Introduction to Machine Learning

ECE 580
Spring 2022

HW #2

Submission Instructions

Submit your work to the corresponding assignment in Gradescope. Although Gradescope accepts multiple file formats, they strongly recommend submitting your assignment as a single PDF file.

You are responsible for tagging the pages that correspond to each question. Pages may be tagged after submission, even if the submission deadline has passed. If you are submitting close to the submission deadline, submit your assignment first then immediately return to tag the pages.

When code is requested, submit a PDF print-out of your code. Submitting a URL for a cloud-based repository is insufficient.

Scoring Information

Individual questions will be scored holistically on the 9-point scale described in the syllabus. The homework assignment score will be the weighted average of the individual scores for each question. (The weight for each question is shown in parentheses to the left of the question number.)

Late Submissions

Late submissions will be accepted up to 5 days after the submission deadline, with the following point penalty applied if its late submission is not excused: [1]

1 day (0^+ to 24 hours) late: 0.2 point deduction

2 days (24^+ to 48 hours) late: 0.5 point deduction

3 days late: 0.8 point deduction

4 days late: 1.6 point deduction

5 days late: 3.2 point deduction

6 or more days late: score = 3.0 (not accepted for credit)

The late policy is designed to be minimally punitive for submissions up to 3 days late, yet encourage staying current with the coursework for our course by not allowing one assignment's late submission to overlap with the next assignment's submission.

A homework score will not drop below 3.0 as a result of applying the late penalty point deduction.

[1] One day = one 24-hour period or fraction thereof.

Structuring and Organizing Your Code

I encourage you to think about how to make your code modular and extensible. All of the regression results for this assignment can be generated using a single code block that accepts:

1. a given data set
2. the desired norm
   (this assignment uses only the L_2 norm; an extensible code block that allows for the norm to be specified could be helpful for you going forward because the L_1 norm is commonly used in regression, as the L_1 norm tends to be less sensitive to high leverage data points (data points with large residuals)).
3. the desired regularization
4. the desired cross-validation parameters

Cross-Validation

Cross-validation is used for both model selection and hyperparameter selection, to ensure the chosen model and/or hyperparameter(s) are not too highly tuned ("overfit") to the data. Here, you are going to explore the impact of cross-validation on your selection of a model that predict a car's price from its characteristics.

1. Continuing with the 13 continuous predictor variables from the Automobile Data Set from the UCI Machine Learning Repository that you used in Homework #1 to predict a car's price from its characteristics, you are going to further explore the 3 models you proposed in problem 2(a) in Homework #1.

(10) (a) For your proposed model #1, perform linear regression (L_2-norm) with 3-10-folds cross-validation (3 independent repetitions of 10-folds cross-validation) to evaluate the consistency in both the estimated model and the model performance. For each of the 30 independent regressions, save the model parameters and the mean-square error (MSE) for the model.

   i. Remind us what your proposed model #1 is.
   (Write down the equation $\hat{price} = f(\text{features}, \hat{\mathbf{w}})$, with the parameters $\hat{\mathbf{w}}$ unspecified.)

   ii. Plot a kernel density estimate for the mean square error (MSE) for this model. (Each fold yields an estimate of MSE; find the kernel density estimate using these 30 samples of MSE.)

   What is $E\{\text{MSE}\}$ for this model?

   Given the kernel density estimate for MSE, do you believe $E\{\text{MSE}\}$ is a representative summary statistic (a good approximation) for MSE?

   iii. Compute the *variance* for this model. (refer to Eqs. 2.26 and 2.28 in Machine Learning, by Liu) [2]

   iv. Determine the $(bias)^2 + noise\ variance$ for this model, assuming $E\{\text{MSE}\}$ is a good approximation for MSE. (Recall: $\text{MSE} = E\{(t-\hat{t})^2\} = variance + (bias)^2 + noise\ variance$)

   v. What is your impression of the *variance* of this model versus the $(bias)^2 + noise\ variance$ for this model? (i.e., does the model appear to have greater consistency and greater systematic error across data sets, or to have greater variability and less systematic error across data sets?)

[2] These equations are also provided in a companion pdf file.

(10) (b) Repeat the above steps for your proposed model #2.

i. Remind us what your proposed model #2 is.

ii. Plot a kernel density estimate for the mean square error (MSE) for this model.

What is $E\{\text{MSE}\}$ for this model?

Given the kernel density estimate for MSE, do you believe $E\{\text{MSE}\}$ is a representative summary statistic (a good approximation) for MSE?

iii. Compute the *variance* for this model.

iv. Determine the $(bias)^2 + noise\ variance$ for this model, assuming $E\{\text{MSE}\}$ is a good approximation for MSE.

v. What is your impression of the *variance* of this model versus the $(bias)^2 + noise\ variance$ for this model?

(10) (c) Repeat the above steps for your proposed model #3.

i. Remind us what your proposed model #3 is.

ii. Plot a kernel density estimate for the mean square error (MSE) for this model.

What is $E\{\text{MSE}\}$ for this model?

