0/20 Questions Answered

Homework 2: Quiz

Q1 Regularization Functions

2 Points

Recall that the goal of a regularization function, $R(w)$, is to be large when the model (with weights w) is likely to overfit, and small otherwise. Which of the following functions achieve this goal?

Q1.1 Sum of Values

0.5 Points

$R(w) = \sum_{j=1}^{D} w_j$ achieves this goal.

- ○ True
- ○ False

Save Answer

Q1.2 Sum of Absolute Values

0.5 Points

$R(w) = \sum_{j=1}^{D} |w_j|$ achieves this goal.

- ○ True
- ○ False

Save Answer

Q1.3 Sum of Values Squared
0.5 Points

$R(w) = \sum_{j=1}^{D} w_j^2$ achieves this goal.

- True
- False

Save Answer

Q1.4 Square of Sum of Values
0.5 Points

$R(w) = \left(\sum_{j=1}^{D} w_j\right)^2$ achieves this goal.

- True
- False

Save Answer

Q2 Ridge Regression
4 Points

For these problems, assume we are using a sufficiently complex model, such that it is possible for it to overfit using its features. Recall that the quality metric (loss) used in Ridge Regression is:

$$MSE(w) + \lambda||w||_2^2$$

Q2.1 Large Penalty
1 Point

In ridge regression, choosing a very large penalty strength λ leads to a model with

- Very small coefficient magnitudes
- Very large coefficient magnitudes

Save Answer

Q2.2 Small Penalty

1 Point

In ridge regression, choosing a very small penalty strength λ tends to lead to a model with

- Low bias, High variance
- High bias, High variance
- Low bias, Low variance
- High Bias, Low variance

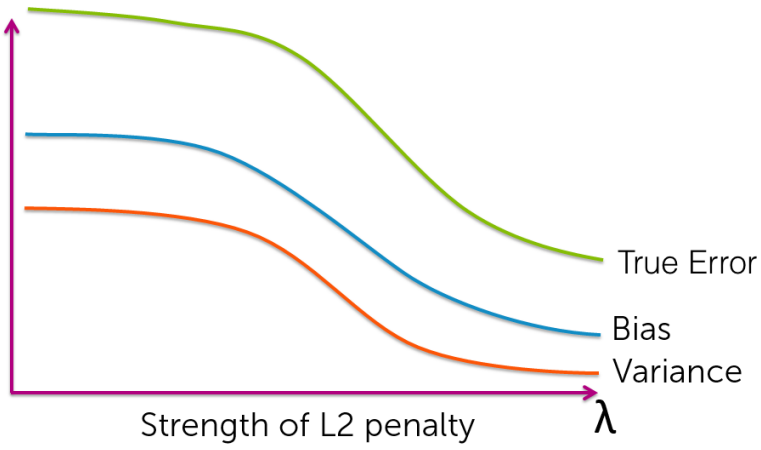
Save Answer

Q2.3 Error Curve

2 Points

Which of the following plots best characterize the trend of bias, variance, and true error as the regularization coefficient, λ, increases? The x-axis starts at $\lambda = 0$ and then shows increasing λ.

