

QF624-2025-W7

Number of participants: 26



**If a model is considered low Bias
1. and High Variance, which description
is usually correct for the model?**

9 correct answers
out of 16 respondents

Complex model;
underfitting



3 votes



Complex model;
overfitting



9 votes

Simple model;
underfitting



4 votes

Simple model;
overfitting



0 votes



2. **When we have very big data set, we usually prefer complex model**

10 correct answers
out of 15 respondents



True



10 votes

False



5 votes

You're comparing two different models on the same prediction task: Model A: A very simple linear model. Model B: A highly flexible model with many parameters (for example, a deep decision tree).



3. **After training both models on the same data, you notice: Both models achieve similar error on the training set. On new (unseen) data, Model A's error is lower than Model B's error. Which of the following explanations is MOST LIKELY correct?**

3 correct answers
out of 15 respondents



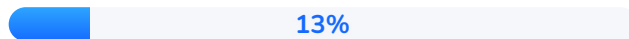
Model B has lower bias than Model A, but its predictions vary so much from sample to sample that its error on new data ends up larger.



20%

3 votes

Model B must have higher bias than Model A, since its error on new data is larger.



13%

2 votes

Because both models fit the training data equally well, they have the same amounts of bias and variance.



0%

0 votes

Model A underfits (makes overly simple assumptions), and Model B overfits (follows noise in the training data); hence Model A generalizes better despite being simpler.



67%

10 votes



4.

Suppose you plot test risk versus the number of parameters in a neural network. You observe a U-shaped curve at first (classical bias–variance tradeoff), followed by a rise around the interpolation threshold, and then a second decline for extremely wide networks. Which explanation best captures why test risk declines again in the “overparameterized” region?

8 correct answers
out of 12 respondents

Overparameterized models always memorize noise perfectly, so they generalize poorly.

0%

0 votes

Extremely large models find “simplest” interpolating solutions (e.g., minimum-norm weights) that generalize well, reducing overfitting.

67%

8 votes

Having more parameters forces the model to underfit, which lowers variance.

25%

3 votes

The second descent happens solely because of using a different optimization algorithm in that regime.

8%

1 vote



5. If penalty weight $f(x)$ is large, all $f(x)$ are restricted to a small space

8 correct answers
out of 10 respondents



True



8 votes

False



2 votes



Consider fitting a linear model to standardized predictors (mean zero, unit variance) using either Lasso or Ridge regression. Which of the following statements about how each method handles correlated predictors is most accurate?

9 correct answers
out of 13 respondents

Lasso tends to distribute nonzero coefficients evenly among a group of highly correlated predictors, whereas Ridge tends to pick only one predictor from the group and set others to zero.



3 votes

✓ Ridge regression tends to shrink coefficients of correlated predictors toward each other (the "grouping effect"), while Lasso often selects exactly one variable from a set of highly correlated predictors and zeros out the rest.



9 votes

Both Lasso and Ridge always assign exactly the same coefficients to perfectly correlated predictors.



1 vote

Lasso and Ridge both perform variable selection by shrinking some coefficients exactly to zero when predictors are highly correlated.



0 votes

You train a Ridge model and a Lasso model on a fixed dataset and then evaluate them on new data.



- 7. Which statement about model stability (i.e., how much the fitted coefficients change if you slightly perturb the training set) is true?**

6 correct answers
out of 12 respondents

Lasso is generally more stable than Ridge, because once it zeroes a feature, small changes in data won't bring it back.



33%

4 votes



Ridge is generally more stable than Lasso, because Ridge's squared penalty produces gradual coefficient changes, while Lasso's piecewise nature makes it jumpy.



50%

6 votes

Both Lasso and Ridge are equally stable, since they both penalize large coefficients.



17%

2 votes

Neither method is stable; both fit arbitrary coefficients that can swing wildly with small data changes.



0%

0 votes

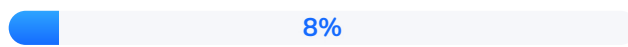
You have three scenarios: The true signal uses only a few predictors (sparse). All p predictors contribute modestly (dense). Many predictors are noisy, but the small true signals are highly correlated. Based on bias–variance tradeoff, which method—Lasso or Ridge—is most suited for each scenario?

9 correct answers
out of 12 respondents



8. are noisy, but the small true signals are highly correlated. Based on bias–variance tradeoff, which method—Lasso or Ridge—is most suited for each scenario?

(1) Lasso ; (2)
Lasso ; (3) Lasso



1 vote

(1) Ridge ; (2)
Ridge ; (3) Ridge



0 votes



(1) Lasso ; (2)
Ridge ; (3) Ridge



9 votes

(1) Ridge ; (2)
Lasso ; (3) Lasso



2 votes



9. Which option is FALSE about train-validation-test method?

6 correct answers
out of 11 respondents

Testing set can only be used once at the end



2 votes

Testing set should contain a wide range of different kinds of observations



2 votes

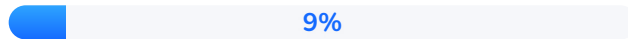


Validation set can only be used once for training



6 votes

Validation set can be used to fine-tune model parameters



1 vote



When one input variable 'area'
10. goes from square meter to square feet, what is the implication?

2 correct answers
out of 10 respondents

Value of 'area'
goes down; less
penalty on
coefficient of 'area'



20%

2 votes

Value of 'area'
goes up; more
penalty on
coefficient of 'area'



50%

5 votes

Value of 'area'
goes down; more
penalty on
coefficient of 'area'



10%

1 vote

Value of 'area'
goes up; less
penalty on
coefficient of 'area'



20%

2 votes



11. What is NOT the purpose of regularization?

0 correct answer
out of 0 respondent

To address overfitting to the training data set

0%

0 votes

To prevent overly complicated model

0%

0 votes

To reduce the coefficients of input variables that are not very relevant

0%

0 votes



To find the perfect model for the training data set

0%

0 votes



12. You want to predict a continuous outcome using a family of models. You compare a very simple model class (e.g., linear functions) versus a very flexible class (e.g., deep neural networks). Which statement best distinguishes approximation error from estimation error?

0 correct answer
out of 0 respondent

Approximation error comes from not having enough data; estimation

0%

0 votes

error comes from choosing a model that is too flexible.

Approximation error is the gap between the best possible predictor in your chosen model class and the true underlying pattern; estimation error is the extra “wobble room” because you only saw a finite sample of data.



0%

0 votes

Approximation error always decreases as you collect more data; estimation error always increases as you collect more data.

0%

0 votes

Approximation error is eliminated by tuning hyperparameters; estimation error is eliminated by adding more hidden layers

0%

0 votes