1 Comparison of Machine Learning Algorithms [Jayant, 20 points]

In this problem, you will review the important aspects of the algorithms we have learned about in class. For every algorithm listed in the two tables on the next pages, fill out the entries under each column according to the following guidelines. Turn in your completed table with your problem set. [$\approx \frac{1}{2}$ **point per entry**]

Guidelines:

1. **Generative or Discriminative** – Choose either "generative" or "discriminative"; you may write "G" and "D" respectively to save some writing.
2. **Loss Function** – Write either the name or the form of the loss function optimized by the algorithm (e.g., "exponential loss").
3. **Decision Boundary / Regression Function Shape** – Describe the shape of the decision surface or regression function, e.g., "linear". If necessary, enumerate conditions under which the decision boundary has different forms.
4. **Parameter Estimation Algorithm / Prediction Algorithm** – Name or concisely describe an algorithm for estimating the parameters or predicting the value of a new instance. Your answer should fit in the provided box.
5. **Model Complexity Reduction** – Name a technique for limiting model complexity and preventing overfitting.

Solution: Completed tables are on the following pages.

Learning Method	Generative or Discriminative?	Loss Function	Decision Boundary	Parameter Estimation Algorithm	Model Complexity Reduction
Gaussian Naïve Bayes	Generative	$-\log P(X, Y)$	Equal variance: linear boundary. Unequal variance: quadratic boundary	Estimate $\hat{\mu}$, $\hat{\sigma}^2$, and $P(Y)$ using maximum likelihood	Place prior on parameters and use MAP estimator
Logistic Regression	Discriminative	$-\log P(Y\|X)$	Linear	No closed form estimate. Optimize objective function using gradient descent.	L_2 regularization
Decision Trees	Discriminative	Either $-\log P(Y\|X)$ or zero-one loss	Axis-aligned partition of feature space	Many algorithms: ID3, CART, C4.5	Prune tree or limit tree depth
K-Nearest Neighbors	Discriminative	zero-one loss	Arbitrarily complicated	Must store all training data to classify new points. Choose K using cross validation.	Increase K
Support Vector Machines (with slack variables, no kernel)	Discriminative	hinge loss: $\|1-y(w^T x)\|_+$	linear (depends on kernel)	Solve quadratic program to find boundary that maximizes margin	Reduce C
Boosting (with decision stumps)	Discriminative	exponential loss: $\exp\{-yf(x)\}$	Axis-aligned partition of feature space	AdaBoost	Reduce the number of iterations

Table 1: Comparison of Classification Algorithms

Learning Method	Loss Function	Regression Function Shape	Parameter Estimation Algorithm	Prediction Algorithm
Linear Regression (assuming Gaussian noise model)	square loss: $(\hat{Y} - Y)^2$	Linear	Solve $\beta = (X^T X)^{-1} X^T Y$	$\hat{Y} = X\beta$
Nadaraya-Watson Kernel Regression	square loss: $(\hat{Y} - Y)^2$	Arbitrary	Store all training data. Choose kernel bandwidth h using cross validation.	$f(x) = \frac{\sum_i y_i K(x_i, x)}{\sum_j K(x_j, x)}$
Regression Trees	square loss: $(\hat{Y} - Y)^2$	Axis-aligned partition of feature space	Many: ID3, CART, C4.5	Move down tree based on x, predict value at the leaf.

Table 2: Comparison of Regression Algorithms

2 Comparison of Machine Learning Algorithms [Jayant, 15 pts]

In this problem, you will review the important aspects of the algorithms we have learned about in class since the midterm. For every algorithm listed in the two tables on the next pages, fill out the entries under each column according to the following guidelines. Do not fill out the greyed-out cells. Turn in your completed table with your problem set.

Guidelines:

1. **Generative or Discriminative** – Choose either "generative" or "discriminative"; you may write "G" and "D" respectively to save some writing.
2. **Loss Function** – Write either the name or the form of the loss function optimized by the algorithm (e.g., "exponential loss"). For the clustering algorithms, you may alternatively write a short description of the loss function.
3. **Decision Boundary** – Describe the shape of the decision surface, e.g., "linear". If necessary, enumerate conditions under which the decision boundary has different forms.
4. **Parameter Estimation Algorithm / Prediction Algorithm** – Name or concisely describe an algorithm for estimating the parameters or predicting the value of a new instance. Your answer should fit in the provided box.
5. **Model Complexity Reduction** – Name a technique for limiting model complexity and preventing overfitting.
6. **Number of Clusters** – Choose either "predetermined" or "data-dependent"; you may write "P" and "D" to save time.
7. **Cluster Shape** – Choose either "isotropic" (i.e., spherical) or "anisotropic"; you may write "I" and "A" to save time.

Solution: Completed tables are on the following pages.

Learning Method	Generative or Discriminative?	Loss Function	Parameter Estimation Algorithm	Prediction Algorithm	Model Complexity Reduction
Bayes Nets	Generative	$-\log P(X,Y)$	MLE	Variable Elimination	MAP
Hidden Markov Models	Generative	$-\log P(X,Y)$	MLE	Viterbi or Forward-Backward, depending on prediction task	MAP
Neural Networks	Discriminative	Sum-squared error	Back-Propagation	Forward Propagation	Reduce number of hidden layers, regularization, early stopping

Table 1: Comparison of Classification Algorithms

Learning Method	Loss Function	Number of clusters: Predetermined or Data-dependent	Cluster shape: isotropic or anisotropic?	Parameter Estimation Algorithm
K-means	Within-class squared distance from mean	Predetermined	Isotropic	K-means
Gaussian Mixture Models (identity covariance)	$-\log P(X)$, (equivalent to within-class squared distance from mean)	Predetermined	Isotropic	Expectation Maximization (EM)
Single-Link Hierarchical Clustering	Maximum distance between a point and its nearest neighbor within a cluster	Data-dependent	Anisotropic	Greedy agglomerative clustering
Spectral Clustering	Balanced cut	Predetermined	Anisotropic	Run Laplacian Eigenmaps followed by K-means or thresholding eigenvector signs

Table 2: Comparison of Clustering Algorithms