Problem 1: General Understanding (12 of 58 points)

How are these methods called?

Example: Minimize the sum of the edge weights of a Hamiltonian cycle of a graph: traveling salesman algorithm.

a) Maximize the variance of the data along one dimension.

b) Minimize the sum of small quadratic and large scaled absolute regression errors.

c) Minimize the weighted sum of the squared norm of the normal vector of the discriminance hyperplane in the kernelized high–dimensional space plus the sum of the penalty terms for all data points that are not beyond the classification margin.

d) Minimize the sum of distances between each data point and the average of the corresponding set of points.

e) Minimize the quadratic reconstruction error after linear projection to a plane.

f) Minimize the sum of the m smallest squared differences between the regression values and the output data.

Problem 2: Principal Component Analysis (16 of 58 points)

Consider a data set X for which one–dimensional PCA will yield

$$X' = \{(3, 4), (3, 4), (-3, -4), (-3, -4)\}$$

with mean square error $e = 1$.

a) Compute the eigenvectors of the covariance matrix of X (not X').

b) Compute the eigenvalues of the covariance matrix of X (not X').

c) Find all data sets X that will lead to these results.

Problem 3: Regression (16 of 58 points)

Consider an n–layer MLP with one neuron per layer, no offset, transfer function

$$y = \frac{x}{\sqrt{1 + x^2}}$$

All weights are initialized as 1. Training is done by minimizing the quadratic output error using gradient descent backpropagation. We use a simple training data set with only one sample: input 1 and output 0.

a) For the untrained network and the given training input, compute the output of layer k, $k = 1, \ldots, n$.

b) For the first training step with the given training data, compute the gradient to update the weight of the edge between layer $k-1$ and layer k, $k = 1, \ldots, n$.

c) Which problem will occur when the number of layers becomes very large?

d) How can this problem be avoided?

Problem 4: Classification (14 of 58 points)

Consider the following data from the UK Office for National Statistics about the mortality reates by vaccination status for deaths involving COVID–19 in England, from September 18–24, 2021.

Vaccination status	Age group	Number of deaths	Population
unvaccinated	10–59	28	7167322
unvaccinated	60+	67	470189
vaccinated	10–59	34	19811645
vaccinated	60+	436	11700409

a) What is the death probability that a Naive Bayes classifier yields for a vaccinated person aged 10–59?

b) What is the death probability that a Naive Bayes classifier yields for an unvaccinated person aged 10–59?

c) Compare and explain the results from (a) and (b)!

d) From the given data (ignoring the Naive Bayes classifier), compute the conditional probability of death for a vaccinated person.

e) From the given data (ignoring the Naive Bayes classifier), compute the conditional probability of death for an unvaccinated person.

f) Compare and explain the results from (d) and (e)!