1. . . . the weights may be reduced to zero.

(a) L1 and L2 (b) L1 (c) L2 (d) None of the above

2. If we want to create an optimal set of weight, we choose to minimize this loss function concerning w over the entire training data set.

3. What is the primary purpose of regularization in deep learning?

(a) to increase computational efficiency
(b) to reduce the number of layers in a neural network
(c) to prevent overfitting
(d) to speed up the training process

4. Which of the following regularization techniques adds a penalty term based on the absolute values of the weights?

(a) L1 regularization (b) L2 regularization (c) Dropout (d) Elastic Net

5. In neural networks, what does L2 regularization encourage?

(a) Sparse weight matrices (b) large weight values
(c) small weight values (d) No impact on weight values

6. How does dropout regularization work in a neural network?

(a) It randomly drops input features during training
(b) It randomly drops entire layers during training
(c) It adds noise to the input data
(d) It introduces a penalty term for large weights.

7. Which regularization technique combines both L1 and L2 penalties?

(a) Dropout (b) Ride regression
(c) Elastic Net (d) Batch Normalization

8. What is the purpose of early stopping as a form of regularization?

(a) To stop the training process when the model is underfitting
(b) To prevent the model from memorizing the training data
(c) To speed up the convergence of the training process
(d) To reduce the impact of outliers in the training data

9. Which of the following statements is true about the bias-variance tradeoff in the context of regularization?

(a) Regularization always increases bias and decreases variance
(b) Regularization always increases both bias and variance
(c) Regularization can help balance bias and variance
(d) Regularization has no impact on the bias-variance tradeoff

10. In the context of neural networks, what does weight decay refer to?

(a) The gradual increase in weight values during training
(b) The gradual decrease in weight values during training
(c) The removal of unnecessary weights from the network
(d) The introduction of noise to the weight values

11. Which of the following is a disadvantage of using a high regularization strength in a neural network?

(a) Increased risk of overfitting
(b) Faster convergence during training
(c) Enhanced generalization to new data
(d) Reduced capacity to capture complex patterns

12. What is the role of the temperature parameter in the context of knowledge distillation as a form of regularization?

- (a) Controls the learning rate
- (b) Adjusts the level of noise in the input data
- (c) Regulates the softness of the target distribution
- (d) Sets the threshold for dropout during training

13. In the context of neural networks, what does dropout rate refer to?

- (a) The percentage of training samples used during each iteration
- (b) The rate at which weight are decayed during training
- (c) The probability of dropping out a unit in the hidden layers during training
- (d) The learning rate for stochastic gradient descent.

14. Which of the following is a technique used for dynamic adjustment of the learning rate during training to improve convergence in deep learning?

- (a) Adversarial training
- (b) Learning rate annealing
- (c) Batch Normalization
- (d) Feature Scaling

15. What is the purpose of adding noise to the input data as a form of regularization?

- (a) To make the training process deterministic
- (b) To improve model interpretability
- (c) To reduce the impact of outliers in the input data
- (d) To prevent the model from memorizing the training data

16. In the context of regularization, what does the term "shrinkage" refer to?

- (a) Reducing the size of the input data
- (b) Reducing the number of hidden layers in the network
- (c) Constraining the magnitude of the weights in the model
- (d) Eliminating unnecessary features from the dataset

17. Which of the following statements is true about the dropout technique?

- (a) Dropout is more effective in shallow networks than deep networks
- (b) Dropout can be applied only to input layers
- (c) Dropout introduces random variations only during testing
- (d) Dropout helps prevent co-adaptation of hidden units

18. What is the primary goal of ensemble methods in machine learning?

- (a) To reduce the computational complexity of models
- (b) To increase the training time of individual models
- (c) To improve the predictive performance of a model by combining multiple models
- (d) To decrease the diversity among base models

19. Which of the following statements is true about bagging (Bootstrap Aggregating)?

- (a) It trains multiple models sequentially.
- (b) It trains multiple models independently on different subsets of the training data.
- (c) It combines models using a weighted average.
- (d) It is not suitable for high-variance models.