- Image 1



- Image 2

Image 2

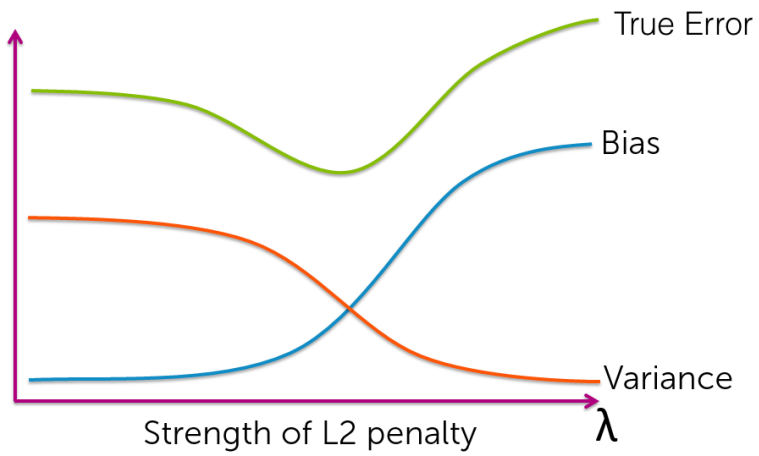


Image 3

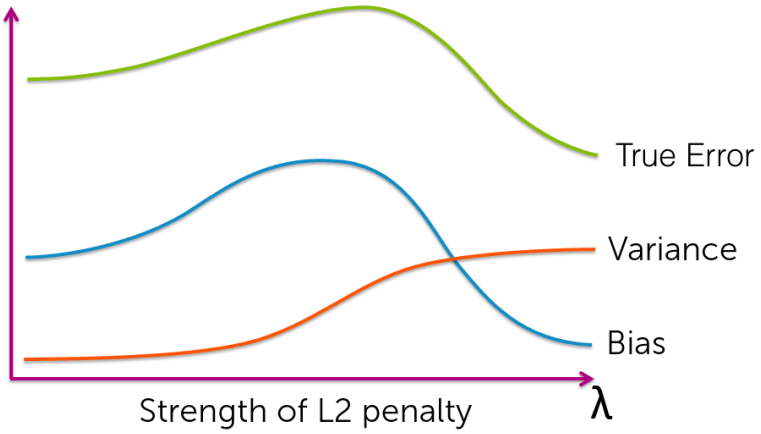
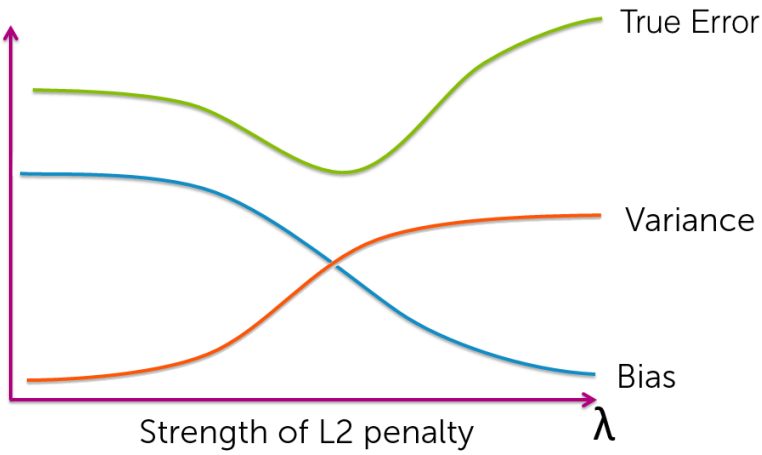


Image 4



Save Answer

Q3 Choosing Hyper-Parameters

2 Points

How should we select which value of λ to use for Ridge Regression?

- ○ Choose the setting of λ that has the smallest $MSE(\hat{w})$ on the **training set**
- ○ Choose the setting of λ that has the smallest $MSE(\hat{w})$ on the **test set**
- ○ Choose the setting of λ that has the smallest $MSE(\hat{w})$ on the **validation set**
- ○ Choose the setting of λ that has the smallest $MSE(\hat{w}) + \lambda||\hat{w}||_2^2$ on the **training set**
- ○ Choose the setting of λ that has the smallest $MSE(\hat{w}) + \lambda||\hat{w}||_2^2$ on the **test set**
- ○ Choose the setting of λ that has the smallest $MSE(\hat{w}) + \lambda||\hat{w}||_2^2$ on the **validation set**
- ○ Choose the setting of λ that results in the smallest coefficients.
- ○ Choose the setting of λ that results in the largest coefficients.

Save Answer

Q4 Using Test Set

1.5 Points

Your friend comes up to you to tell you that they thought of a new way to decide what the right model complexity is without using a validation set. Their idea:

> We don't need to use a validation set because we can just use the test set! Just do a train/test split like before and then choose the model with the lowest test error. That way, we'll have a larger train set.

Consider this approach to model selection when answering the below questions.

Q4.1
0.5 Points

The resulting model will be unlikely to overfit to the training data.

- ○ True
- ○ False

Save Answer

Q4.2
0.5 Points

The resulting model's test error will be an overly optimistic assessment of the true error.

