in the linear data generating process, multiple linear regression performs better than or at least comparably to KNN regression in terms of test MSE.*

---

*Suppose the data in Step (iii) is generated as :*

$$yi = \frac{1}{(-2 + 1.4_{x1i} - 2.6x_{2i} + 2.9x_{1i}^2)} + 3.1sin(x_{2i}) - 1.5x1ix_{2i}^2 + \epsilon_i.$$

Work out the problems in (1) and (2). Compare and comment on the results.

```
rm(list=ls())
set.seed(123)
n=1000
x1i=rnorm(n, mean=0,sd=2)
x2i=rpois(n, lambda=1.5)
error=rnorm(n, mean=0,sd=1)
yi= 1/((-2) + 1.4*x1i - 2.6*x2i + 2.9*(x1i)^2) + 3.1*sin(x2i) - 1.5*x1i*(x2i)^2 + error
data=data.frame(x1i,x2i,yi)
train_data=data[1:800,]
test_data=data[801:1000,]
# (a) fitting the multiple linear regression model and calculating Test MSE
model1=lm(yi~x1i+x2i,data=train_data);summary(model1)

##
## Call:
## lm(formula = yi ~ x1i + x2i, data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -241.819   -5.150   -0.051    5.566  155.662
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.2369     1.0272  -0.231  0.81764
## x1i          -6.3747     0.3373 -18.901  < 2e-16 ***
## x2i           1.3849     0.5283   2.621  0.00892 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.7 on 797 degrees of freedom
## Multiple R-squared:  0.3146, Adjusted R-squared:  0.3129
## F-statistic: 182.9 on 2 and 797 DF,  p-value: < 2.2e-16

pred_model1=predict(model1,newdata=test_data);summary(pred_model1)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -33.082  -6.081   1.741   1.373  10.377  35.892

mse.model1=mean((pred_model1-test_data$yi)^2);mse.model1
```

```
## [1] 205.1776
```

```
# (b) Fit a KNN model with k = 1, 2, 5, 9, 15. Calculate test MSE for each choice of k.
library(caret)
set.seed(123)
knnreg(train_data[,c("x1i","x2i")],train_data$yi,k=1)
```

```
## 1-nearest neighbor regression model
```

```
tr_kr_1=knnreg((yi)~(x1i)+(x2i),data=train_data,k=1);tr_kr_1
```

```
## 1-nearest neighbor regression model
```

```
y_knn_1=predict(tr_kr_1,test_data);y_knn_1
```

```
##   [1]    4.37412146    5.18263603   -0.15035182    1.62670968   -1.61298224
##   [6]   -0.53355419    2.07486250  -59.10389818   -1.61298224    7.17809534
##  [11]   62.10880036    8.99241162    3.79546316   -6.27298342   -6.32502272
##  [16]    0.53730373    0.14941122 -116.56079851    5.74970718   17.04143919
##  [21]    5.82996631    1.23853778    3.79546316  -25.61525688   -0.99196214
##  [26]    4.98521239    5.69004910   -0.14006821   16.57307283    7.43245154
##  [31]   -9.60965424   11.11957574   -6.50138935  -44.13840119    3.30004111
##  [36]    2.57005128   48.71622062    5.79901178    1.94028311   -6.83495502
##  [41]    7.48563262   -6.89234111   11.89160659   -0.80596111   -7.53143832
##  [46]    3.39721447    9.05368657    0.03207358   -2.48987918   -1.24132571
##  [51]  -29.70776317    2.06871172   -2.61194094    0.10201275  -22.31646321
##  [56]    4.02384018    8.42185261   10.63708727   -0.04280016  -15.95934027
##  [61]   -7.48487079   -0.79105742    0.54270542    7.06627671    4.77237089
##  [66]  -22.31646321   24.69864143  -30.18939008   -3.32561639   -1.56167742
##  [71]   15.10639567   -0.79105742    1.49390068    0.21772704  -14.46019530
##  [76]  -52.98317381   -1.70909808   -2.71977177   -1.30427252    7.17809534
##  [81]   -0.12038546   14.08994627    6.04482329  -13.71694418    3.44779111
##  [86]  -10.58291610   17.07565298    1.27109552   12.22117279  -11.32730334
##  [91]    0.78690575  -29.70776317    1.49390068    0.56121607    0.90388643
##  [96]    4.74788687    2.30027983    6.23357312    2.60279358   -1.94338366
## [101]   47.09905307   12.22117279   -1.14193704   -1.56167742   -8.88215704
## [106]    9.71674507  -23.49800845    2.47986422    2.77765303   -0.41681871
## [111]  -22.31646321    6.46703782    1.09335629    8.16318878  -16.39403190
## [116]   -8.88215704  -14.17117189   26.61110839   -2.31401919  -16.85496318
## [121]    0.93476661    4.21875484    3.01276799    2.60279358   -9.48626183
## [126]    0.01246410    4.60280439  -11.17887519  -22.31646321   -0.79338885
## [131]    1.82176393   -2.44352731    3.48947090   -0.99968307    2.01568846
## [136]    2.92223563    9.21486854    2.12411778   -0.99968307   -1.21526901
## [141]    0.03207358   10.63708727    1.02876270    0.16457292   -1.51289394
## [146]  -10.07798434    5.18263603  -20.61964856   -0.38862536    5.60397989
## [151]   -0.97202809   14.19393245   -1.67942289   -1.94338366   -0.50704142
## [156]   39.50747064   -1.94338366   -1.14193704    7.43245154   -0.94556169
## [161]  -25.61525688    0.56343316   -0.53355419    2.19862901    0.28601217
## [166]   -2.74446768   48.71622062    1.02876270   -0.44894669    2.88038255
## [171]    1.94028311    8.37643864    6.68547601    1.49390068   -0.75700738
```

```
## [176]    4.04922397   39.97284370   -2.74446768   10.63708727   -0.79105742
## [181]  -22.31646321    5.95424520   -2.62277997   -2.05787674   -1.26415677
## [186]   21.25586362   24.48631749   -1.70909808   -0.34166389  -15.95934027
## [191]  -15.17510917    4.63420438  -10.58291610    3.42232337   -8.58291275
## [196]   -0.44894669  -28.92793316   -2.44352731    4.14059600    7.12360739
```

```
mse_knn_1=mean((test_data$yi-y_knn_1)^2);mse_knn_1
```

```
## [1] 47.5249
```

```
tr_kr_2=knnreg((yi)~(x1i)+(x2i),data=train_data,k=2);tr_kr_2
```

```
## 2-nearest neighbor regression model
```

```
y_knn_2=predict(tr_kr_2,test_data);y_knn_2
```

```
##   [1]   2.524456233   4.537893875   0.457395084   0.941700208  -1.014900474
##   [6]  -0.200931547   2.305367942 -56.963551303  -1.014900474   7.882451304
##  [11]  47.688249794   7.803940524   4.449052222  -6.512560295  -6.579988874
##  [16]   0.679461371  -0.247946057 -84.728806589   6.131651804  18.229845178
##  [21]   6.207876261   1.530135845   4.449052222 -22.383185117  -0.690293749
##  [26]   3.932797469   6.825983199  -0.477228584  16.807256014   7.089353279
##  [31]  -9.763667792  10.529381362  -5.949711859 -44.453970316   3.253084040
##  [36]   1.308223114  51.518159269   6.453227920   0.567161000  -7.708933888
##  [41]   7.430991910  -6.141986406  13.382255730  -0.169864229  -8.508850075
##  [46]   3.204991229   9.627991324  -1.289933683  -2.090886811  -0.999255716
##  [51] -28.710672215   1.762695649  -1.289933683  -0.069980489 -22.055869193
##  [56]   4.504526285   7.584053813  10.528317716   1.401327042 -16.407151725
##  [61]  -7.796741980  -0.195147026   0.090316327   7.005373714   4.932262466
##  [66] -22.055869193  25.416870979 -30.202664358  -4.253022211  -2.025778295
##  [71]  15.839734255  -0.195147026   0.545201903   0.009067862 -15.465094306
##  [76] -52.398703103  -1.127205664  -2.782869621  -1.217104127   7.882451304
##  [81]  -0.324104411  14.874738886   5.720882031 -12.621816474   4.526487361
##  [86] -10.955109721  15.634792712   0.680468960  12.230924005 -10.884424260
##  [91]   0.517263037 -28.710672215   0.545201903  -0.341601250   0.566046265
##  [96]   4.422496378   0.429526800   7.548791857   1.010734535  -2.227893806
## [101]  47.851643459  12.167476837  -1.377459642  -2.025778295  -9.245905639
## [106]   9.385215820 -24.556632662   3.157564950   2.740507837  -1.014900474
## [111] -22.055869193   6.363181781   1.245317646   7.413928227 -16.401170286
## [116]  -9.634282055 -14.315683597  26.354255827  -1.585160857 -16.407151725
## [121]   0.574565727   3.646675515   3.204991229   1.047876081 -12.330685499
## [126]   0.501534362   3.729072517 -10.730762115 -22.055869193  -1.553704022
## [131]   2.141097168  -2.130884406   3.032635677  -0.899786551   2.706451463
## [136]   3.546372407   9.785056688   2.046521048  -1.313602004  -1.078907911
## [141]  -0.348358567  10.528317716   0.445382506   1.491223440  -1.277975500
## [146] -10.180316694   4.537893875 -21.218220254  -0.563108376   6.094883052
## [151]  -0.631277133  14.926731973  -2.501601770  -2.227893806   1.047876081
## [156]  40.408755838  -2.227893806  -1.111212508   7.269255722  -0.913688684
## [161] -22.383185117   0.391619695   0.417692029   2.329976131   0.689684226
## [166]  -2.698182396  51.518159269   0.445382506  -0.873670089   2.901309090
```

```
## [171]   1.759125561   8.985044038   5.930004801   0.545201903  -0.041343091
## [176]   5.766491176  39.158034572  -2.478522888   9.197574193  -1.889918046
## [181] -22.055869193   5.957865553  -2.560484805  -2.531737456  -0.683625204
## [186]  20.586975407  29.376028855  -1.127205664  -0.104256427 -16.603884871
## [191]  -7.634025186   6.058981381 -10.955109721   2.665714879  -7.708933888
## [196]  -0.873670089 -29.349941914  -2.130884406   3.765826618   5.915204658
```

```
mse_knn_2=mean((test_data$yi-y_knn_2)^2);mse_knn_2
```

```
## [1] 54.48942
```

```
tr_kr_5=knnreg((yi)~(x1i)+(x2i),data=train_data,k=5);tr_kr_5
```

```
## 5-nearest neighbor regression model
```

```
y_knn_5=predict(tr_kr_5,test_data);y_knn_5
```

```
##   [1]   6.404932396   4.417615870   0.359785824   0.131553253  -0.462536878
##   [6]   0.894472745   2.889626754 -56.797686801  -0.462536878   5.509041692
##  [11]  46.543860768   7.312415521   4.289680703  -5.253087456  -6.709167536
##  [16]   0.159597573  -0.664908324 -67.432665574   6.189017098  17.970824524
##  [21]   6.140700688   0.593796096   4.289680703 -22.309151216  -0.501960611
##  [26]   3.796435986   7.507501275   0.007607714  17.458188903   6.976230533
##  [31]  -9.761792399   9.961893074  -6.693152030 -37.748378999   3.050594374
##  [36]   2.054176551  47.919582704   6.692783848   1.211069325  -6.709167536
##  [41]   4.795783836  -3.214061145  12.362963186   0.595984620  -8.654555842
##  [46]   3.281027233  10.064000789  -0.765124812  -1.863678639  -0.467388890
##  [51] -31.012462081   2.540690344  -0.455031616   0.121755788 -21.591608815
##  [56]   3.795793919   6.892113754  11.221387980   2.165046872 -16.427189521
##  [61]  -7.537975598  -1.486842233  -0.547298163   6.741091211   4.002643322
##  [66] -15.561115276  26.188482854 -29.943049314  -9.963287944  -1.863678639
##  [71]  17.042962368  -1.486842233   0.371243752  -0.191411040 -15.123796045
##  [76] -52.742514728  -0.366999144  -3.122264328  -0.574847307   5.509041692
##  [81]   0.051694533  15.746482515   5.955180278 -12.228917882  -6.670042347
##  [86] -13.030821039  16.056453066   3.456479786  11.280679398  -9.553926480
##  [91]   0.452714284 -31.012462081   0.196780741  -0.036670139   0.390953918
##  [96]   3.144732358   0.096254663   7.170280277   1.168815041  -5.022639091
## [101]  52.227384475  12.443471145  -0.842001916  -1.863678639 -10.351278911
## [106]   9.861353885 -23.297951351   2.556120362   1.742207684  -0.343895342
## [111] -15.561115276   6.149219195   1.142892291   7.985030672 -16.627937753
## [116] -10.351278911 -15.123796045  26.188482854  -1.658869768 -16.198584307
## [121]   0.430391052   4.779226248   3.464755072   0.681707775  -7.618910221
## [126]  -0.144209426   4.289680703 -10.011411572 -18.306576765  -1.417444913
## [131]   1.926892016  -0.547298163   2.832520422  -0.994694238   2.576092578
## [136]   3.578195122  10.857877804   3.330089833  -0.854768573  -0.980186003
## [141]  -0.765124812  11.221387980   0.308034902   0.661817983  -0.806778734
## [146]  -6.446819041   4.417615870 -20.932867544  -0.885539473   6.149219195
## [151]  -0.720076045  16.056453066  -2.731235144  -5.022639091   0.399426455
## [156]  41.110475100  -3.390513389  -1.218229328   6.976230533  -0.752325488
## [161] -21.438601864   0.559780809   1.005213556   2.290696102   0.487383800
```

