Batch 4 - Module B

Set 1 – Answer Justifications

Question 1: C) Linear Regression **Justification:** Linear regression is the appropriate initial model to try when a linear relationship is suspected between the independent (square footage) and dependent (house prices) variables.

Question 2: A) LinearRegression() **Justification:** The code imports the LinearRegression class from sklearn.linear_model. To instantiate a linear regression model, you need to call the class constructor using LinearRegression().

Question 3: 60 **Justification:** Substituting the distance (20 km) into the given linear model equation: time = 2.5 × 20 + 10 = 50 + 10 = 60 minutes.

Question 4: B) Overfitting **Justification:** Using a high-degree polynomial (degree-10) that fits the training data very well but performs poorly on unseen data (increased test error) is a classic sign of overfitting.

Question 5: B) Try Polynomial Regression **Justification:** A clear curved upward trend in the scatter plot suggests a non-linear relationship between revenue and time. Polynomial regression is designed to model such non-linear relationships.

Question 6: predict **Justification:** After fitting a linear regression model in scikit-learn, the predict() method is used to generate predictions on new or training data (X_test).

Question 7: B) Logistic Regression **Justification:** Logistic regression is specifically designed for binary classification problems where the goal is to predict a categorical outcome with two classes (Yes/No, 1/0).

Question 8: 0.5 **Justification:** The sigmoid function is defined as $1+e^{-z}1$. When $z=0$, the sigmoid value is $1+e^{0}1=1+11=21=0.5$. Also can be answered by intuition.

Question 9: A, C **Justification:** A) The sigmoid function's output always lies between 0 and 1, inclusive. C) The output of the sigmoid function in logistic regression is interpreted as the probability of the positive class.

Question 10: B) 1 **Justification:** In logistic regression with a threshold of 0.5, if the predicted probability (0.7) is greater than the threshold, the final prediction is classified as 1.

Question 11: B) To determine step size when updating weights **Justification:** The learning rate in gradient descent is a hyperparameter that controls the size of the steps taken to update the model's weights during the optimization process.

Question 12: learning_rate **Justification:** The gradient update rule in gradient descent involves subtracting a fraction of the gradient from the current weight. This fraction is determined by the learning rate.

Question 13: A) L1 (Lasso) **Justification:** L1 regularization, also known as Lasso, adds a penalty proportional to the absolute value of the coefficients to the loss function.

Question 14: A, B, D **Justification:** A) L2 regularization (Ridge) adds a penalty proportional to the square of the coefficients, which shrinks their values towards zero but generally doesn't make them exactly zero. B) By penalizing large coefficients, L2 regularization helps to prevent overfitting. D) Shrinking the coefficients makes the model simpler and often leads to better generalization on unseen data.

Question 15: True **Justification:** In Ridge regression, increasing the alpha value increases the strength of the regularization penalty. If alpha is too high, it can overly constrain the model, preventing it from fitting the data adequately and leading to underfitting.

Question 16: C) PCA **Justification:** Principal Component Analysis (PCA) is a dimensionality reduction technique that can reduce a high-dimensional dataset (200 features) into a lower-dimensional representation (e.g., 2D) for visualization while retaining most of the variance in the data.

Question 17: B) Most of the data's structure is captured by 10 features **Justification:** If 10 principal components explain 95% of the variance in the data, it implies that these 10 components capture most of the underlying structure and information present in the original 1000 features.

Question 18: C) Importance of web pages based on links **Justification:** The PageRank algorithm is designed to measure the importance or authority of web pages based on the number and quality of links pointing to them.

Question 19: A, B, D **Justification:** A) The number of inbound links (links from other pages) directly influences a page's PageRank. B) Links from high-quality or high-PageRank websites contribute more to the PageRank of the linked page. D) The overall structure of links within the web network is fundamental to how PageRank is calculated and distributed.

Question 20: A) Founder A (most referenced and referencing) **Justification:** In the given graph, Founder A has two outgoing links (to B and C) and is also linked to by C. This central position, being both a source and a recipient of links, especially within a small interconnected network, tends to give A a higher PageRank.

Question 21: B) Reduced representation of data **Justification:** The pca.fit_transform(X_scaled) method applies PCA to the scaled data and returns a new array X_pca where the original data has been transformed into the lower-dimensional space defined by the principal components. The shape of X_pca will be (number of samples, n_components), representing the reduced data.

Question 22: convergence **Justification:** When the gradient norm becomes very small and the model parameters stop changing significantly over several iterations, it indicates that the optimization process has likely reached a minimum or a flat region in the loss landscape, which is referred to as convergence.

Question 23: C **Justification:** C) The coefficient 5 represents the slope of the relationship between age and price

Question 24: C) Polynomial Regression (degree 2) **Justification:** A U-shaped trend indicates a non-linear relationship where the dependent variable (coral reef survival) initially decreases and then increases with the independent variable (ocean temperature). Polynomial regression, particularly a degree-2 (quadratic) polynomial, can model such curved relationships.

Question 25: C) Severe overfitting **Justification:** Achieving a near-perfect R^2 on the training data with a very high-degree polynomial (15th degree) fitted to a small number of data points (12) strongly suggests that the model has memorized the training data, including its noise. A negative test R^2 indicates that the model performs worse than simply predicting the mean of the test data, which is a clear sign of severe overfitting.

Question 26: A, B, C **Justification:** A) Stochastic Gradient Descent (SGD) updates the model weights after processing each individual training example. B) Mini-batch Gradient Descent uses a small random subset (batch) of the training data to compute the gradient and update the weights in each iteration, offering a balance between the efficiency of SGD and the stability of Batch GD. C) Batch Gradient Descent calculates the gradient of the loss function over the entire training dataset before updating the model weights in each epoch.

D) is incorrect. Mini-batch Gradient Descent is slower than Batch Gradient Descent but more accurate. Incorrect for two reasons: Mini-batch GD is faster than Batch GD in most practical cases because: Batch GD requires computing gradients over all millions of records per update and Mini-batch only uses a small subset (e.g., 32, 64 samples). Accuracy depends on many factors — learning rate, model architecture, convergence — and Mini-batch GD often performs better due to efficient GPU use and smoother convergence.

Question 27: B) It gives higher rank to bloggers linked by other influential bloggers. **Justification:** PageRank assigns importance based on the principle that a page is important if it is linked to by other important pages. In the context of travel bloggers, PageRank would give higher scores to bloggers who are linked to by other influential bloggers, effectively identifying those with greater authority or recognition within the network.

Question 28: B) Only manually tagged images are labeled data. **Justification:** Labeled data refers to data that has been manually tagged with the correct output or classification (in this case, healthy/unhealthy crop status). The drone images manually labeled with crop health status are therefore the labeled data.

Question 29: B) 50% chance of disease **Justification:** When z=0, the sigmoid function evaluates to $1+e^{-0}1=1+11=0.5$. In the context of logistic regression for binary classification, the sigmoid output represents the predicted probability of the positive class (having the disease). Therefore, a sigmoid output of 0.5 corresponds to a 50% chance of having the disease.

Question 30: A, C **Justification:** A) Polynomial regression can model non-linear relationships between variables, including U-shaped or other curved patterns observed in the data. C) By using polynomial features, the model can capture more complex trends in the data compared to a simple linear model.