- ○ True
- ○ False

Save Answer

Q4.3
0.5 Points

This approach to model selection is **not** generally recommended.

- ○ True
- ○ False

Save Answer

Q5 Cross Validation
1 Point

10-fold cross validation is more computationally expensive than leave-one-out (LOO) cross validation. Assume we have a really large dataset (i.e., billions of rows of data).

- True
- False

Save Answer

Q6 Feature Selection

1 Point

Suppose we have many features to use for a model and we are trying to train different models using different size subsets of the features, with the "All Subsets" approach from lecture. The best fit model of size 5 (i.e., with 5 features) always contains the set of features from the best fit model of size 4.

- True
- False

Save Answer

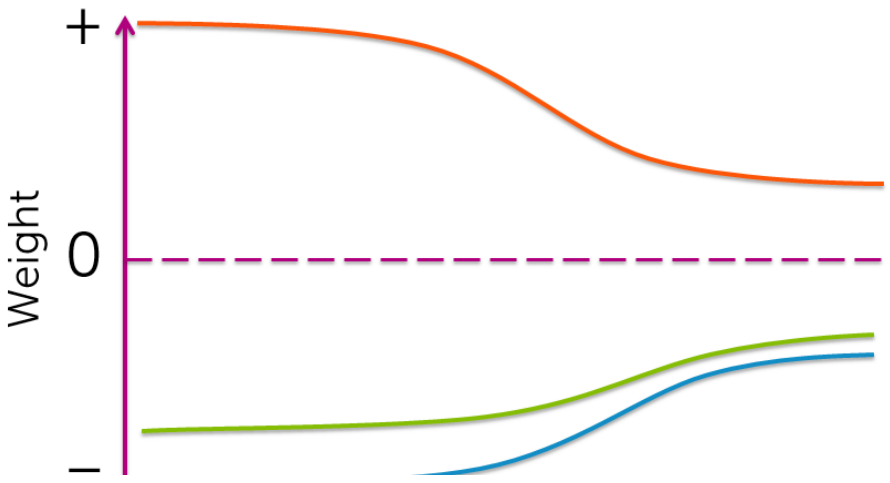
Q7 LASSO Coefficients

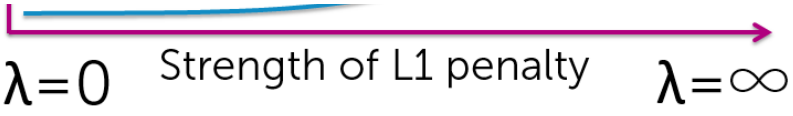
1 Point

Which plot could correspond to a LASSO coefficient path?

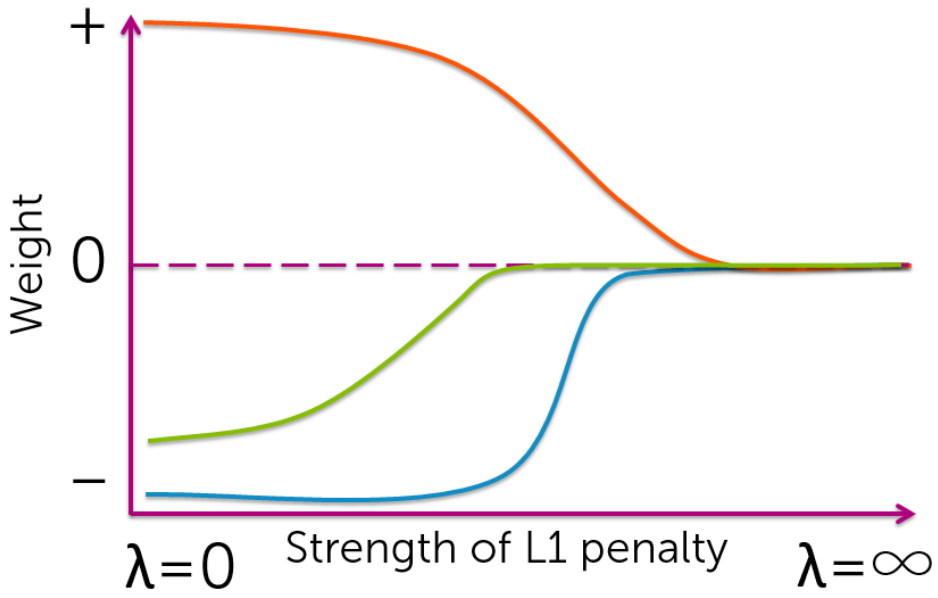
Hint: Notice we include when $\lambda = \infty$ in the plot as the right-most value of λ.

