HW 18-19

Imports

```
In [1]:   import warnings

          import numpy as np
          import pandas as pd
          from matplotlib import pyplot as plt

          warnings.filterwarnings("ignore")
```

Bias-Variance Trade

```
In [2]:   # It's the same dataset for both HW 18 and HW 19
          data: np.ndarray = pd.read_csv("bias-variance-trade.txt", delimiter=" ", header=None).values.astype(np.floa
          x1 = data[:, 0]
          y1 = data[:, 1]
          x2 = data[:, 2]
          y2 = data[:, 3]
```

HW 18

In this problem, we will use the data in bias variance trade.txt to practice using cross-validation in model evaluation. The first column is one sample of the independent variable x, while the second is the dependent variable y that corresponds to those points. The third and fourth columns are the same for a different sample of the independent variable. You will want to complete the Bias-Variance Tradeoff problem before proceeding.

```
In [3]:   # (a) Fit the data from the first two columns using Legendre polynomials of orders 1 through 10.
          # For each fit, preform leave-one-out cross validation and report the average error to the point
```

```python
# left out. Note that this means for each order of polynomial, you will perform 20 fits. If you
# have not found a shortcut in your programming language of choice already, do so before
# proceeding.

# a design matrix is created as follows:
def design_mat(k: int, vec) -> np.ndarray:
    """ copied from HW 16 """
    A = np.stack([np.polynomial.legendre.Legendre.basis(i)(vec) for i in range(k)]).T
    return A


optimal_fit_k = -1
optimal_fit_mean_mse = np.inf
optimal_test_k = -1
optimal_test_resid = np.inf


for k in range(1,12):
    A = design_mat(k, x1)
    test_resids = []
    fit_mses = []
    for i in range(A.shape[0]):
        A_tmp = np.concatenate([A[:i], A[i+1:]])
        y1_tmp = np.concatenate([y1[:i], y1[i+1:]])
        A_i = A[i]
        y1_i = y1[i]

        theta_hat, sse_crossval, _, _ = np.linalg.lstsq(A_tmp, y1_tmp)
        mse_crossval = sse_crossval/A_tmp.shape[0]
        y1_i_hat = A_i@theta_hat
        test_resid = abs(y1_i_hat-y1_i)
        test_resids.append(test_resid)
        fit_mses.append(mse_crossval)

        print(f"With K={k}, leaving out row {i}, our fit MSE is {mse_crossval} and test residual is {test_
    mean_test_resid = np.mean(test_resids)
    mean_fit_mse = np.mean(fit_mses)
    print("="*20)
    print(f"Across all leave-one-out runs for K={k}:\n- Our average test residual is {mean_test_resid} and

    if mean_test_resid < optimal_test_resid:
        optimal_test_resid = mean_test_resid
        optimal_test_k = k
    if mean_fit_mse < optimal_fit_mean_mse:
```

```
            optimal_fit_mean_mse = mean_fit_mse
            optimal_fit_k = k

print("="*40)
print(f"For this dataset and bucket of models:\n- The ideal fit by test residual is K={optimal_test_k}, wh
```

```
With K=1, leaving out row 0, our fit MSE is [0.15328997] and test residual is 0.3683567385819108
With K=1, leaving out row 1, our fit MSE is [0.15396343] and test residual is 0.34959616727527376
With K=1, leaving out row 2, our fit MSE is [0.15694634] and test residual is 0.250118431909715
With K=1, leaving out row 3, our fit MSE is [0.15894036] and test residual is 0.15059502042414732
With K=1, leaving out row 4, our fit MSE is [0.15816128] and test residual is 0.1956029581896502
With K=1, leaving out row 5, our fit MSE is [0.15939403] and test residual is 0.11664257965398506
With K=1, leaving out row 6, our fit MSE is [0.14927922] and test residual is 0.46465214080537465
With K=1, leaving out row 7, our fit MSE is [0.15916202] and test residual is 0.13507648407008474
With K=1, leaving out row 8, our fit MSE is [0.15400719] and test residual is 0.3483422065965093
With K=1, leaving out row 9, our fit MSE is [0.14367974] and test residual is 0.5726178588970473
With K=1, leaving out row 10, our fit MSE is [0.11978988] and test residual is 0.8976015485981512
With K=1, leaving out row 11, our fit MSE is [0.15811015] and test residual is 0.19819977213052942
With K=1, leaving out row 12, our fit MSE is [0.15923434] and test residual is 0.12961216405596582
With K=1, leaving out row 13, our fit MSE is [0.16004365] and test residual is 0.02476212313970115
With K=1, leaving out row 14, our fit MSE is [0.12549402] and test residual is 0.8316282785624198
With K=1, leaving out row 15, our fit MSE is [0.15999217] and test residual is 0.04053004277839256
With K=1, leaving out row 16, our fit MSE is [0.15987557] and test residual is 0.06304477114805646
With K=1, leaving out row 17, our fit MSE is [0.12401573] and test residual is 0.8492181757187022
With K=1, leaving out row 18, our fit MSE is [0.15958055] and test residual is 0.09937296837860246
With K=1, leaving out row 19, our fit MSE is [0.16002717] and test residual is 0.03070354114558615
====================
Across all leave-one-out runs for K=1:
- Our average test residual is 0.3058136986029903 and our average fit mse is 0.15164933987639292
With K=2, leaving out row 0, our fit MSE is [0.15314715] and test residual is 0.37165575671635476
With K=2, leaving out row 1, our fit MSE is [0.15389458] and test residual is 0.3568495593647903
With K=2, leaving out row 2, our fit MSE is [0.15690382] and test residual is 0.25719603872289126
With K=2, leaving out row 3, our fit MSE is [0.15858366] and test residual is 0.1797795411188805
With K=2, leaving out row 4, our fit MSE is [0.15815998] and test residual is 0.19654575632515747
With K=2, leaving out row 5, our fit MSE is [0.15920936] and test residual is 0.13184502612171367
With K=2, leaving out row 6, our fit MSE is [0.1477693] and test residual is 0.5176594230070772
With K=2, leaving out row 7, our fit MSE is [0.15914496] and test residual is 0.13333936927518242
With K=2, leaving out row 8, our fit MSE is [0.15366184] and test residual is 0.3607762675880889
With K=2, leaving out row 9, our fit MSE is [0.14347844] and test residual is 0.5762470727062117
With K=2, leaving out row 10, our fit MSE is [0.11926507] and test residual is 0.9128319014688485
With K=2, leaving out row 11, our fit MSE is [0.15793943] and test residual is 0.22539276271597036
With K=2, leaving out row 12, our fit MSE is [0.15923205] and test residual is 0.12833700728330633
With K=2, leaving out row 13, our fit MSE is [0.16001578] and test residual is 0.01617260335167495
With K=2, leaving out row 14, our fit MSE is [0.1248838] and test residual is 0.8422135078363235
With K=2, leaving out row 15, our fit MSE is [0.15987391] and test residual is 0.05943008292196647
With K=2, leaving out row 16, our fit MSE is [0.15984364] and test residual is 0.060842378276132765
With K=2, leaving out row 17, our fit MSE is [0.12399179] and test residual is 0.8489922258885867
With K=2, leaving out row 18, our fit MSE is [0.15957434] and test residual is 0.0971155342399018
```

