Machine Learning Midterm Examination

Total Marks: 100

Time Allowed: 3 hours

Section A: Short Answer Questions (5 × 4 = 20 marks)

(Answer all questions concisely)

1. **Regularization**:
   Explain why L1 regularization (Lasso) induces sparsity in feature weights, while L2 regularization (Ridge) does not. Use a geometric intuition.
2. **KNN vs. K-Means**:
   Contrast K-Nearest Neighbors (supervised) with K-Means clustering (unsupervised) in terms of objectives, input data, and output.
3. **PCA**:
   Given a dataset with 100 features, you perform PCA and retain 95% variance. The reduced dataset has 12 principal components. Interpret this result.
4. **ROC Curve**:
   Define TPR (True Positive Rate) and FPR (False Positive Rate). Why is the ROC curve invariant to class imbalance?
5. **Decision Trees**:
   Describe how Gini impurity and entropy differ as splitting criteria. When would you prefer one over the other?

Section B: Medium Answer Questions (5 × 8 = 40 marks)

(Derivations and numerical problems)

1. **Gradient Descent**:
   Consider the loss function $J(\theta) = \theta^2 + 5\theta + 6$.
   - Derive the gradient update rule for θ with learning rate $\eta = 0.1$.
   - Perform 3 iterations starting from $\theta_0 = 0$.
2. **Logistic Regression**:
   For binary classification with sigmoid activation $\sigma(z) = \frac{1}{1+e^{-z}}$:
   - Derive the log-likelihood loss $J(\theta)$.
   - Compute the gradient $\frac{\partial J}{\partial \theta_j}$.
3. **SVM & Hinge Loss**:
   Given a linearly separable dataset and the hinge loss: $L(y, \hat{y}) = \max(0, 1 - y \cdot \hat{y})$,
   - Explain how maximizing the margin relates to minimizing hinge loss.
   - Solve: If $y = 1$ and the classifier outputs $\hat{y} = 0.8$, compute the hinge loss.

4. **K-Means Numerical**:
   Cluster points: $A(1,2), B(1,0), C(4,1), D(5,2)$ into $K=2$ clusters. Initialize centroids at $\mu_1 = (1,0)$ and $\mu_2 = (5,2)$.
   - Assign points to clusters.
   - Recompute centroids after one iteration.
5. **Naive Bayes**:
   Predict if an email is spam ($Y=1$) based on words "win" (X_1) and "prize" (X_2). Given:
   - $P(Y=1) = 0.3$, $P(X_1=1|Y=1) = 0.6$, $P(X_2=1|Y=1) = 0.4$
   - $P(X_1=1|Y=0) = 0.1$, $P(X_2=1|Y=0) = 0.2$
     Compute $P(Y=1|X_1=1, X_2=1)$.

Section C: Long Answer Questions (2 × 20 = 40 marks)

(Answer both questions with detailed derivations/analysis)

1. **Ensemble Methods**:
   - Explain how Random Forests (bagging) reduce variance, while AdaBoost (boosting) reduces bias. (6 marks)
   - Describe the working of XGBoost, including its regularization strategy and split-finding optimization. (8 marks)
   - Numerically: Train a Gradient Boosting model for 2 iterations on this data (MSE loss, learning rate $\eta = 0.5$):

X	y
1	2
2	3
3	5

     (Base model: $F_0(X) = \bar{y}$). (6 marks)
2. **Clustering & Model Evaluation**:
   - Compare DBSCAN, Hierarchical Clustering, and BIRCH in terms of:
     - Handling outliers.
     - Scalability to large datasets.
     - Shape flexibility of clusters.
       (8 marks)
   - Given a classifier's confusion matrix:

	Predicted +	Predicted −
Actual +	40	10
Actual −	5	45

     Calculate: Accuracy, Precision, Recall, F1-score. Sketch the ROC curve (show AUC intuition). (8 marks)
   - Why is cross-validation crucial for hyperparameter tuning? Contrast 5-fold CV with LOOCV. (4 marks)

Answer Key Highlights

- **Section A**: Tests conceptual clarity (e.g., ROC invariance due to TPR/FPR ratios).

- **Section B**: Requires step-by-step calculations (e.g., K-Means centroid updates).
- **Section C**: Focuses on synthesis (e.g., XGBoost's weighted quantile sketch) and analysis (e.g., density-based vs. centroid-based clustering).

The paper balances:

- **Theory**: Regularization geometry, ensemble principles.
- **Derivations**: Logistic regression gradients, boosting updates.
- **Numericals**: K-Means, Naive Bayes, GBM iterations.