**Data 100, Spring 2024**

Exam Prep Section #7

# Cross-Validation

1. Justify whether the following statements are true or false.

   (a) Cross-validation always yields the true optimal solution when the loss function is convex.

   (b) Cross-validation requires the loss function to be differentiable.

   (c) Cross-validation should be used to solve for the optimal parameters for a linear regression model.

   (d) Cross-validation will always yield a similar error to the true test set error.

# The English Alphabet Goes $\alpha$, $|\beta|$, C(ross-validation)

2. Brandon is using stochastic gradient descent to optimize a model with 1,000 parameters that can detect vehicles and pedestrians in images. However, Brandon is having trouble choosing hyperparameters such as the learning rate, $\alpha$, and the batch size, $\beta$.

   (a) Suppose Brandon employs k-fold cross-validation to compute optimal values of the hyperparameters $\alpha$ and $|\beta|$. Which of the following is true?

   ◯ A. k-fold cross-validation will not yield useful information for selecting $\alpha$ or $|\beta|$, and is instead intended only for selecting the regularization parameter $\lambda$ for regularized models.

   ◯ B. k-fold cross-validation may yield suboptimal hyperparameters for $\alpha$ and $|\beta|$.

   ◯ C. k-fold cross-validation cannot output an optimal set of hyperparameters if the empirical risk is non-convex.

   ◯ D. To use cross-validation to find the optimal $\alpha$ or $|\beta|$, the loss must be differentiable with respect to $\alpha$ and $|\beta|$.

   (b) Regardless of your answer to the previous question, suppose Brandon uses k-fold cross-validation with $\alpha$ chosen from 0.1, 0.2, 0.4 and $|\beta|$ chosen from 32, 64, 128. The average cross-validated loss is shown in the below table for each combination of $\alpha$ and $|\beta|$.

| $\lvert\beta\rvert$ | $\alpha$ | | |
|---|---|---|---|
| | 0.1 | 0.2 | 0.4 |
| 32 | 0.0022 | 0.7031 | 0.0370 |
| 64 | 0.0051 | 0.9075 | 0.0471 |
| 128 | 0.0018 | 0.6007 | 0.0157 |

   What is the most optimal pair of $\alpha$ and $|\beta|$?

# Regularizing Model Adventures

3. Choose the best option for the following questions

   (a) Jessica is training a model and finds that as she adds more features her training error is decreasing along with her validation error. What should she do?

   ◯ A. Increase regularization.
   ◯ B. Decrease regularization.
   ◯ C. No additional regularization changes are needed.
   ◯ D. Not enough informatoin.

   (b) Jessica is training a model and her training error is rapidly decreasing but her validation error is increasing. What should she do?

   ◯ A. Increase regularization.
   ◯ B. Decrease regularization.
   ◯ C. No additional regularization changes are needed.
   ◯ D. Not enough information.

   (c) Suppose you are interested in finding the minimal set of explanatory features, which form of regularization would be most appropriate?

   ◯ A. L1 regularization.
   ◯ B. L2 regularization.
   ◯ C. No regularization.
   ◯ D. Not enough information.

   (d) Consider a simple setting in which we are predicting the height of a person in centimeters based on their weight. Suppose we include the weight measured in kilograms (kg) and milligrams (mg) as two separate features and we tune the coefficient of the L1 regularization to include only one feature.
   Without normalizing the data before training, which feature would be selected after the model is trained?

   ◯ A. Weight in mg
   ◯ B. Weight in kg
   ◯ C. Not enough information

# Extreme Tradeoffs

4. Rahul wants to model extreme precipitation events, which are historically difficult to predict accurately. To attempt to create the world's best model, he tries training multiple models using regularization.

   (a) He decides to change his models to add L2 regularization. What behavior is expected in the training set compared to the unregularized models?

   ☐ A. The model bias will decrease.

   ☐ B. The model bias will increase.

   ☐ C. The model variance will decrease.

   ☐ D. The model variance will increase.

   (b) Regardless of your answer to the previous question, assume that after implementing regularization, the model bias is too high. Which of these solutions helps reduce the model bias?

   ☐ A. Add an intercept term.

   ☐ B. Increase the regularization hyperparameter.

   ☐ C. Decrease the regularization hyperparameter.

   ☐ D. None of the above.

   (c) Assume that we fixed the previous issue by changing to a different unspecified regression model, and the model bias decreased. Which of the following could have happened as a result?

   ☐ A. The model variance increased.

   ☐ B. The model variance decreased.

   ☐ C. The model variance stayed the same.

   ☐ D. The observational variance decreased.