With K=2, leaving out row 19, our fit MSE is [0.15993638] and test residual is 0.044979833010498224
====================
Across all leave-one-out runs for K=2:
- Our average test residual is 0.31591279857081 and our average fit mse is 0.15142546475927457
With K=3, leaving out row 0, our fit MSE is [0.15271669] and test residual is 0.3924648657252174
With K=3, leaving out row 1, our fit MSE is [0.15389432] and test residual is 0.3568010896249205
With K=3, leaving out row 2, our fit MSE is [0.15685525] and test residual is 0.2599700498070736
With K=3, leaving out row 3, our fit MSE is [0.15840919] and test residual is 0.20528594167642836
With K=3, leaving out row 4, our fit MSE is [0.15815986] and test residual is 0.19647941184997675
With K=3, leaving out row 5, our fit MSE is [0.15920852] and test residual is 0.1316657318807541
With K=3, leaving out row 6, our fit MSE is [0.14766069] and test residual is 0.5213071237435812
With K=3, leaving out row 7, our fit MSE is [0.15905258] and test residual is 0.14417421238482922
With K=3, leaving out row 8, our fit MSE is [0.15354798] and test residual is 0.365935988040416
With K=3, leaving out row 9, our fit MSE is [0.14246386] and test residual is 0.6140163226212358
With K=3, leaving out row 10, our fit MSE is [0.11747056] and test residual is 0.9509603946508585
With K=3, leaving out row 11, our fit MSE is [0.15725316] and test residual is 0.3065827265993309
With K=3, leaving out row 12, our fit MSE is [0.15923074] and test residual is 0.128155192536844
With K=3, leaving out row 13, our fit MSE is [0.16001382] and test residual is 0.01463949723628729
With K=3, leaving out row 14, our fit MSE is [0.12350586] and test residual is 0.8733106332297803
With K=3, leaving out row 15, our fit MSE is [0.15984808] and test residual is 0.07434243366888657
With K=3, leaving out row 16, our fit MSE is [0.15981216] and test residual is 0.06719213124326431
With K=3, leaving out row 17, our fit MSE is [0.1217816] and test residual is 0.9034980786155655
With K=3, leaving out row 18, our fit MSE is [0.15957032] and test residual is 0.097062550240943389
With K=3, leaving out row 19, our fit MSE is [0.15992359] and test residual is 0.04750132003254135
====================
Across all leave-one-out runs for K=3:
- Our average test residual is 0.3325672847704368 and our average fit mse is 0.1510189412603708
With K=4, leaving out row 0, our fit MSE is [0.15265973] and test residual is 0.39679648677430895
With K=4, leaving out row 1, our fit MSE is [0.1535268] and test residual is 0.37465298715471207
With K=4, leaving out row 2, our fit MSE is [0.15673981] and test residual is 0.26666754937112475
With K=4, leaving out row 3, our fit MSE is [0.15821452] and test residual is 0.22433482295550272
With K=4, leaving out row 4, our fit MSE is [0.15800817] and test residual is 0.2083457893296134
With K=4, leaving out row 5, our fit MSE is [0.15918304] and test residual is 0.13843102161611387
With K=4, leaving out row 6, our fit MSE is [0.14733438] and test residual is 0.5374809471327835
With K=4, leaving out row 7, our fit MSE is [0.15897029] and test residual is 0.15213242214600303
With K=4, leaving out row 8, our fit MSE is [0.1533422] and test residual is 0.38014787367195696
With K=4, leaving out row 9, our fit MSE is [0.14229904] and test residual is 0.6211870500112732
With K=4, leaving out row 10, our fit MSE is [0.11372905] and test residual is 1.0301576672553059
With K=4, leaving out row 11, our fit MSE is [0.15533826] and test residual is 0.5389434948189504
With K=4, leaving out row 12, our fit MSE is [0.15914827] and test residual is 0.13682711015366988
With K=4, leaving out row 13, our fit MSE is [0.16000419] and test residual is 0.014593863528355722
With K=4, leaving out row 14, our fit MSE is [0.1225766] and test residual is 0.89783738293258

```
With K=4, leaving out row 15, our fit MSE is [0.15967182] and test residual is 0.1277300784548458
With K=4, leaving out row 16, our fit MSE is [0.15977849] and test residual is 0.07207802472483057
With K=4, leaving out row 17, our fit MSE is [0.12171614] and test residual is 0.9044496404799797
With K=4, leaving out row 18, our fit MSE is [0.15951591] and test residual is 0.10371617967644925
With K=4, leaving out row 19, our fit MSE is [0.15991922] and test residual is 0.046462230554837836
====================
Across all leave-one-out runs for K=4:
- Our average test residual is 0.358649028354166 and our average fit mse is 0.15058379726283694
With K=5, leaving out row 0, our fit MSE is [0.14987762] and test residual is 0.34974521400947783
With K=5, leaving out row 1, our fit MSE is [0.15059597] and test residual is 0.3275373707391907
With K=5, leaving out row 2, our fit MSE is [0.15369363] and test residual is 0.198381092911881