20. What is the purpose of random forests in ensemble learning?

- (a) To create a forest of decision trees with high correlation
- (b) To reduce the number of trees in the ensemble
- (c) To introduce randomness by considering a random subset of features for each tree
- (d) To eliminate the need for decision trees in the ensemble

21. In boosting, how are the weights assigned to misclassified instances during training?

- (a) Equally to all instances
- (b) Proportional to the difficulty of the instance
- (c) Sequentially, with higher weights for misclassified instances
- (d) Inversely proportional to the number of features

22. Which ensemble method combines the predictions of base models by taking a weighted average, where the weights are learned based on the performance of each model?

(a) Bagging (b) Stacking (c) Boosting (d) Random Forest

23. What is the primary advantage of ensemble methods over individual base models?

- (a) Ensemble methods are always faster than individual models.
- (b) Ensemble methods can handle only linear relationships.
- (c) Ensemble methods often generalize better and have improved robustness.
- (d) Ensemble methods are more prone to overfitting.

24. In the context of boosting, what does the term "weak learner" refer to?

- (a) A model with high training accuracy
- (b) A model that performs slightly better than random chance
- (c) A model with a large number of parameters
- (d) A model that is highly overfit

25. Which ensemble method is known for building a sequence of weak learners, each correcting the errors of its predecessor?

(a) Bagging (b) AdaBoost (c) Random Forest (d) Gradient Boosting

26. Which ensemble method trains multiple models independently on different subsets of the training data?

(a) Boosting (b) Stacking (c) Bagging (d) Random Forest

27. What is bagging short for in the context of ensemble methods?

(a) Bootstrap Aggregating (b) Boosting Algorithm (c) Bagged Aggregation (d) Batch Aggregation

28. In boosting, how are the weights assigned to misclassified instances during training?

- (a) Equally to all instances
- (b) Proportional to the difficulty of the instance
- (c) Sequentially, with higher weights for misclassified instances
- (d) Randomly assigned

29. Which ensemble method combines the predictions of base models by taking a weighted average?

- (a) Bagging
- (b) Stacking
- (c) Boosting
- (d) Random Forest

30. Which ensemble method is known for building a sequence of weak learners, each correcting the errors of its predecessor?

(a) Bagging
(b) AdaBoost
(c) Random Forest
(d) Gradient Boosting

31. What is the primary advantage of ensemble methods over individual base models?

(a) Faster training time
(b) Improved generalization and robustness
(c) Lower computational complexity
(d) Higher sensitivity to outliers

32. Which ensemble method is based on constructing a forest of decision trees with high diversity?

(a) Bagging (b) AdaBoost (c) Random Forest (d) Stacking

33. What does the acronym "LSTM" stand for in the context of deep learning?

(a) Long Short-Term Memory (b) Linear Short-Term Memory
(c) Limited Short-Term Memory (d) Lasting Short-Term Memory

34. In boosting, what is the purpose of the learning rate parameter?

(a) It controls the number of weak learners It adjusts the amount by which weights are updated during each iteration
(b) It determines the depth of decision trees
(c) It sets the threshold for feature selection

35. What distinguishes Random Forest from traditional bagging techniques?

(a) Random Forest uses a single decision tree
(b) Random Forest trains models sequentially
(c) Random Forest introduces randomness by considering a random subset of features for each tree
(d) Random Forest assigns equal weights to all instances

36. How does stacking differ from bagging and boosting in ensemble methods?

(a) Stacking trains models independently on different subsets
(b) Stacking combines predictions using a weighted average
(c) Stacking builds a sequence of weak learners
(d) Stacking uses multiple base models to form a meta-model

37. What role does the concept of "bias-variance tradeoff" play in ensemble methods?

(a) Ensemble methods eliminate the bias-variance tradeoff
(b) Ensemble methods intensify the bias-variance tradeoff
(c) Ensemble methods help balance bias and variance
(d) Ensemble methods have no impact on bias and variance

38. What is the primary limitation of using too many weak learners in boosting?

(a) Increased risk of overfitting (b) Decreased computational complexity
(c) Improved generalization (d) Faster training time

39. In bagging, how are the subsets of the training data created for each base model?

(a) Randomly and with replacement
(b) Randomly and without replacement
(c) Sequentially and with replacement
(d) Sequentially and without replacement

40. What is the primary advantage of using gradient boosting over traditional AdaBoost?

(a) Faster convergence
(b) Better handling of outliers
(c) Reduced risk of overfitting
(d) Simplicity in implementation

