Data Science Thesaurus

~ Storm Schutte | 2022/03/13

1. Accuracy: Accuracy is a measure of how well a model is able to predict the correct labels for a given dataset.
2. Activation function: An activation function is a non-linear function applied to the output of a neuron to introduce non-linearity in the model.
3. Active Learning: Active learning is a technique where a model is able to request additional labeled data from an oracle (human or another model) in order to improve its performance.
4. Adam: Adam is an optimization algorithm that combines the ideas of gradient descent, momentum and adaptive learning rates.
5. Adversarial example: An adversarial example is a sample that has been slightly modified to fool a machine learning model into making a wrong prediction.
6. Algorithm: An algorithm is a set of instructions used to train a machine learning model and make predictions.
7. Anomaly Detection: Anomaly detection is a task of identifying data points that deviate from the normal or expected behavior.
8. Anomaly Detection: Anomaly detection is the task of identifying data points that are unusual or do not conform to a typical pattern.
9. Attention Mechanism: Attention Mechanism is a technique that allows a model to focus on certain parts of the input when making predictions, rather than using the entire input equally.
10. Attention: Attention is a mechanism used in some neural networks to focus on certain parts of the input when making predictions.
11. AUC-ROC: AUC-ROC (Area Under the Receiver Operating Characteristic Curve) is a metric that measures the performance of a binary classification model.
12. Auto-encoder: An autoencoder is a type of neural network that is trained to reconstruct its input, typically used for unsupervised learning tasks such as dimensionality reduction and anomaly detection.
13. AutoML: AutoML is a method of automating the process of selecting, designing, and tuning a machine learning model, thus reducing the need for human intervention.
14. Backpropagation: Backpropagation is an algorithm used to calculate the gradient of the loss function with respect to the parameters of a neural network.
15. Backpropagation: Backpropagation is an algorithm used to train neural networks, it is used to compute the gradient of the loss function with respect to the model's parameters.
16. Batch Normalization: Batch normalization is a technique used to improve the stability and accuracy of neural networks by normalizing the inputs to each layer across a batch of data.
17. Batch normalization: Batch normalization is a technique used to normalize the activations of a neural network, typically applied before or after the activation

function, in order to reduce internal covariate shift and improve the stability and performance of the model.

18. Batch Normalization: Batch normalization is a technique used to stabilize the training of deep neural networks by normalizing the activations of the neurons for each mini-batch of data.
19. Bayesian model: Bayesian models are a class of models that use Bayes' theorem to update beliefs about the state of the world based on new evidence.
20. Capsule Networks: A Capsule Network is a neural network architecture that uses "capsules" to represent features of an input. Capsules are a group of neurons that represent a specific feature and its properties.
21. Confusion Matrix: A confusion matrix is a table that is used to define the performance of a classification algorithm. It helps to visualize the performance of an algorithm.
22. Convolutional Neural Networks (CNN): Convolutional Neural Networks (CNN) is a type of neural network architecture specifically designed to process data with a grid-like topology, such as images, videos and speech signals.
23. Convolutional Neural Networks (CNN): Convolutional Neural Networks are a type of neural network that are particularly well suited for image recognition and other tasks that involve processing grid-like data.
24. Counterfactual Analysis: Counterfactual analysis is a method of explaining the decisions made by a model by identifying the minimal changes in the input data that would lead to a different prediction.
25. Cross-validation: Cross-validation is a technique used to estimate a model's generalization performance by training and evaluating the model multiple times using different subsets of the data.
26. Curse of dimensionality: The curse of dimensionality refers to the difficulty of working with high-dimensional data, as the number of samples required to train a model increases exponentially with the number of features.
27. Data augmentation: Data augmentation is a technique used to artificially increase the size of a dataset by applying random transformations to the existing data samples.
28. Data Augmentation: Data augmentation is the process of artificially increasing the size of a dataset by applying random transformations to the existing data.
29. Data Imputation: Data imputation is the process of filling in missing values in a dataset.
30. Deep Learning: Deep learning is a subfield of machine learning that uses deep neural networks with multiple layers to learn representations of data.
31. Dimensionality Reduction: Dimensionality reduction is the process of reducing the number of features in a dataset while retaining as much information as possible. Techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are commonly used for dimensionality reduction.
32. Dropout: Dropout is a regularization technique used to prevent overfitting in deep neural networks by randomly dropping out (i.e., setting to zero) some of the activations during training.
33. Dropout: Dropout is a regularization technique used to prevent overfitting in neural networks by randomly "dropping out" or ignoring a certain number of neurons during training.
34. Dropout: Dropout is a technique used to prevent overfitting by randomly dropping out a certain proportion of neurons from the network during training.