```
## [166]  -2.021355301  47.919582704   0.701311665  -0.744058083   3.578195122
## [171]   1.211069325   8.403488424   6.892113754   0.371243752   0.099452167
## [176]   5.616583567  34.482047562  -1.807286331  10.445579710  -1.486842233
## [181] -21.591608815   5.616583567  -1.934414892  -2.794747369  -1.095294911
## [186]  20.111661474  27.555549420  -0.463349681   0.371243752 -17.301628410
## [191]  -7.404956324   5.456083286 -13.030821039   2.888344075  -7.680070507
## [196]  -0.937663514 -28.817735502  -0.547298163   3.682188825   4.387229756
```

```
mse_knn_5=mean((test_data$yi-y_knn_5)^2);mse_knn_5
```

```
## [1] 59.77963
```

```
tr_kr_9=knnreg((yi)~(x1i)+(x2i),data=train_data,k=9);tr_kr_9
```

```
## 9-nearest neighbor regression model
```

```
y_knn_9=predict(tr_kr_9,test_data);y_knn_9
```

```
##   [1]   1.019343157   3.665620952   0.185892387  -0.260726528  -0.367990839
##   [6]   1.200390957   2.609729622 -53.430944656  -0.367990839   7.586495474
##  [11]  37.850820364   7.201752347   3.638988710  -5.257861641  -6.815778796
##  [16]  -0.152241701  -0.789621688 -52.732514546   6.338341539  17.838320615
##  [21]   6.250149985   0.247987204   3.638988710 -21.536115359  -1.109354914
##  [26]   3.838657204   7.318131231   0.255370249  17.838320615   6.828800698
##  [31] -10.215214650   9.873093424  -6.215749282 -39.064003730   3.195748685
##  [36]   1.736984748  43.240500797   6.830432857   1.255952826  -7.593311866
##  [41]   4.600950072  -4.721326559  12.829167196   1.255952826  -6.790029536
##  [46]   3.146850536  10.007912252  -0.545830339  -1.439227204  -0.806110126
##  [51] -22.635017932   2.633728665  -0.220687586   0.129543163 -28.345897074
##  [56]   3.838657204   7.095588763   9.843004311   1.638822060 -16.141351010
##  [61]  -7.438366714  -1.887393080  -0.056443686   6.637552253   4.471321497
##  [66] -22.544612955  26.311517565 -30.494464811  -1.817291227  -1.885749860
##  [71]  17.758277401  -1.887393080   0.305642651  -0.158389981 -14.375107306
##  [76] -51.264462250  -0.125359571  -2.688681184  -0.759329627   4.174655323
##  [81]  -0.441822121  14.317782941   5.004358470 -12.336064471  -3.603190808
##  [86] -12.623969818  15.192489744  -0.017628957  11.464921390  -8.072193275
##  [91]   0.346036010 -22.635017932   0.305642651   0.031525116   0.319427701
##  [96]   3.776737961   0.118650979   6.952721863   1.101737680  -2.209989353
## [101]  41.417314393  11.492055313  -0.683761686  -1.885749860 -10.215214650
## [106]  10.106819326 -22.434153480   2.573235996   1.573212295  -0.160335164
## [111] -22.544612955   5.903944037   1.086969685   7.750375053 -16.878530331
## [116] -10.215214650 -13.873232752  26.528772898  -1.664451639 -16.141351010
## [121]  -0.192686890   4.990284318   3.146850536   0.869660580  -6.620037562
## [126]  -0.082460306   3.849605036 -14.604547457 -22.544612955  -1.725042945
## [131]   1.769479770  -0.211688410   2.795499846  -0.616880115   3.146850536
## [136]   3.866151012  10.549778953   4.532463878  -0.616880115  -1.082358991
## [141]  -0.545830339  10.886149210   0.646729131   0.403037193  -0.873971783
## [146]  -6.472742230   3.665620952 -20.074959948  -1.109354914   6.192203251
## [151]  -0.871083456  14.799659818  -3.561713937  -2.209989353   0.869660580
## [156]  39.031746909  -4.658453756  -1.069951826   6.828800698  -0.483056312
```

```