41. Which ensemble method is prone to becoming computationally expensive as the number of models increases?

(a) Bagging
(b) Stacking
(c) Boosting
(d) Random Forest

42. What does the term "stacking" refer to in ensemble learning?

(a) Combining models using a weighted average
(b) Training models independently on different subsets
(c) Constructing a sequence of weak learners
(d) Using multiple base models to form a meta-model

43. Which ensemble method is known for its ability to handle both linear and non-linear relationships in the data?

(a) Bagging
(b) Stacking
(c) Random Forest
(d) Gradient Boosting

44. Explain the concept of "out-of-bag" error in the context of bagging.

(a) It is the error rate calculated on the training set
(b) It is the error rate on the validation set
(c) It is an estimate of the test error obtained from the unused samples during training
(d) It is a measure of the model's performance on out-of-distribution data

45. What is the role of the hyperparameter "max depth" in decision trees within a Random Forest?

(a) It controls the number of trees in the forest
(b) It limits the maximum depth of individual decision trees
(c) It sets the learning rate for boosting
(d) It adjusts the weights assigned to misclassified instances

46. In the context of ensemble methods, what is "early stopping," and how does it contribute to regularization?

(a) Early stopping involves terminating the training process when the model is underfitting, contributing to model simplicity.
(b) Early stopping prevents overfitting by stopping the training process when the model starts to memorize the training data.
(c) Early stopping introduces noise to the input data during training, preventing overfitting.
(d) Early stopping is not related to regularization in ensemble methods.

47. What is the impact of increasing the number of base models on the computational complexity of stacking?

(a) The computational complexity decreases linearly
(b) The computational complexity increases linearly
(c) The computational complexity remains constant
(d) The computational complexity depends on the type of base models used

48. Explain the concept of "adversarial training" in the context of ensemble methods.

(a) Adversarial training involves training models to be robust against adversarial attacks.
(b) Adversarial training focuses on maximizing the accuracy on the training set.
(c) Adversarial training eliminates the need for ensemble methods.
(d) Adversarial training refers to using adversarial examples as additional training data.

49. How does the concept of "stacking with cross-validation" address the risk of overfitting in stacking?

(a) It eliminates the need for cross-validation in stacking.
(b) It uses multiple cross-validated models, reducing overfitting.
(c) It increases the depth of individual base models.
(d) It has no impact on the risk of overfitting.

50. What is the primary drawback of using a high learning rate in boosting algorithms?

(a) Slower convergence
(b) Increased risk of overfitting
(c) Decreased model performance
(d) Improved generalization

51. Explain the concept of "feature importance" in the context of Random Forest.

(a) Feature importance represents the number of times a feature is selected by a base model.
(b) Feature importance indicates the relevance of a feature in predicting the target variable.
(c) Feature importance is not applicable to ensemble methods.
(d) Feature importance measures the computational cost of using a specific feature.

52. What is the role of the "n estimators" hyperparameter in ensemble methods such as Random Forest and Gradient Boosting?

(a) It controls the learning rate in boosting algorithms.
(b) It sets the maximum depth of individual decision trees.
(c) It specifies the number of base models in the ensemble.
(d) It determines the subset of features considered for each base model.

53. Explain the concept of "stacking with meta-features" in the context of ensemble methods.

(a) Stacking with meta-features involves using the output of base models as features for a meta-model.
(b) Stacking with meta-features eliminates the need for multiple base models.
(c) Stacking with meta-features refers to combining models using a weighted average.
(d) Stacking with meta-features involves using only one type of base model in the ensemble.

54. What is Dropout in the context of neural networks?

(a) Adding noise to input features
(b) Removing random neurons during training
(c) Reducing the learning rate
(d) Increasing the number of hidden layers

55. What is the main purpose of Dropout in neural networks?

(a) To increase overfitting
(b) To speed up the training process
(c) To prevent co-adaptation of neurons
(d) To eliminate the need for activation functions

56. Which of the following statements is true about the application of Dropout during training?

(a) Dropout is only applied to input layers
(b) Dropout is applied to all layers except the output layer
(c) Dropout is applied during both training and testing
(d) Dropout is never applied to neural networks

57. How does Dropout contribute to regularization in neural networks?

(a) By increasing the number of parameters
(b) By introducing noise to the input data
(c) By reducing the model's capacity
(d) By promoting co-adaptation of neurons

58. In terms of training, what does it mean if a neuron is "dropped out"?

(a) The neuron's weights are set to zero
(b) The neuron is removed from the network temporarily
(c) The neuron's activation function is bypassed
(d) The neuron's output is squared