With K=5, leaving out row 3, our fit MSE is [0.15355707] and test residual is 0.22620929848976423
With K=5, leaving out row 4, our fit MSE is [0.15420492] and test residual is 0.16197084229681497
With K=5, leaving out row 5, our fit MSE is [0.15525356] and test residual is 0.05722872345857338
With K=5, leaving out row 6, our fit MSE is [0.14624161] and test residual is 0.4829464578668591
With K=5, leaving out row 7, our fit MSE is [0.15503396] and test residual is 0.09028128438096136
With K=5, leaving out row 8, our fit MSE is [0.14801982] and test residual is 0.40015444004250644
With K=5, leaving out row 9, our fit MSE is [0.1309935] and test residual is 0.7549526557333004
With K=5, leaving out row 10, our fit MSE is [0.11041964] and test residual is 1.016612167384852
With K=5, leaving out row 11, our fit MSE is [0.15496346] and test residual is 0.3005085601796941
With K=5, leaving out row 12, our fit MSE is [0.15505107] and test residual is 0.08633875284622902
With K=5, leaving out row 13, our fit MSE is [0.15499648] and test residual is 0.09919490678442977
With K=5, leaving out row 14, our fit MSE is [0.12023683] and test residual is 0.8746169244938655
With K=5, leaving out row 15, our fit MSE is [0.15504153] and test residual is 0.14880707495360146
With K=5, leaving out row 16, our fit MSE is [0.15538508] and test residual is 0.0041145154114381015
With K=5, leaving out row 17, our fit MSE is [0.10596806] and test residual is 1.0669963995269627
With K=5, leaving out row 18, our fit MSE is [0.15530632] and test residual is 0.04232037607709728
With K=5, leaving out row 19, our fit MSE is [0.15524957] and test residual is 0.05878109688206885
====================
Across all leave-one-out runs for K=5:
- Our average test residual is 0.33738490772347846 and our average fit mse is 0.14600448513136435
With K=6, leaving out row 0, our fit MSE is [0.14889758] and test residual is 0.4043790402527211
With K=6, leaving out row 1, our fit MSE is [0.15052462] and test residual is 0.3303417257967663
With K=6, leaving out row 2, our fit MSE is [0.15330456] and test residual is 0.2351125622433497
With K=6, leaving out row 3, our fit MSE is [0.15341652] and test residual is 0.2491911268341379
With K=6, leaving out row 4, our fit MSE is [0.15418999] and test residual is 0.1621198664902348
With K=6, leaving out row 5, our fit MSE is [0.15522406] and test residual is 0.06146824300309153
With K=6, leaving out row 6, our fit MSE is [0.1461645] and test residual is 0.4886946106417849
With K=6, leaving out row 7, our fit MSE is [0.15503362] and test residual is 0.090925529526161583
With K=6, leaving out row 8, our fit MSE is [0.14763708] and test residual is 0.41707366706905663
With K=6, leaving out row 9, our fit MSE is [0.1307445] and test residual is 0.7609042663929086
With K=6, leaving out row 10, our fit MSE is [0.10432668] and test residual is 1.1611420197943936
```

```
With K=6, leaving out row 11, our fit MSE is [0.15444668] and test residual is 0.7682314919959743
With K=6, leaving out row 12, our fit MSE is [0.1550345] and test residual is 0.08684982766580746
With K=6, leaving out row 13, our fit MSE is [0.15487423] and test residual is 0.11959103519986855
With K=6, leaving out row 14, our fit MSE is [0.11506155] and test residual is 0.9965701686089126
With K=6, leaving out row 15, our fit MSE is [0.1549578] and test residual is 0.20256007930196804
With K=6, leaving out row 16, our fit MSE is [0.15537296] and test residual is 0.0005750691644094211
With K=6, leaving out row 17, our fit MSE is [0.10372332] and test residual is 1.1192240075626658
With K=6, leaving out row 18, our fit MSE is [0.15528487] and test residual is 0.04484659463325581
With K=6, leaving out row 19, our fit MSE is [0.15520155] and test residual is 0.06838827822717974
====================
Across all leave-one-out runs for K=6:
- Our average test residual is 0.3884094488070051 and our average fit mse is 0.145171059195411
With K=7, leaving out row 0, our fit MSE is [0.13830655] and test residual is 0.44911729716325
With K=7, leaving out row 1, our fit MSE is [0.13981516] and test residual is 0.3822344811516015
With K=7, leaving out row 2, our fit MSE is [0.14558356] and test residual is 0.135073365535529774
With K=7, leaving out row 3, our fit MSE is [0.12778471] and test residual is 0.9905640292024365
With K=7, leaving out row 4, our fit MSE is [0.14387103] and test residual is 0.23155605425266143
With K=7, leaving out row 5, our fit MSE is [0.14197843] and test residual is 0.3730932435412918
With K=7, leaving out row 6, our fit MSE is [0.13473289] and test residual is 0.5495709576816451
With K=7, leaving out row 7, our fit MSE is [0.14557425] and test residual is 0.12720497375877945
With K=7, leaving out row 8, our fit MSE is [0.14127077] and test residual is 0.3408132412587204