## [161] -21.536115359  -0.668966882   1.064137905   2.483390426   1.086969685
## [166]  -1.550497594  46.062065080   0.622061957  -0.720206196   3.838657204
## [171]   1.255952826   8.012842501   6.908372578   0.305642651  -0.182000866
## [176]   5.560287779  44.508648577  -1.300188022   9.133237594  -1.887393080
## [181] -28.345897074   5.769167900  -1.533401778  -1.817291227  -0.483634115
## [186]  20.488469907  12.879722551  -0.003579855   0.346036010 -16.736917713
## [191]  -6.458353642   5.560287779 -12.623969818   3.091802134  -7.593311866
## [196]  -0.695124150 -29.014581442  -0.056443686   3.618573761   5.004800450
```

```
mse_knn_9=mean((test_data$yi-y_knn_9)^2);mse_knn_9
```

```
## [1] 62.10913
```

```
tr_kr_15=knnreg((yi)~(x1i)+(x2i),data=train_data,k=15);tr_kr_15
```

```
## 15-nearest neighbor regression model
```

```
y_knn_15=predict(tr_kr_15,test_data);y_knn_15
```

```
##   [1]   0.07208759   3.72703050   0.25791873  -0.30979295  -0.52672299
##   [6]   0.81121331   2.63548599 -47.92029452  -0.52672299   4.29196233
##  [11]  38.48057697   7.26434206   3.53611156  -4.39011157  -5.66361932
##  [16]  -0.03933436  -0.63606177 -47.95506737   6.27375813  17.66487507
##  [21]   6.20008813   0.21711345   3.53611156 -20.94071413  -1.24367996
##  [26]   3.84439683   7.02909265   0.36413831  17.23520981   6.73212438
##  [31]  -9.20853263   9.70372903  -6.07792003 -38.36524349   2.86112799
##  [36]   1.63112949  41.45331783   5.59160347   1.44815975  -6.17192657
##  [41]   5.39269936  -8.55554467  11.86462520   1.18610851  -6.03267029
##  [46]   2.48725263  10.41270889  -0.27450807  -1.55473939  -0.91810406
##  [51] -24.93448177   2.52095016  -0.13405809  -0.02000382 -26.69508133
##  [56]   3.92618825   6.92168650   9.36482261   1.59790242 -15.69012150
##  [61]  -7.85658965  -1.59669024   0.24320203   6.13076830   4.51191939
##  [66] -27.82019083  24.70049515 -28.70301621  -1.47349760  -1.80530431
##  [71]  17.26124408  -2.11320685  -0.02000382  -0.27041337 -14.65039377
##  [76] -46.84924373   0.19552759  -1.89912235  -0.76239277   3.89319075
##  [81]  -2.78902457  14.34216939   4.31051215 -11.34799662  -1.90487460
##  [86] -11.66966438  15.29117493  -0.49666688  11.19565801  -9.93988830
##  [91]   0.45299824 -24.93448177  -0.02000382  -0.17803051   1.52759406
##  [96]   4.24428641  -0.25414458   7.02909265   1.10490187  -0.70263160
## [101]  36.14643594  11.26820132  -0.84822974  -1.55473939 -10.18117985
## [106]  10.22330173 -22.41598744   2.68773177   1.47277765  -0.34295638
## [111] -27.82019083   5.96769659   1.14299183   7.50602553 -17.63852460
## [116] -10.44448533 -14.04911745  24.70049515  -1.66246100 -14.65039377
## [121]  -0.57526999   5.16051295   2.48725263   1.10490187  -6.25302469
## [126]  -0.25034566   3.53611156 -11.08425650 -26.69508133  -1.76076171
## [131]   1.90918665   0.13414031   2.61756040  -0.81006998   2.48725263
## [136]   3.84664570  10.52006489   4.84860093  -0.70832588  -0.77755560
## [141]  -0.27450807  10.19976448   0.37157641   0.63867231  -0.93469992
## [146]  -4.79384739   3.72703050 -22.87413100  -1.24367996   6.20008813
## [151]  -0.77421596  14.34216939  -3.80881461  -0.80760000   1.10490187
```

```
## [156]  38.19118586  -1.03555057  -1.03041525   6.78739923  -0.27755775
## [161] -20.94071413  -1.33904670   0.81121331   2.58701757   1.14299183
## [166]  -1.36444050  42.60722388   0.37157641  -0.86695213   3.84664570
## [171]   1.44815975  10.76381513   7.00618619  -0.02000382  -0.13756345
## [176]   5.78420610  47.83991485  -1.43287182   8.39104028  -2.11320685
## [181] -26.69508133   5.78420610  -1.11371355  -1.47349760  -0.37559242
## [186]  21.67911418   4.81626985   0.19552759   0.35079227 -16.30193345
## [191]  -6.29080597   5.73085627 -10.80416078   2.89332673  -6.17192657
## [196]  -0.70837615 -28.70301621   0.13414031   3.92792416   5.70966854
```

```
mse_knn_15=mean((test_data$yi-y_knn_15)^2);mse_knn_15
```

```
## [1] 63.72303
```