35. Early stopping: Early stopping is a technique used to prevent overfitting by interrupting the training process when the performance of the model on a validation set starts to decrease.
36. Embedding: Embedding is a technique used to represent categorical variables or words in a continuous and dense vector space.
37. Ensemble learning: Ensemble learning is a technique used to improve the performance of a model by combining the predictions of multiple models.
38. Ensemble Learning: Ensemble learning is a technique where multiple models are trained and combined to make a final prediction. Common ensemble techniques include bagging, boosting, and stacking.
39. Explainability: Explainability refers to the degree to which a model's predictions can be understood and explained by humans.
40. Explainable AI (XAI): Explainable AI (XAI) is a field of research that aims to make the decisions and predictions made by AI models more transparent and interpretable to humans.
41. Explainable AI (XAI): Explainable AI is a field of AI research focused on creating models that can be understood and interpreted by humans.
42. F1-Score: F1-Score is a metric that combines precision and recall into a single number.
43. Feature: A feature is an input variable that is used by the model to make predictions.
44. Generalization: Generalization refers to a model's ability to make accurate predictions on unseen data, beyond the data it was trained on.
45. Generative Adversarial Networks (GAN): Generative Adversarial Networks are a type of neural network that consist of two parts: a generator and a discriminator. The generator is trained to generate new data samples, while the discriminator is trained to distinguish between real and generated samples.
46. Generative Adversarial Networks (GANs): Generative Adversarial Networks (GANs) are a type of neural network architecture composed of two networks: a generator network and a discriminator network. The generator network generates new data samples that are similar to the training data, while the discriminator network tries to distinguish the generated samples from the real samples.
47. Generative model: A generative model is a model that can generate new data samples that are similar to the ones it was trained on.
48. Generative Pre-training: Generative Pre-training is a technique where a generative model is trained on a large dataset and then used as a pre-trained model for other tasks.
49. Gradient descent: Gradient descent is an optimization algorithm used to find the minimum of a function by iteratively moving in the direction of steepest descent.
50. Gradient Descent: Gradient descent is an optimization algorithm used to minimize a loss function by iteratively adjusting the parameters of a model in the direction of steepest decrease of the loss function.
51. Hyperparameter tuning: Hyperparameter tuning is the process of finding the optimal set of hyperparameters for a machine learning model. Hyperparameters are the parameters that are set before training a model, as opposed to the parameters learned during training.
52. Hyperparameter tuning: Hyperparameter tuning refers to the process of adjusting the hyperparameters of a model to optimize its performance on a given dataset.

Techniques such as grid search and random search are commonly used for hyperparameter tuning.

53. Hyperparameter tuning: Hyperparameter tuning, also known as model selection or hyperparameter optimization, is the process of selecting the best set of hyperparameters for a given model and dataset.
54. Hyperparameter: Hyperparameter is a parameter of a machine learning model that is set before training, as opposed to learned during training. Examples include learning rate, number of hidden layers, and regularization strength.
55. Hyperparameters: Hyperparameters are parameters that are set before training a model and are not learned during the training process.
56. Instance Segmentation: Instance segmentation is the task of detecting and segmenting individual objects of the same class in an image.
57. Interpretability: Interpretability refers to the degree to which the reasons behind a model's predictions can be understood by humans.
58. Label: A label is the output variable that the model is trying to predict.
59. Learning rate: The learning rate is a hyperparameter that controls the step size of the optimization algorithm.
60. Long Short-Term Memory (LSTM): Long Short-Term Memory (LSTM) is a type of RNN architecture that is designed to address the problem of vanishing gradients in traditional RNNs.
61. Long Short-term Memory (LSTM): Long Short-term Memory is a type of RNN that is able to capture long-term dependencies in sequential data by using a memory cell and gates to control the flow of information.
62. Loss function: Loss function is a function that measures the difference between the predicted outputs and the true outputs, used to guide the optimization process while training.
63. Machine Translation: Machine Translation is the task of automatically translating text from one language to another.
64. MCMC: Markov Chain Monte Carlo (MCMC) is a class of algorithms used to approximate complex probability distributions by sampling from them.
65. Meta-Learning: Meta-learning is a subfield of machine learning that involves learning how to learn, or learning how to improve the learning process itself.
66. Mini-batch Gradient Descent: Mini-batch Gradient Descent is another variant of gradient descent that uses a small fixed-size subset of the data to compute the gradient at each step.
67. Model Compression: Model compression is the process of reducing the size and computational requirements of a trained model without significant loss of accuracy.
68. Model: A model is a mathematical representation of a system or process that is used to make predictions or decisions.
69. Natural Language Processing (NLP): Natural Language Processing (NLP) is the field of machine learning that deals with the analysis, understanding, and generation of human language.
70. Natural Language Processing (NLP): Natural Language Processing is a field of artificial intelligence that focuses on the interaction between computers and human languages, including speech and text.
71. Neuron: A neuron is the basic building block of a neural network. It receives input, performs computations and produces output.

