DATA.STAT.770 Dimensionality Reduction and Visualization, Spring 2025, Exercise set 2

Part A: Forward Selection and Variable Ranking

In this week's exercises we will use feature selection for a prediction (regression) task, using nearest-neighbor predictor on a face image data set. The basic information is first given below, then the exercises are given as Problem A1 and A2.

Goodness measure. Given a finite training data set of N data points where the output values of each data point are known, the goodness of the nearest neighbor predictor can be estimated by *leave-one-out validation error*: for each data point, predict its output values using the $N - 1$ other points as the set of potential neighbors, and count the sum of squared errors between the predictions and the true values (sum over all output variables and all data points where predictions were made).

Data set. The Sculpt Faces data set "noisy_sculpt_faces.txt" is provided in the same archive as this description. It is a data set of 100 face images of a statue taken from different directions with different lighting directions; each image is a 16×16 grayscale image, where the $16 \cdot 16 = 256$ pixel values are the features. The file is an ASCII file where each row is an image: in each row, the first 256 values are the pixel values of the 16×16 image listed column by column, and the rest of the values describe how the image was taken: the 257th and 258th values are two pose angles (left-right and top-down) describing which direction the face is looking at and the 259th value is the angle of lighting. The images in this data set have been downscaled from original 64×64 sized images to reduce the amount of computation time needed to solve the exercise, and noise has been added to the pixels to simulate a poor-quality grainy camera. The original data set, and a picture of a nonlinear embedding of the original images, can be seen at `http://web.archive.org/web/20160913051505/http://isomap.stanford.edu/datasets.html`. However, the images in your data set are heavily corrupted and it is hard for a human observer to see the original faces anymore—but it may be possible for a computer to notice useful features (pixels that are not too noisy and contain useful information for classification).

Task. We want to predict the orientation of the face and the direction of lighting based on the content of the noisy low-resolution image (the pixel values): the target values are the two pose angles and lighting angle. We will use nearest neighbor prediction, and use squared error of leave-one-out prediction as a performance measure. We want to choose the best features for the task: that is, we want to find out which of the 16×16 pixels are most helpful for predicting the target values. The amount of noise in the different pixels varies, and also each pixel shows a different part of the faces, so it is reasonable to expect that some pixels will be more useful for the prediction than others.

Problem A1: Forward Selection

In this exercise, we will use *forward selection* to select the best features for nearest neighbor prediction of the angles in the face images, by minimizing the leave-one-out validation error.

a) Implement the nearest neighbor predictor, and compute the leave-one-out error of nearest neighbor prediction for the Sculpt Faces data set when using all the data features. This number is the *baseline* performance that we must improve, in order to have any benefit from feature selection.

   Hint 1: the leave-one-out predictor simply means that for each face i, you find the most similar other face j, and return the pose and lighting angles of that face j as your predictions.

   Hint 2: To find the most similar other face, compute the sum of squared differences between pixels of face i and corresponding pixels of face j. The face with smallest sum of squared differences is the nearest face.

   Hint 3: To compute the leave-one-out error, create the predictions (the three predicted angles) for each face i as in Hint 1. Then, for each face i, compute the squared error (sum of squared differences) between the predicted angles and the true angles of face i. Lastly, sum the squared error over all the faces.

b) Implement forward selection of features, as described in slide 28 of lecture 2. Run forward selection on the Sculpt Faces data, stopping when the perfomance does not improve anymore. What is the leave-one-out validation error that you reached? Is it better than the corresponding value using all features? How many features were needed?

c) Implement a variant of the forward selection: instead of stopping when the performance measure does not improve, add the best feature in each iteration even if it does not improve the performance, and continue to run the algorithm in this way until all the features have been added to the feature set. Run this variant on the Sculpt Faces data set: record the order in which the features were added, and the performance achieved with each number of features.

d) Analyze the results of the previous step:

   - Did you, at any point in the algorithm, reach a better solution than in step b)? Why?
   - Plot a curve of the performance measure with respect to the number of features.

Problem A2: Variable Ranking

1. Discuss the use of simple variable ranking methods for the Sculpt Faces data set: could, for example, ranking by Pearson correlation be used, how?

2. As a more complicated ranking function, it was suggested to "make a non-linear fit of the target with simple variables, rank by goodness of fit" (lecture 2, slide 9). Let us try this, using squared error of leave-one-out nearest neighbor prediction as the goodness of fit. In other words, compute the squared error of leave-one-out nearest-neighbor prediction using each variable alone, and rank the variables by those errors. How does this ranking differ from the order that the variables were added in problem 1, step c)?