With K=7, leaving out row 9, our fit MSE is [0.12690693] and test residual is 0.6917170781084172
With K=7, leaving out row 10, our fit MSE is [0.10324949] and test residual is 1.111216695165346
With K=7, leaving out row 11, our fit MSE is [0.13923015] and test residual is 3.688801662500599
With K=7, leaving out row 12, our fit MSE is [0.14524915] and test residual is 0.14952608475109172
With K=7, leaving out row 13, our fit MSE is [0.14051139] and test residual is 0.45798851839494664
With K=7, leaving out row 14, our fit MSE is [0.10834296] and test residual is 0.9684952781098219
With K=7, leaving out row 15, our fit MSE is [0.12775394] and test residual is 2.0135875652723083
With K=7, leaving out row 16, our fit MSE is [0.1461467] and test residual is 0.04638176523166292
With K=7, leaving out row 17, our fit MSE is [0.10243019] and test residual is 1.069921767117001
With K=7, leaving out row 18, our fit MSE is [0.1459684] and test residual is 0.07816705819579488
With K=7, leaving out row 19, our fit MSE is [0.14525765] and test residual is 0.16611697674698933
====================
Across all leave-one-out runs for K=7:
- Our average test residual is 0.701057404647983 and our average fit mse is 0.13449821394002276
With K=8, leaving out row 0, our fit MSE is [0.13821815] and test residual is 0.43233357126706173
With K=8, leaving out row 1, our fit MSE is [0.13907587] and test residual is 0.36313802769340453
With K=8, leaving out row 2, our fit MSE is [0.14429316] and test residual is 0.1134994692213854
With K=8, leaving out row 3, our fit MSE is [0.11606616] and test residual is 1.9789049367033682
With K=8, leaving out row 4, our fit MSE is [0.14277774] and test residual is 0.21284641218432523
With K=8, leaving out row 5, our fit MSE is [0.13503842] and test residual is 0.6543621611395692
With K=8, leaving out row 6, our fit MSE is [0.12973984] and test residual is 0.6527402284890258
```

```
With K=8, leaving out row 7, our fit MSE is [0.14347513] and test residual is 0.18152534933246883
With K=8, leaving out row 8, our fit MSE is [0.13998702] and test residual is 0.3339990997251288
With K=8, leaving out row 9, our fit MSE is [0.12658837] and test residual is 0.6781970322309566
With K=8, leaving out row 10, our fit MSE is [0.09937845] and test residual is 1.14885825379254
With K=8, leaving out row 11, our fit MSE is [0.08841988] and test residual is 21.177700219241977
With K=8, leaving out row 12, our fit MSE is [0.14403424] and test residual is 0.1282104999721821
With K=8, leaving out row 13, our fit MSE is [0.13161338] and test residual is 0.824171137754102
With K=8, leaving out row 14, our fit MSE is [0.10802911] and test residual is 0.9947396570407225
With K=8, leaving out row 15, our fit MSE is [0.10676893] and test residual is 5.449274782703361
With K=8, leaving out row 16, our fit MSE is [0.14439757] and test residual is 0.09496260595507874
With K=8, leaving out row 17, our fit MSE is [0.09601202] and test residual is 1.1497914794807391
With K=8, leaving out row 18, our fit MSE is [0.14464051] and test residual is 0.05008205011599272
With K=8, leaving out row 19, our fit MSE is [0.14362153] and test residual is 0.17859480214951756
====================
Across all leave-one-out runs for K=8:
- Our average test residual is 1.8398965888096452 and our average fit mse is 0.12810877315143115
With K=9, leaving out row 0, our fit MSE is [0.04628131] and test residual is 0.416875430974694544
With K=9, leaving out row 1, our fit MSE is [0.04811246] and test residual is 0.3141204821977115
With K=9, leaving out row 2, our fit MSE is [0.031703] and test residual is 0.8747127399518615
With K=9, leaving out row 3, our fit MSE is [0.03249302] and test residual is 2.8877583788779417
With K=9, leaving out row 4, our fit MSE is [0.04826212] and test residual is 0.3073320473494035
With K=9, leaving out row 5, our fit MSE is [0.04659367] and test residual is 0.5053442490495397
With K=9, leaving out row 6, our fit MSE is [0.05232239] and test residual is 0.0309741847041905
With K=9, leaving out row 7, our fit MSE is [0.0517591] and test residual is 0.12393213972131961
With K=9, leaving out row 8, our fit MSE is [0.05146676] and test residual is 0.15193802362637965
With K=9, leaving out row 9, our fit MSE is [0.04924574] and test residual is 0.2909947671189159
With K=9, leaving out row 10, our fit MSE is [0.04849241] and test residual is 0.39425693673607853
With K=9, leaving out row 11, our fit MSE is [0.05107347] and test residual is 5.944730477291114
With K=9, leaving out row 12, our fit MSE is [0.05095094] and test residual is 0.17980319619658006
With K=9, leaving out row 13, our fit MSE is [0.04716841] and test residual is 0.6308562049694412
With K=9, leaving out row 14, our fit MSE is [0.04020609] and test residual is 0.5964119612420273