# RESULTS OF NON LINEAR MODEL

$$y_i = \frac{1}{-2 + 1.4x_{1i} - 2.6x_{2i} + 2.9x_{1i}^2} + 3.1\sin(x_{2i}) - 1.5x_{1i}x_{2i}^2 + \epsilon_i$$

*Interpretation:*

The data generating process in this case is nonlinear and includes higher-order and interaction terms. A linear regression model is therefore *misspecified* and unable to capture the true functional relationship between the predictors and the response. This is reflected in the high test Mean Squared Error (MSE) of approximately 205.18. In contrast, K-nearest neighbours (KNN) regression, being a flexible non-parametric method, adapts to the nonlinear structure of the data and achieves substantially lower test MSE (approximately 47 for the optimal choice of $k$).

*Thus, in the presence of strong nonlinearity, KNN regression clearly outperforms the linear regression model in terms of predictive accuracy.*

---

*Overall Comparison*

The results demonstrate that model performance depends critically on the underlying data generating process.

-In the linear setting, where the true relationship between the predictors and the response is linear, the multiple linear regression model is correctly specified and achieves the lowest test Mean Squared Error (MSE). KNN regression, although flexible, does not improve predictive accuracy in this case.

-In contrast, under the nonlinear data generating process, the linear regression model is misspecified and produces a substantially higher test MSE. KNN regression, being a non-parametric and flexible method, adapts to the complex nonlinear structure and achieves significantly lower prediction error.

Overall, the findings illustrate that parametric models perform best when their structural assumptions are correct, whereas flexible non-parametric methods are advantageous when the true relationship is nonlinear.

---