59. What challenge does Dropout aim to address in neural networks?

(a) Underfitting (b) Overfitting (c) Vanishing gradients (d) Exploding gradients

60. How does Dropout affect the training time of a neural network?

(a) Slows down the training process
(b) Speeds up the training process
(c) No impact on training time
(d) Depends on the type of activation function used

61. What is the recommended range for Dropout rates in neural networks?

(a) 0.0 to 0.1
(b) 0.2 to 0.5
(c) 0.5 to 0.8
(d) 0.9 to 1.0

62. How does Dropout contribute to model generalization?

(a) By memorizing the training data
(b) By promoting co-adaptation of neurons
(c) By reducing the sensitivity of neurons to specific input features
(d) By increasing the number of hidden layers

63. When applying Dropout, which phase is used for adjusting the weights of the neural network?

(a) Training phase
(b) Testing phase
(c) Both training and testing phases
(d) Neither training nor testing phases

64. Explain the term "co-adaptation of neurons" in the context of neural networks and how Dropout addresses it.

(a) Co-adaptation refers to neurons relying too much on each other, and Dropout breaks these dependencies by randomly dropping neurons during training.
(b) Co-adaptation is a form of regularization, and Dropout exacerbates co-adaptation by introducing noise.
(c) Co-adaptation occurs when neurons are independent, and Dropout enforces co-adaptation by removing dependencies.
(d) Co-adaptation is unrelated to Dropout; Dropout only affects the learning rate.

65. How does the effectiveness of Dropout vary with the size and complexity of a neural network?

(a) Dropout is more effective in small and simple networks
(b) Dropout is more effective in large and complex networks
(c) Dropout is equally effective across all network sizes and complexities
(d) Dropout is irrelevant to network size and complexity

66. What is the relationship between Dropout and the concept of ensemble learning?

- (a) Dropout is a type of ensemble learning
- (b) Ensemble learning and Dropout are unrelated concepts
- (c) Dropout and ensemble learning achieve the same result in terms of model diversity
- (d) Dropout eliminates the need for ensemble learning

67. Explain the trade-off between using a high Dropout rate and a low Dropout rate in neural networks.

- (a) High Dropout rates lead to overfitting, while low Dropout rates may result in underfitting.
- (b) High Dropout rates always improve model generalization, while low Dropout rates reduce model capacity.
- (c) There is no trade-off; the Dropout rate does not impact model performance.
- (d) The trade-off depends on the type of activation function used in the network.

68. How does Dropout contribute to mitigating the vanishing gradient problem in deep neural networks?

- (a) a. By increasing the learning rate
- (b) By preventing co-adaptation of neurons
- (c) By introducing noise to the input data
- (d) By reducing the sensitivity of neurons to specific input features

69. What is the primary goal of data augmentation in machine learning?

- (a) To decrease the size of the dataset
- (b) To increase the computational complexity
- (c) To improve model performance by increasing the diversity of the training data
- (d) To eliminate the need for validation data

70. Which of the following is a common technique used in data augmentation for image data?

- (a) Principal Component Analysis (PCA)
- (b) Feature scaling
- (c) Image rotation
- (d) Lasso regularization

71. How does data augmentation contribute to preventing overfitting in machine learning models?

- (a) By reducing the size of the training dataset
- (b) By increasing the number of layers in the model
- (c) By introducing noise to the input data
- (d) By providing a more diverse set of training examples

72. In text data augmentation, what technique involves replacing words with their synonyms?

- (a) Tokenization
- (b) Embedding
- (c) Word substitution
- (d) Lemmatization

73. Which of the following is a disadvantage of data augmentation?

- (a) Increased model generalization
- (b) Potential introduction of unrealistic patterns
- (c) Improved model robustness
- (d) Decreased computational efficiency

74. What is the purpose of random cropping in image data augmentation?

- (a) To decrease the image resolution
- (b) To remove irrelevant features from the image
- (c) To create variations in the spatial location of objects
- (d) To increase the image contrast

75. Which type of data augmentation is commonly used for time series data?

(a) Image rotation (b) Time warping (c) Word substitution (d) Feature scaling

76. Explain the concept of "jittering" in the context of data augmentation.

(a) Jittering refers to the introduction of noise to input features
(b) Jittering involves the random selection of a subset of data points
(c) Jittering is a synonym for image rotation
(d) Jittering is irrelevant to data augmentation

77. In the context of image data augmentation, what is the purpose of horizontal flipping?

(a) To rotate images clockwise
(b) To create mirror images
(c) To adjust the image brightness
(d) To resize images

78. How does data augmentation differ from feature engineering?

(a) Data augmentation focuses on creating new samples, while feature engineering manipulates existing features.
(b) Feature engineering is limited to image data, while data augmentation is applicable to all data types.
(c) Data augmentation involves scaling features, while feature engineering involves randomization.
(d) Feature engineering and data augmentation are synonymous terms.