72. Object Detection: Object detection is the task of detecting and locating objects of interest in an image.
73. Overfitting: Overfitting occurs when a model is too complex and is able to fit the noise in the data instead of the underlying pattern, leading to poor generalization performance.
74. Overfitting: Overfitting occurs when a model is trained too well on the training data and performs poorly on new, unseen data. This is often caused by having too many features or a complex model, or too little data.
75. Partial dependence plot: Partial dependence plot is a technique used to visualize the relationship between a feature and the target variable, while holding other features constant.
76. Permutation Importance: Permutation importance is a technique used to measure the importance of a feature for a given model by shuffling the feature and measuring the change in the model's performance.
77. Precision: Precision is a metric that measures the proportion of true positive predictions among all positive predictions made by a model.
78. Q-Learning: Q-Learning is a specific algorithm used in RL that learns the optimal action-value function for an agent in an environment.
79. Q-Learning: Q-Learning is a type of reinforcement learning algorithm that is used to estimate the optimal action-value function (Q-function) for a given Markov Decision Process.
80. Recall: Recall is a metric that measures the proportion of true positive predictions among all actual positive instances.
81. Recommender Systems: Recommender systems are used to make personalized recommendations to users, such as suggesting products or movies they might like.
82. Recurrent Neural Networks (RNN): Recurrent Neural Networks (RNN) is a type of neural network architecture that can process sequential data such as time series, speech, and text.
83. Recurrent Neural Networks (RNN): Recurrent Neural Networks are a type of neural network that are particularly well suited for sequential data such as time series, speech, or text.
84. Regularization: Regularization is a technique used to prevent overfitting by adding a penalty term to the loss function that discourages the model from having too many parameters. Commonly used regularization techniques include L1 and L2 regularization.
85. Regularization: Regularization is a technique used to prevent overfitting by adding a penalty term to the model's objective function.
86. Reinforcement Learning (RL): Reinforcement Learning (RL) is a type of machine learning where an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or punishments.
87. Reinforcement Learning: Reinforcement learning is a type of learning where an agent learns to make a sequence of decisions by interacting with its environment and receiving feedback in the form of rewards or penalties.
88. Reinforcement Learning: Reinforcement learning is a type of machine learning where an agent learns to make decisions by interacting with an environment and receiving rewards or penalties for its actions.
89. Semantic Segmentation: Semantic segmentation is the task of classifying each pixel in an image into one of multiple predefined classes.

90. Semi-supervised Learning: Semi-supervised learning is a technique where a model is trained on a dataset that contains both labeled and unlabeled examples.
91. Sentiment Analysis: Sentiment analysis is a specific task within NLP that involves determining the sentiment (e.g. positive, negative, neutral) expressed in a piece of text.
92. Simulation-based Learning: Simulation-based learning is a subfield of machine learning that focuses on learning from simulations of the real world.
93. Speech Recognition: Speech recognition is the process of converting spoken words into text.
94. Stochastic Gradient Descent (SGD): Stochastic Gradient Descent is a variant of gradient descent that uses only a small random subset of the data to compute the gradient at each step.
95. Testing: Testing is the process of evaluating a trained model using a separate dataset to estimate its performance.
96. Training: Training is the process of using a dataset to learn the parameters of a model so that it can make accurate predictions.
97. Transfer Learning: Transfer learning is a technique where a pre-trained model on a large dataset is used as a starting point for training a model on a smaller dataset.
98. Transfer learning: Transfer learning is the process of using a pre-trained model on one task as a starting point for training a model on a different but related task.
99. Transfer Learning: Transfer learning is the process of using a pre-trained model on one task as a starting point for training a model on a new, but related task. This can help to reduce the amount of data and computation needed to train a new model.
100. Transferability: Transferability refers to the degree to which an adversarial example can be used to fool different models.
101. Underfitting: Underfitting occurs when a model is not able to learn the underlying pattern in the data and performs poorly on both training and new, unseen data. This is often caused by having too few features or a too simple model.
102. Underfitting: Underfitting occurs when a model is too simple and is unable to capture the complexity of the data, leading to poor performance on the training and test data.
103. Unsupervised Learning: Unsupervised learning is a technique where a model is trained on a dataset that contains only unlabeled examples.
104. Validation: Validation is the process of evaluating a model's performance using a separate dataset, typically used to tune the model's hyperparameters. [Original source: https://studycrumb.com/alphabetizer]