With K=9, leaving out row 15, our fit MSE is [0.0340562] and test residual is 7.528773607093412
With K=9, leaving out row 16, our fit MSE is [0.05197828] and test residual is 0.09808558977341278
With K=9, leaving out row 17, our fit MSE is [0.03979902] and test residual is 0.6326289028255502
With K=9, leaving out row 18, our fit MSE is [0.05188273] and test residual is 0.10558859054746804
With K=9, leaving out row 19, our fit MSE is [0.04639177] and test residual is 0.4151026243960264
====================
Across all leave-one-out runs for K=9:
- Our average test residual is 1.1215110267321535 and our average fit mse is 0.046011943282771285
With K=10, leaving out row 0, our fit MSE is [0.04476118] and test residual is 0.34825491666457553
With K=10, leaving out row 1, our fit MSE is [0.04501707] and test residual is 0.2893649597798163
With K=10, leaving out row 2, our fit MSE is [0.02922672] and test residual is 1.1562704970135067
```

```
With K=10, leaving out row 3, our fit MSE is [0.01515446] and test residual is 7.881457285320115
With K=10, leaving out row 4, our fit MSE is [0.04568428] and test residual is 0.26302296541168124
With K=10, leaving out row 5, our fit MSE is [0.03585503] and test residual is 0.8306874280086476
With K=10, leaving out row 6, our fit MSE is [0.04808842] and test residual is 0.14145778661190728
With K=10, leaving out row 7, our fit MSE is [0.04720543] and test residual is 0.1928539785559728
With K=10, leaving out row 8, our fit MSE is [0.04840519] and test residual is 0.0709931569470788
With K=10, leaving out row 9, our fit MSE is [0.04670006] and test residual is 0.23220448164485452
With K=10, leaving out row 10, our fit MSE is [0.04682442] and test residual is 0.2877874084129701
With K=10, leaving out row 11, our fit MSE is [0.01684762] and test residual is 59.351307988029234
With K=10, leaving out row 12, our fit MSE is [0.04778627] and test residual is 0.13734852696269212
With K=10, leaving out row 13, our fit MSE is [0.0263886] and test residual is 1.731660610828235
With K=10, leaving out row 14, our fit MSE is [0.03759717] and test residual is 0.5698622397415452
With K=10, leaving out row 15, our fit MSE is [0.0130878] and test residual is 24.811585954322258
With K=10, leaving out row 16, our fit MSE is [0.04749468] and test residual is 0.17148618993884154
With K=10, leaving out row 17, our fit MSE is [0.03156236] and test residual is 0.7586362229933148
With K=10, leaving out row 18, our fit MSE is [0.0480719] and test residual is 0.11037911546598976
With K=10, leaving out row 19, our fit MSE is [0.03834877] and test residual is 0.5721004697658572
====================
Across all leave-one-out runs for K=10:
- Our average test residual is 4.9954361094679465 and our average fit mse is 0.03800537095759345
With K=11, leaving out row 0, our fit MSE is [0.00701223] and test residual is 0.12408541397240291
With K=11, leaving out row 1, our fit MSE is [0.00677646] and test residual is 0.13013232220222026
With K=11, leaving out row 2, our fit MSE is [0.00713685] and test residual is 0.23080653389292416
With K=11, leaving out row 3, our fit MSE is [0.00747783] and test residual is 0.037883285156537605
With K=11, leaving out row 4, our fit MSE is [0.00711586] and test residual is 0.0943788184719673
With K=11, leaving out row 5, our fit MSE is [0.00500093] and test residual is 0.5324005555840042
With K=11, leaving out row 6, our fit MSE is [0.00644299] and test residual is 0.21195391677325875
With K=11, leaving out row 7, our fit MSE is [0.00730476] and test residual is 0.06885588374201584
With K=11, leaving out row 8, our fit MSE is [0.00685601] and test residual is 0.1327536730216144
With K=11, leaving out row 9, our fit MSE is [0.00735394] and test residual is 0.06183171865266912
With K=11, leaving out row 10, our fit MSE is [0.0052528] and test residual is 0.356320525271532
With K=11, leaving out row 11, our fit MSE is [0.0072613] and test residual is 11.762281911217348
With K=11, leaving out row 12, our fit MSE is [0.00742399] and test residual is 0.036296907701273196
With K=11, leaving out row 13, our fit MSE is [0.00747064] and test residual is 0.045745546688871414
With K=11, leaving out row 14, our fit MSE is [0.00675652] and test residual is 0.15897864757348912
With K=11, leaving out row 15, our fit MSE is [0.00746446] and test residual is 1.252130681609544
With K=11, leaving out row 16, our fit MSE is [0.00716579] and test residual is 0.09208359282144338
With K=11, leaving out row 17, our fit MSE is [0.00464293] and test residual is 0.3383257613512497
With K=11, leaving out row 18, our fit MSE is [0.00489747] and test residual is 0.250210328542111
With K=11, leaving out row 19, our fit MSE is [0.00724903] and test residual is 0.09485697961848333
====================
Across all leave-one-out runs for K=11:
```