79. What is the role of dropout in the context of data augmentation?

(a) Dropout is not related to data augmentation
(b) Dropout enhances data augmentation by randomly removing features during training
(c) Dropout is a type of data augmentation technique
(d) Dropout prevents data augmentation from introducing unrealistic patterns

80. Which data augmentation technique is commonly used for audio data to introduce variations in pitch?

(a) Time warping
(b) Spectrogram augmentation
(c) Random cropping
(d) Jittering

81. What is the purpose of elastic deformation in image data augmentation?

(a) To adjust the image contrast
(b) To introduce non-linear distortions to the image
(c) To resize the image
(d) To rotate the image

82. In natural language processing, which technique involves randomly removing words from sentences during data augmentation?

(a) Tokenization
(b) Word substitution
(c) Sentence splitting
(d) Sentence dropout

83. Explain the concept of "adversarial training" in the context of data augmentation and how it addresses robustness.

(a) Adversarial training focuses on creating adversarial examples to test the model's robustness against unseen patterns introduced by data augmentation.
(b) Adversarial training is irrelevant to data augmentation.
(c) Adversarial training involves increasing the size of the training set.
(d) Adversarial training enhances data augmentation by introducing adversarial noise during the augmentation process.

84. How does data augmentation contribute to handling class imbalance in classification tasks?

(a) Data augmentation exacerbates class imbalance
(b) Data augmentation is not related to class imbalance
(c) Data augmentation generates additional samples for minority classes, addressing class imbalance
(d) Data augmentation reduces the need for addressing class imbalance

85. What challenges might arise when applying data augmentation to non-image data types, such as tabular data?

(a) Difficulty in implementing data augmentation for non-image data
(b) Limited applicability of data augmentation to non-image data
(c) The potential introduction of unrealistic patterns
(d) No challenges; data augmentation is equally effective for all data types

86. Explain the term "mixup" in the context of data augmentation and how it differs from traditional augmentation techniques.

(a) Mixup involves blending two or more samples, creating new synthetic samples with averaged labels.
(b) Mixup is a synonym for image rotation.
(c) Mixup refers to the addition of random noise to input features.
(d) Mixup is irrelevant to data augmentation.

87. How does data augmentation impact the interpretability of machine learning models?

(a) Data augmentation improves model interpretability by providing more diverse training examples.
(b) Data augmentation has no impact on model interpretability.
(c) Data augmentation reduces model interpretability due to the introduction of synthetic samples.
(d) Data augmentation improves model interpretability by eliminating the need for validation data.

88. What is the role of "cutout" in image data augmentation?

(a) To remove random portions from images
(b) To blur the edges of images
(c) To rotate images
(d) To resize images

89. In the context of data augmentation, explain how the technique of "shearing" is applied to image data.

(a) Shearing involves adjusting the brightness of images.
(b) Shearing is irrelevant to data augmentation.
(c) Shearing introduces non-linear distortions to the image by tilting it along one of its axes.
(d) Shearing is a synonym for image rotation.

