

1. What is the main objective of regularization in a machine learning model?
 - (a) Reduce the model's variance
 - (b) Improve the model's ability to overfit
 - (c) Increase the number of parameters
 - (d) Maximize the variance of errors
2. In the context of cross-validation, why is it necessary to tune the regularization hyperparameter?
 - (a) To improve the model's fit to the training set
 - (b) To find the best trade-off between bias and variance
 - (c) To increase the complexity of the model
 - (d) To speed up model training
3. What is the effect of a very large lambda (λ) in Ridge regularization?
 - (a) The model becomes more complex
 - (b) The model becomes simpler, with lower variance
 - (c) The model stops learning completely
 - (d) The model converges to a local maximum
4. What happens if the regularization parameter (λ) in Lasso regression is set too high?
 - (a) The model captures too much noise from the data
 - (b) Most coefficients are shrunk to zero, leading to an overly simplified model
 - (c) The model retains all parameters but with reduced values
 - (d) The model becomes more likely to overfit

Transformation of inputs :

5. When applying one-hot encoding to a categorical variable with 4 levels, how many new predictors (features) will be created if you set one level as the reference?
 - (a) 3
 - (b) 4
 - (c) 5
 - (d) 2
6. Standardization transforms data by:
 - (a) Shifting the mean to 0 and scaling the standard deviation to 1
 - (b) Normalizing the data to a range between 0 and 1
 - (c) Removing outliers
 - (d) Converting categorical variables into numerical form
7. If a predictor represents an angle (e.g., 0° to 360°), what is the recommended transformation to respect neighborhood relationships?
 - (a) One-hot encoding
 - (b) Square root transformation

- (c) Use sine and cosine transformations
- (d) Log transformation

Gradient descent :

8. In gradient descent, what happens if the learning rate is too high?
 - (a) The model converges faster
 - (b) The model may not converge and oscillate around the minimum
 - (c) The model converges but to the wrong minimum
 - (d) The model stops learning
9. What is a key advantage of stochastic gradient descent (SGD) over standard gradient descent?
 - (a) SGD guarantees better accuracy
 - (b) SGD can find the global minimum for non-convex loss functions
 - (c) SGD typically converges faster by updating after each batch
 - (d) SGD avoids overfitting more effectively
10. Why is early stopping used in training models with gradient descent?
 - (a) To reduce overfitting by stopping before the model fits the training data too closely
 - (b) To ensure the model reaches the global minimum
 - (c) To adjust the learning rate dynamically
 - (d) To allow the model to train indefinitely until perfect accuracy is achieved
11. True or false: Linear regression can always capture the structure in any dataset if enough data points are used.
12. True or false: In K-nearest neighbor models, increasing the number of neighbors (K) always leads to a decrease in both bias and variance.
13. In the bias-variance trade-off, increasing the flexibility of a model tends to:
 - (a) Increase bias and decrease variance
 - (b) Increase both bias and variance
 - (c) Decrease bias and increase variance
14. In K-nearest neighbor models, what is the main consequence of using a very large value for K?
 - (a) The model will have low bias and high variance.
 - (b) The model will smooth over local data patterns, potentially leading to underfitting.
 - (c) The model will overfit the data by memorizing the training set.
15. True or false: For hyper-parameter selection, cross-validation is less prone to variability than the validation set approach.
16. Given a dataset of 10000 datapoints, the dataset is split with 70% train, 10% validation and 20% test sets in a nested cross-validation manner. How many different splits of the data are performed in the process?

17. You can create more quiz questions and even conceptual exercises using generative AI, such as ChatGPT or Claude. A useful strategy is to provide examples of quiz questions or exercises from the course or to say that you would like to have exercises in the style of the textbook "Introduction to Statistical Learning" and then specify on which specific topic you would like to have an exercise, e.g. "gradient descent", "multi-layer perceptrons", etc. You can also specify the difficulty level.

CAVEAT: Unfortunately, the quality of exercises and solutions produced by some generative AI is highly variable. Class colleagues may be better at generating good quiz questions and exercises :)

Solutions

1. a, 2. b, 3. b, 4. b, 5. a, 6. a, 7. c, 8. b, 9. c, 10. a, 11. false, 12. false, 13. c, 14. b, 15. true, 16. 40.