- Our average test residual is 0.8006156501932402 and our average fit mse is 0.006703138983798042
=======================================
For this dataset and bucket of models:
- The ideal fit by test residual is K=1, which has a mean test residual of 0.3058136986029903
- The ideal fit by fit mse is 11, which has a mean fit mse of 0.006703138983798042

```
In [4]:  # (b) Identify the order of polynomial that yields the lowest average error. Is it the same as the
         # best fitting models from the previous problem? What does this tell you about predictive
         # modeling generally?

         # As shown in the output from the cell above, the order of polynomial that yields the lowest
         # mse in modeling is the tenth-order polynomial (11 terms, k=11). However, the order of polynomial
         # that yields the lowest average test residual is the zero-th-order polynomial (1 term, k=1).
         # This highlights that while the Legendre polynomials were effective at fitting the training data
         # effectively, they did not effectively capture the underlying physics.
```

HW 19

In this problem, we will revisit the data in bias variance trade.txt to demonstrate generalized aliasing. Remember that the first and third columns are samples of the independent variable x, while the second and fourth columns are samples of the dependent variable y that correspond to those points.

```
In [5]:  # (a) Combine the two samples into one dataset and generate a system matrix assuming a basis of
         # sines and cosines with frequency that increases with column number (i.e. cos(x),cos(2x),cos(3x),
         # etc.). Your system matrix should have 1,000 columns. Do not print this matrix.

         # a design matrix is created as follows:
         def design_mat(k: int, vec) -> np.ndarray:
             """ adapted from HW 18 """
             columns = []
             for i in range(k):
                 con_plus = i+2
                 con_pmod = con_plus%2
                 con_split = con_plus-con_pmod
                 constant = con_split/2
                 if i%2==0:
                     columns.append(np.sin(constant*vec))
                 else:
                     columns.append(np.cos(constant*vec))
```

```
    A = np.stack(columns).T
    return A


x_full = np.concatenate((x1, x2))
y_full = np.concatenate((y1, y2))
A_full = design_mat(1000, x_full)

# we won't print the full matrix, but just the first 5 columns of the first five rows to get a picture of
pd.DataFrame(A_full[:10, :10])
```

Out[5]:

	0	1	2	3	4	5	6	7	8	9
0	-0.160021	0.987114	-0.315917	0.948787	-0.463672	0.886007	-0.599476	0.800393	-0.719830	0.694150
1	-0.397202	0.917731	-0.729049	0.684462	-0.940940	0.338573	-0.998012	-0.063024	-0.890874	-0.454251
2	-0.472548	0.881305	-0.832918	0.553396	-0.995561	0.094116	-0.921867	-0.387506	-0.629330	-0.777138
3	-0.616901	0.787041	-0.971053	0.238866	-0.911614	-0.411047	-0.463902	-0.885886	0.181395	-0.983410
4	-0.380407	0.924819	-0.703615	0.710581	-0.921027	0.389499	-0.999951	0.009851	-0.928522	-0.371278
5	0.406080	0.913838	0.742182	0.670199	0.950388	0.311068	0.994818	-0.101667	0.867818	-0.496883
6	0.489518	0.871993	0.853712	0.520745	0.999345	0.036178	0.889132	-0.457650	0.551290	-0.834314
7	0.128753	0.991677	0.255363	0.966845	0.377722	0.925919	0.493793	0.869579	0.601645	0.798764
8	-0.319307	0.947651	-0.605183	0.796086	-0.827698	0.561173	-0.963556	0.267507	-0.998532	-0.054166
9	0.065082	0.997880	0.129887	0.991529	0.194142	0.980973	0.257574	0.966259	0.319914	0.947447