Solutions to the Exercises

1. (**b**) L1
2.
3. (**c**) to prevent overfitting
4. (**a**) L1 regularization
5. (**c**) small weight values
6. (**b**) It randomly drops entire layers during training
7. (**c**) Elastic Net
8. (**b**) To prevent the model from memorizing the training data
9. (**c**) Regularization can help balance bias and variance
10. (**b**) The gradual decrease in weight values during training
11. (**d**) Reduced capacity to capture complex patterns
12. (**c**) Regulates the softness of the target distribution
13. (**c**) The probability of dropping out a unit in the hidden layers during training
14. (**b**) Learning rate annealing
15. (**c**) To reduce the impact of outliers in the input data
16. (**c**) Constraining the magnitude of the weights in the model
17. (**d**) Dropout helps prevent co-adaptation of hidden units
18. (**c**) To improve the predictive performance of a model by combining multiple models
19. (**b**) It trains multiple models independently on different subsets of the training data.
20. (**c**) To introduce randomness by considering a random subset of features for each tree
21. (**c**) Sequentially, with higher weights for misclassified instances
22. (**b**) Stacking
23. (**c**) Ensemble methods often generalize better and have improved robustness.
24. (**b**) A model that performs slightly better than random chance
25. (**b**) AdaBoost
26. (**c**) Bagging
27. (**a**) Bootstrap Aggregating
28. (**c**) Sequentially, with higher weights for misclassified instances
29. (**b**) Stacking
30. (**b**) AdaBoost
31. (**b**) Improved generalization and robustness
32. (**c**) Random Forest
33. (**a**) Long Short-Term Memory
34. (**a**) It adjusts the amount by which weights are updated during each iteration
35. (**c**) Random Forest introduces randomness by considering a random subset of features for each tree
36. (**d**) Stacking uses multiple base models to form a meta-model
37. (**c**) Ensemble methods help balance bias and variance
38. (**a**) Increased risk of overfitting
39. (**a**) Randomly and with replacement
40. (**b**) Better handling of outliers
41. (**c**) Boosting
42. (**d**) Using multiple base models to form a meta-model
43. (**c**) Random Forest
44. (**c**) It is an estimate of the test error obtained from the unused samples during training
45. (**b**) It limits the maximum depth of individual decision trees
46. (**b**) Early stopping prevents overfitting by stopping the training process when the model starts to memorize the training data.
47. (**b**) The computational complexity increases linearly
48. (**a**) Adversarial training involves training models to be robust against adversarial attacks.
49. (**b**) It uses multiple cross-validated models, reducing overfitting.
50. (**b**) Increased risk of overfitting
51. (**b**) Feature importance indicates the relevance of a feature in predicting the target variable.
52. (**c**) It specifies the number of base models in the ensemble.
53. (**a**) Stacking with meta-features involves using the output of base models as features for a meta-model.
54. (**b**) Removing random neurons during training
55. (**c**) To prevent co-adaptation of neurons
56. (**b**) Dropout is applied to all layers except the output layer
57. (**c**) By reducing the model's capacity
58. (**b**) The neuron is removed from the network temporarily
59. (**b**) Overfitting
60. (**a**) Slows down the training process
61. (**b**) 0.2 to 0.5
62. (**c**) By reducing the sensitivity of neurons to specific input features
63. (**a**) Training phase
64. (**a**) Co-adaptation refers to neurons relying too much on each other, and Dropout breaks these dependencies by randomly dropping neurons during training.
65. (**b**) Dropout is more effective in large and complex networks
66. (**c**) Dropout and ensemble learning achieve the same result in terms of model diversity
67. (**a**) High Dropout rates lead to overfitting, while low Dropout rates may result in underfitting.
68. (**c**) By introducing noise to the input data
69. (**c**) To improve model performance by increasing the diversity of the training data
70. (**c**) Image rotation
71. (**d**) By providing a more diverse set of training examples
72. (**c**) Word substitution
73. (**b**) Potential introduction of unrealistic patterns
74. (**c**) To create variations in the spatial location of objects
75. (**b**) Time warping
76. (**a**) Jittering refers to the introduction of noise to input features
77. (**b**) To create mirror images
78. (**a**) Data augmentation focuses on creating new samples, while feature engineering manipulates existing features.

79. (**b**) Dropout enhances data augmentation by randomly removing features during training
80. (**b**) Spectrogram augmentation
81. (**b**) To introduce non-linear distortions to the image
82. (**d**) Sentence dropout
83. (**a**) Adversarial training focuses on creating adversarial examples to test the model's robustness against unseen patterns introduced by data augmentation.
84. (**c**) Data augmentation generates additional samples for minority classes, addressing class imbalance
85. (**c**) The potential introduction of unrealistic patterns
86. (**a**) Mixup involves blending two or more samples, creating new synthetic samples with averaged labels.
87. (**c**) Data augmentation reduces model interpretability due to the introduction of synthetic samples.
88. (**a**) To remove random portions from images
89. (**c**) Shearing introduces non-linear distortions to the image by tilting it along one of its axes.