Practice microlearnings

Fill in the blanks to test your knowledge

What is Data Science?

- [ ] _____ is the exploration and quantitative analysis of structured and unstructured data to extract knowledge and drive actionable results.
- [ ] It is _____, involving statistics, computer science, business knowledge, and more.
- [ ] It involves using data to _____ and requires multiple skill sets across domains.

Types of Analytics

- [ ] _____ explains what happened.
- [ ] _____ uses past data to forecast future outcomes.
- [ ] _____ recommends actions based on predictive models.
- [ ] Other types include _____ and _____.

Data Science Process Models

- [ ] _____ includes stages like selection, preprocessing, transformation, mining, and evaluation.
- [ ] _____ is a cross-industry process model with phases like business understanding, data understanding, and modeling.
- [ ] _____ and _____ include collection, analysis, visualization, and decision-making steps.

Big Data Concepts

- [ ] Big Data is characterized by the ___: Volume, Velocity, and Variety.
- [ ] Modern tools like _____ help process large-scale datasets.

- [ ] _____ emerged prominently with NASA in 1997 and expanded in the 2010s.

Machine Learning Introduction

- [ ] _____ is the process of teaching computers to learn from data without being explicitly programmed.
- [ ] ML includes _____, _____, and _____ learning.
- [ ] _____ and _____ are core ML tasks.

Supervised Learning

- [ ] _____ uses labeled data to train models.
- [ ] _____ predicts discrete labels (e.g., spam or not spam).
- [ ] _____ predicts continuous values (e.g., stock prices).

Unsupervised Learning

- [ ] _____ deals with unlabeled data.
- [ ] _____ (like K-Means) groups similar data points together.
- [ ] Evaluation is more difficult than in supervised learning due to lack of ground truth.

Statistical Learning Theory

- [ ] _____ focuses on model generalization and error minimization.
- [ ] Key ideas include _____ and minimizing _____ as a proxy for _____.
- [ ] _____, _____, _____, and _____ are core concepts.

Model Evaluation and Overfitting

- [ ] _____ occurs when a model learns training data too well and performs poorly on new data.

- [ ] _____ occurs when a model is too simple to capture patterns.
- [ ] Good models balance training error and complexity to generalize well.

Model Building and Application

- [ ] Data scientists use data to _____ that guide decisions.
- [ ] Common tasks: spam detection, stock forecasting, medical diagnosis, and sentiment analysis.
- [ ] _____ suggest items based on user history and preferences.

Answered microlearnings

Review sheet with highlighted answers

What is Data Science?

- [ ] **Data Science** is the exploration and quantitative analysis of structured and unstructured data to extract knowledge and drive actionable results.
- [ ] It is **interdisciplinary**, involving statistics, computer science, business knowledge, and more.
- [ ] It involves using data to **make decisions** and requires multiple skill sets across domains.

Types of Analytics

- [ ] **Descriptive analytics** explains what happened.
- [ ] **Predictive analytics** uses past data to forecast future outcomes.
- [ ] **Prescriptive analytics** recommends actions based on predictive models.
- [ ] Other types include **retrospective** and **real-time analytics**.

Data Science Process Models

- [ ] **KDD (Knowledge Discovery in Databases)** includes stages like selection, preprocessing, transformation, mining, and evaluation.
- [ ] **CRISP-DM** is a cross-industry process model with phases like business understanding, data understanding, and modeling.
- [ ] **CCC Pipeline** and **modern pipelines** include collection, analysis, visualization, and decision-making steps.

Big Data Concepts

- [ ] Big Data is characterized by the **3Vs**: Volume, Velocity, and Variety.
- [ ] Modern tools like **Hadoop, Spark, and NoSQL** help process large-scale datasets.

- [ ] **Big Data** emerged prominently with NASA in 1997 and expanded in the 2010s.

Machine Learning Introduction

- [ ] **Machine Learning (ML)** is the process of teaching computers to learn from data without being explicitly programmed.
- [ ] ML includes **supervised**, **unsupervised**, and **reinforcement** learning.
- [ ] **Classification, regression, clustering,** and **recommender systems** are core ML tasks.

Supervised Learning

- [ ] **Supervised learning** uses labeled data to train models.
- [ ] **Classification** predicts discrete labels (e.g., spam or not spam).
- [ ] **Regression** predicts continuous values (e.g., stock prices).

Unsupervised Learning

- [ ] **Unsupervised learning** deals with unlabeled data.
- [ ] **Clustering** (like K-Means) groups similar data points together.
- [ ] Evaluation is more difficult than in supervised learning due to lack of ground truth.

Statistical Learning Theory

- [ ] **Statistical learning theory** focuses on model generalization and error minimization.
- [ ] Key ideas include **Occam's Razor** and minimizing **empirical risk** as a proxy for **expected risk**.
- [ ] **Input (x)**, **output (y)**, **hypothesis (h)**, and **loss function (ℓ)** are core concepts.

Model Evaluation and Overfitting

- [ ] **Overfitting** occurs when a model learns training data too well and performs poorly on new data.
- [ ] **Underfitting** occurs when a model is too simple to capture patterns.
- [ ] Good models balance training error and complexity to generalize well.

Model Building and Application

- [ ] Data scientists use data to **build predictive models** that guide decisions.
- [ ] Common tasks: spam detection, stock forecasting, medical diagnosis, and sentiment analysis.
- [ ] **Recommender systems** suggest items based on user history and preferences.