In [6]:
```
# (b) Use your programming language of choice to create an integer list of terms that are roughly
# equidistant in logarithmic space between 1 and 1,000. For each number of terms, produce
# a fit to the combined data set. Plot a representative subsample of these fits with the data
# used to produce them.

integer_list = np.logspace(1, 3, 3)
x_for_plotting = np.linspace(-1, 1, 1000)

plt.figure(figsize=(10, 10))
plt.scatter(x_full, y_full, marker="x", color="black")

for x_int in integer_list:
```
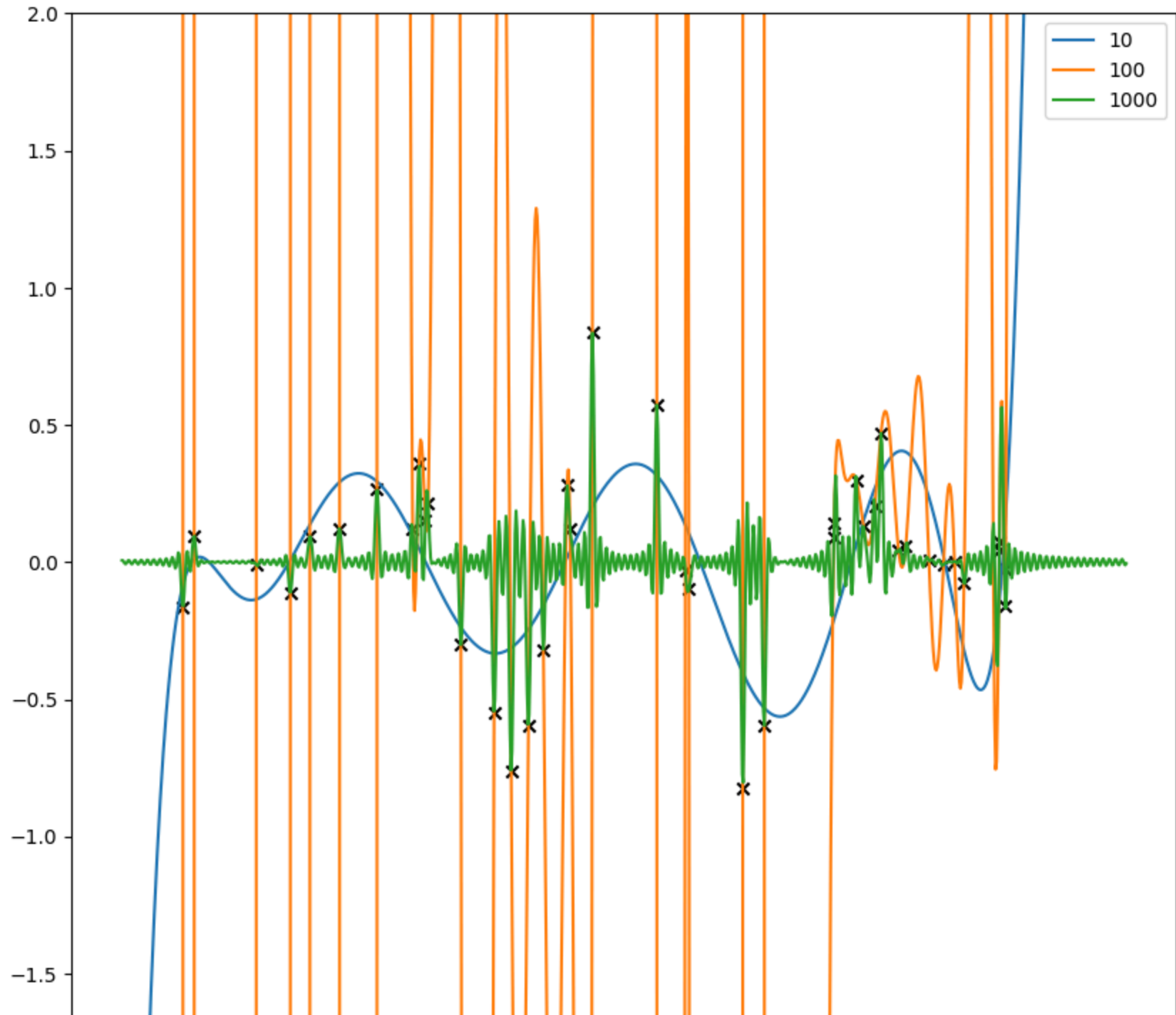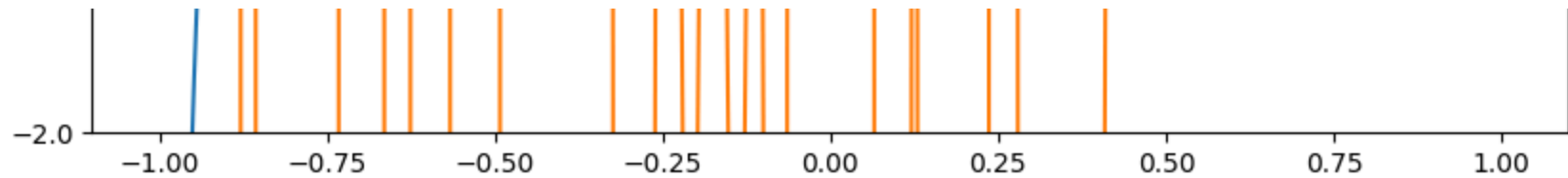
```
    A_tmp = design_mat(int(x_int), x_full)
    theta_hat, sse, rank, s = np.linalg.lstsq(A_tmp, y_full)
    A_tmp = design_mat(int(x_int), x_for_plotting)
    y_for_plotting = A_tmp@theta_hat
    plt.plot(x_for_plotting, y_for_plotting, label=str(int(x_int)))

plt.legend()
plt.ylim(-2, 2)
```

