Solution week-2

1. Regularized Logistic Regression – Gradient Update Term

   The cost function in L2-regularized logistic regression adds a penalty on the weights.

   Which term in the gradient update reflects the effect of L2 regularization?

   A) Subtract learning_rate * weight

   B) Subtract learning_rate * weight / m

   C) Subtract learning_rate * weight * (lambda / m)

   D) Add learning_rate * weight * (lambda / m)

   Answer: C

2. Non-linear Data Handling

   Which algorithm can handle non-linearly separable data by implicitly mapping to a higher-dimensional space?

   A) Perceptron

   B) k-Means

   C) Kernel SVM

   D) Naive Bayes

   Answer: C

3. Bias-Variance Trade-off (Select two)

   Which of the following are true about the bias-variance trade-off?

   A) High bias models underfit

   B) High variance models generalize well

   C) Increasing model complexity lowers bias but increases variance

   D) Regularization increases variance and decreases bias

   Answer: A, C

4. Simple Linear Regression Intercept

   What does the intercept term in a simple linear regression represent?

   A) The slope of the line

   B) The predicted value of Y when X = 0

   C) The average of all Y values

   D) The average residual

   Answer: B

5. Logistic Function

   Which function does logistic regression use to map the linear combination of features to a probability?

   A) Identity

B) Softmax

C) Sigmoid

D) Tanh

Answer: C

6. Error Metric Sensitivity

   Which error metric is most sensitive to large individual errors because it squares the residuals?

   A) Mean Absolute Error (MAE)

   B) Mean Squared Error (MSE)

   C) Accuracy

   D) Area Under the ROC Curve (AUC)

   Answer: B

7. Role of Learning Rate

   In gradient descent, what role does the learning rate play?

   A) It sets the initial values of the parameters

   B) It determines the size of each update step

   C) It selects the loss function

   D) It decides when to stop training

   Answer: B

8. Naive Bayes Independence Assumption

   Which algorithm assumes that all features are conditionally independent given the class label?

   A) k-Nearest Neighbors

   B) Decision Tree

   C) Naive Bayes

   D) Support Vector Machine

   Answer: C

9. L1 (Lasso) Regularization (Select two)

   Which of the following are characteristics of L1 regularization?

   A) Promotes sparse solutions (many coefficients become exactly zero)

   B) Penalizes the square of coefficients

   C) Shrinks coefficients continuously toward zero

   D) Uses the absolute value of coefficients in the penalty

   Answer: A, D

10. Imbalanced-Data Metrics (Select three)

    When evaluating a classifier on highly imbalanced data, which metrics should you consider?

    A) Accuracy

    B) Precision

C) Recall

D) F1-score

Answer: B, C, D

11. Naive Bayes Distributions (Select three)

    Which probability distributions are commonly used in Naive Bayes variants?

    A) Gaussian (for continuous features)

    B) Bernoulli (for binary features)

    C) Multinomial (for word counts)

    D) Exponential (for time-to-event data)

    Answer: A, B, C

12. Overfitting Indicators (Select two)

    Which statements about overfitting are true?

    A) It occurs when model complexity is too low

    B) It leads to low training error but high test error

    C) Regularization techniques can help reduce it

    D) It always improves model generalization

    Answer: B, C

13. Gradient Descent Use Cases (Select two)

    For which tasks can gradient descent be used?

    A) Convex optimization problems

    B) Discrete combinatorial search

    C) Parameter estimation in linear regression

    D) Exact global optimization of arbitrary functions

    Answer: A, C

14. Threshold Classifier Maximum Accuracy

    Table of Feature X and Label values for binary classification:

Sample	Feature X	Label
1	2.0	+1
2	-1.0	-1
3	0.5	+1
4	-0.5	-1
5	1.0	+1

    A simple threshold classifier predicts +1 if X > t, else –1.

    Which threshold t maximizes accuracy?

    A) 0.75

    B) 0.25

C) –0.25

D) 0.0

Answer: D

15. Gradient Descent Update for Linear Model (MSE)

    Given points (x, y) = (1, 2), (2, 3) and model y = w x (no intercept), starting w = 0.

    After one GD step with α = 0.1 under MSE cost, what is the updated w?

    A) 0.2

    B) 0.4

    C) 0.6

    D) 0.8

    Answer: B

16. Logistic Regression Predicted Probability

    For w = 2, b = –1, what is P(y = 1 | x = 1)? (σ(z)=1/(1+e^(–z)))

    A) 0.27

    B) 0.50

    C) 0.73

    D) 0.88

    Answer: C

17. Naive Bayes Posterior Probability

    P(Spam)=0.4, P(NotSpam)=0.6; P(Free|Spam)=0.8; P(Free|NotSpam)=0.1.

    If “Free”=true, what is P(Spam|Free)?

    A) 0.20

    B) 0.50

    C) 0.67

    D) 0.84

    Answer: D

18. Linear Regression with scikit-learn

    What does this code print?

```
import numpy as np
from sklearn.linear_model import LinearRegression
X = np.array([[1],[2],[3],[4]])
y = np.array([2,4,6,8])
lr = LinearRegression()
lr.fit(X,y)
print(lr.coef_, lr.intercept_)
```

    A) [2.] 0.0

B) [0.5] 10.0

C) [2.] 1.0

D) [4.] –2.0

Answer: A

19. Ordinary Least Squares Singularity

If one feature is constant across all samples, what happens to the normal equation?

A) Its coefficient becomes zero.

B) X^TX is singular (non-invertible), no unique solution.

C) GD ignores that feature.

D) The model fits perfectly regardless of targets.

Answer: B

20. Confusion Matrix Accuracy

	Actual Positive	Actual Negative
Pred Positive	40	10
Pred Negative	5	45

What is the accuracy?

A) 0.85

B) 0.90

C) 0.88

D) 0.95

Answer: A

21. Mean Absolute Error (MAE)

Sample	Actual	Predicted
1	100	95
2	110	120
3	90	85

Compute MAE.

A) 6.00

B) 6.67

C) 7.00

D) 8.00

Answer: B