Given the kernel density estimate for MSE, do you believe $E\{\text{MSE}\}$ is a representative summary statistic (a good approximation) for MSE?

iii. Compute the *variance* for this model.

iv. Determine the $(bias)^2 + noise\ variance$ for this model, assuming $E\{\text{MSE}\}$ is a good approximation for MSE.

v. What is your impression of the *variance* of this model versus the $(bias)^2 + noise\ variance$ for this model?

(5) (d) What is your best guess as to how your 3 models are ordered from least complex (high bias, low variance) to most complex (low bias, high variance)? Explain how you arrived at your conclusion.[3]

(10) (e) Now that you have used cross-validation to evaluate the consistency of each model and as well as each model's performance, which of your three proposed models would you select? Why? Comment on how your conclusion is different from or similar to the conclusion you came to without cross-validation in Homework #1.

(5) (f) Submit a PDF print-out of your code. (Submitting a URL for a cloud-based repository is insufficient.)

[3] Developing an understanding of the trade-off between bias and variance, and how these two factors may influence your choice of model is a skill that is directly transferable to the "real world." Even if you may not explicitly evaluate the bias-variance trade-off as we are doing here, you will implicitly evaluate this trade-off when you compare and select models.

Regularization

Regularization is also used to ensure the chosen model and/or hyperparameter(s) are not too highly tuned ("overfit") to the data by constraining the regression so that the model parameters do not become too large. Here, you are going to explore the impact of regularization on your selection of a model that predict a car's price from its characteristics.

2. Continuing with the 13 continuous predictor variables from the Automobile Data Set from the UCI Machine Learning Repository that you used in Homework #1 to predict a car's price from its characteristics, you are going to continue to explore the 3 models you proposed in problem 2(a) in Homework #1.

(10) (a) For your proposed model #1, perform linear regression (L_2-norm) with L_2-norm regularization on the model weights (Ridge regression) and 3-10-folds cross-validation (3 independent repetitions of 10-folds cross-validation) to evaluate the consistency in both the estimated model and the model performance. Choose a weighting for the regularization term that seems reasonable to you. (You may want to experiment a bit with this value.) For each of the 30 independent regressions, save the model parameters and the mean-square error (MSE) for the model.

i. Remind us what your proposed model #1 is.
(Write down the equation $\hat{price} = f(\text{features}, \hat{\mathbf{w}})$, with the parameters $\hat{\mathbf{w}}$ unspecified.)

ii. Plot a kernel density estimate for the mean square error (MSE) for this model. (Each fold yields an estimate of MSE; find the kernel density estimate using these 30 samples of MSE.)

What is $E\{\text{MSE}\}$ for this model?

Given the kernel density estimate for MSE, do you believe $E\{\text{MSE}\}$ is a representative summary statistic (a good approximation) for MSE?

iii. Compute the *variance* for this model. (refer to Eqs. 2.26 and 2.28 in Machine Learning, by Liu)

iv. Determine the $(bias)^2 + noise\ variance$ for this model, assuming $E\{\text{MSE}\}$ is a good approximation for MSE. (Recall: $\text{MSE} = E\{(t - \hat{t})^2\} = variance + (bias)^2 + noise\ variance$)

v. What is your impression of the *variance* of this model versus the $(bias)^2 + noise\ variance$ for this model? (i.e., does the model appear to have greater consistency and greater systematic error across data sets, or to have greater variability and less systematic error across data sets?)

(10) (b) Repeat the above steps for your proposed model #2.

i. Remind us what your proposed model #2 is.

ii. Plot a kernel density estimate for the mean square error (MSE) for this model.

What is $E\{\text{MSE}\}$ for this model?

Given the kernel density estimate for MSE, do you believe $E\{\text{MSE}\}$ is a representative summary statistic (a good approximation) for MSE?

iii. Compute the *variance* for this model.

iv. Determine the $(bias)^2 + noise\ variance$ for this model, assuming $E\{\text{MSE}\}$ is a good approximation for MSE.

v. What is your impression of the *variance* of this model versus the $(bias)^2 + noise\ variance$ for this model?

(10) (c) Repeat the above steps for your proposed model #3.

i. Remind us what your proposed model #3 is.

ii. Plot a kernel density estimate for the mean square error (MSE) for this model.

What is $E\{\text{MSE}\}$ for this model?

Given the kernel density estimate for MSE, do you believe $E\{\text{MSE}\}$ is a representative summary statistic (a good approximation) for MSE?

iii. Compute the *variance* for this model.

iv. Determine the $(bias)^2 + noise\ variance$ for this model, assuming $E\{\text{MSE}\}$ is a good approximation for MSE.

v. What is your impression of the *variance* of this model versus the $(bias)^2 + noise\ variance$ for this model?

(5) (d) What is your best guess as to how your 3 models are ordered from least complex (high bias, low variance) to most complex (low bias, high variance)? Explain how you arrived at your conclusion. Comment on how the results using L_2 regularization (Ridge regression) differ from, or are similar to, the results using no regularization.

(10) (e) Now that you have used cross-validation to evaluate the consistency of each model as well as each model's performance under both L_2 regularization (Ridge regression) and no regularization, which model-regularization combination would you select? Why? (You have three proposed models and you have considered two regularizations, so there are 6 model-regularization combinations under consideration.)

(5) (f) Submit a PDF print-out of your code. (Submitting a URL for a cloud-based repository is insufficient.)