Out[6]: (-2.0, 2.0)

```
In [7]:  # (c) Assuming that the true signal is characterized by f(x) = e−6x2 cos(4πx), calculate the Eu−
         # clidean norm of the difference between the true signal and each of your models for 100 equally
         # spaced samples between −1 and 1. Report the cost versus number of terms in a log−log plot.

         integer_list = np.logspace(1, 3, 10) # in this plot, we'll use more models
         x_spaced = np.linspace(−1, 1, 100)
         true_y = np.exp(−6*(x_spaced**2))*np.cos(4*np.pi*x_spaced)
         costs = []

         for x_int in integer_list:
             x_int = int(x_int)
             A_tmp = design_mat(x_int, x_full)
             theta_hat, sse, rank, s = np.linalg.lstsq(A_tmp, y_full)
             A_tmp = design_mat(x_int, x_spaced)
             y_hat = A_tmp@theta_hat
             norm_tmp = np.linalg.norm(np.abs(y_hat−true_y), ord=2)
             costs.append(norm_tmp)

         plt.loglog(integer_list, costs, marker='o') # plotting call from chatGPT
```

Out[7]: [<matplotlib.lines.Line2D at 0x112973b10>]
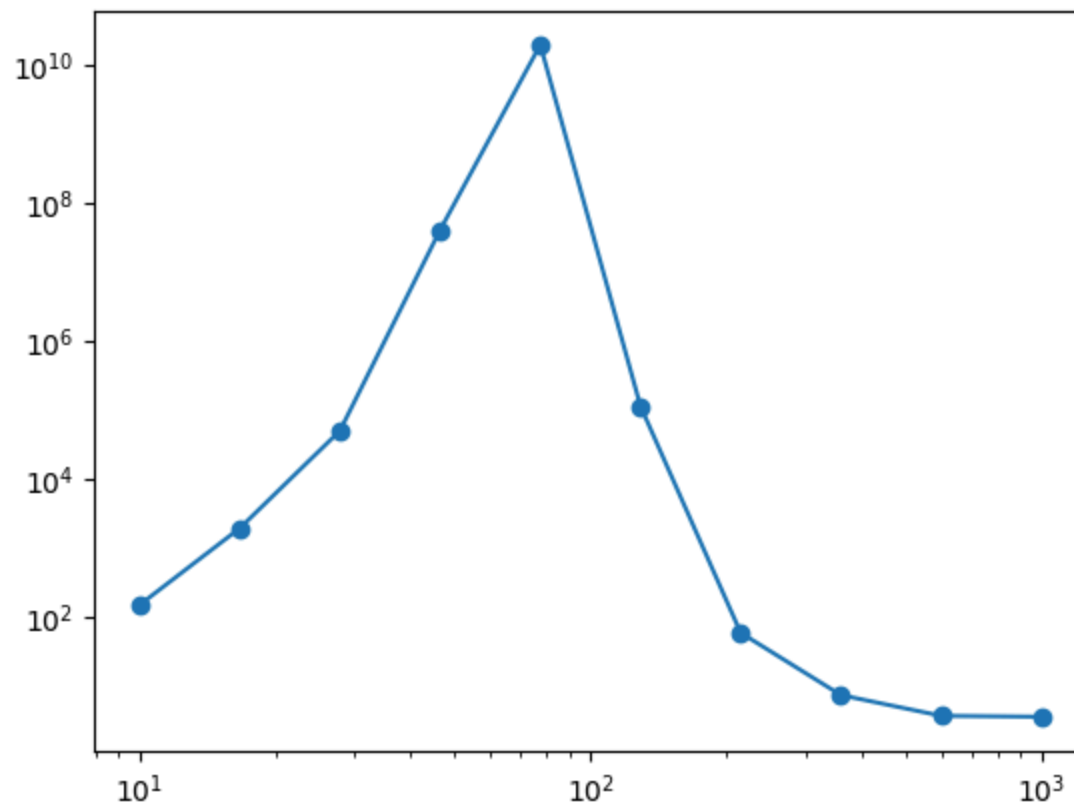
```
In [8]:   # (d) Using the language of the General Aliasing Decomposition, explain the shape of the plot
          # you produced in the prior part.

          # In our plot we can observe double descent, with an interpolation threshold at 10^2.
          # As we add additional/redundant parameters, they "absorb" the error to smooth out the pattern and
          # better fit the underlying true relationship
```

Acknowledgment

Work in this repository and with associated assignments and projects may be adapted or copied from similar files used in my prior academic and industry work (e.g., using a LaTeX file or Dockerfile as a starting point). Those files and any other work in this repository may have been developed with the help of LLM's like ChatGPT. For example, to provide context, answer questions, refine writing, understand function call syntax, and assist with repetitive tasks. In these cases, deliverables and

associated work reflect my best efforts to optimize my learning and demonstrate my capacity, while using available resources and LLM's to facilitate the process.

ChatGPT Conversation