- Image 1

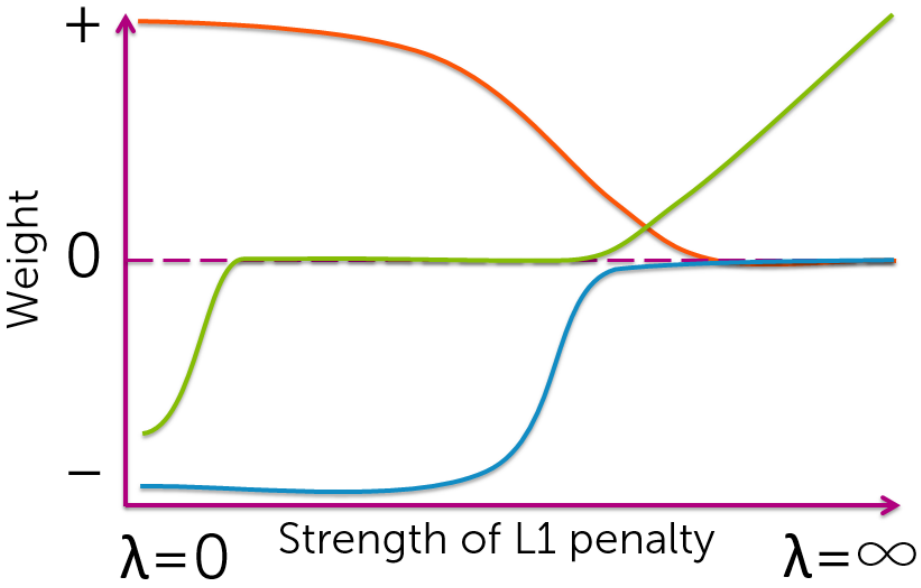




○ Image 2



○ Image 3



Save Answer

Q8 Comparing models

2 Points

Q8.1
1 Point

Select all the following options that apply

Which model would you use for feature selection:

- [ ] Linear Regression
- [ ] Ridge Regression
- [ ] LASSO Regression
- [ ] None of the above

Save Answer

Q8.2
1 Point

Which model leads to computationally more efficient predictors?

NOTE: This question is not referring to the computational efficiency of training the model. Instead, it is referring to the computational efficiency of using a trained model to predict outputs.

- ○ Linear Regression
- ○ Ridge Regression
- ○ LASSO Regression

Save Answer

Q9 ML Practitioner Scenarios
16 Points

Consider the below scenarios, and determine whether the practices are "Correct" or "Incorrect". If "Incorrect," please explain why.

NOTE: Please explicitly right "Correct" or "Incorrect." If "Correct," an explanation is not necessary.

Q9.1
4 Points

Wuwei's housing dataset has 1000 features, but she wants to use fewer features in her regression model. She decides to use Ridge Regression (L2-regularization) in order to identify and select the more prominent features.

Enter your answer here

Save Answer

Q9.2
4 Points

Max is aware of how time-consuming it is to find the globally optimal subset of features for a regression task. Therefore, he decides to use the forward stepwise algorithm, to significantly reduce the runtime, while still finding a good enough set of features.

Enter your answer here

Save Answer

Q9.3
4 Points

After training various models with LASSO regression (L1-regularization), Tanmay notices that the training error is almost 0, but the validation

error is greater than $10,000$. He thinks the best way to fix this is to decrease the regularization parameter λ.

Enter your answer here

Save Answer

Q9.4

4 Points

Karman's housing price dataset has a feature "area of the house." Some of the houses in his dataset are from America and use "square feet," while others are from France and use "square meters." To account for this, Karman decides to normalize the feature.

Enter your answer here

Save Answer

Save All Answers

Submit & View Submission >