CMSC 25300 / 35300
Homework 2

1. Let

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

a) What is the rank of $\boldsymbol{X}$?

b) Suppose that $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{w}$. Derive an explicit formula for $\boldsymbol{w}$ in terms of $\boldsymbol{y}$.

2. Answer the following questions. Make sure to explain your reasoning.

a) Are the columns of the following matrix linearly independent?

$$\boldsymbol{X} = \begin{bmatrix} +0.92 & +0.92 \\ -0.92 & +0.92 \\ +0.92 & -0.92 \\ -0.92 & -0.92 \end{bmatrix}$$

b) Are the columns of the following matrix linearly independent?

$$\boldsymbol{X} = \begin{bmatrix} +1 & +1 & +1 \\ -1 & +1 & -1 \\ +1 & -1 & -1 \end{bmatrix}$$

c) Are the columns of the following matrix linearly independent?

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 & 2 \\ 3 & 4 & 5 \\ 5 & 6 & 8 \end{bmatrix}$$

d) What is the rank of the following matrix?

$$\boldsymbol{X} = \begin{bmatrix} +5 & +2 \\ -5 & +2 \\ +5 & -2 \end{bmatrix}$$

e) Suppose the matrix in part d) is used in the least squares optimization $\min_{\boldsymbol{w}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2$. Does a unique solution exist?

3. Consider the following matrix and vector:

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}.$$

a) Find the solution $\widehat{\boldsymbol{w}}$ to $\min_{\boldsymbol{w}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2$.

b) Make a sketch of the geometry of this particular problem in $\mathbb{R}^3$, showing the columns of $\boldsymbol{X}$, the plane they span, the target vector $\boldsymbol{y}$, the residual vector and the solution $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{w}}$.

4. Design classifier to detect if a face image is happy.

Consider the two faces below. It is easy for a human, like yourself, to decide which is happy and which is not. Can we get a machine to do it?

The key to this classification task is to find good features that may help to discriminate between happy and mad faces. What features do we pay attention to? The eyes, the mouth, maybe the brow?

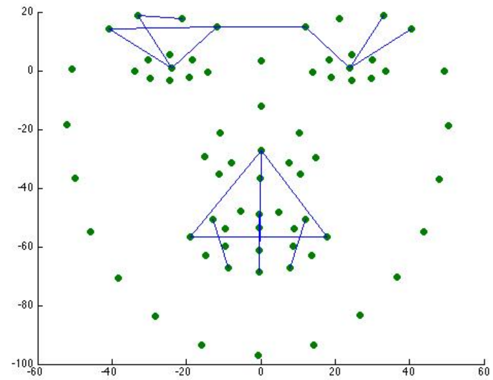
The image below depicts a set of points or "landmarks" that can be automatically detected in a face image (notice there are points corresponding to the eyes, the brows, the nose, and the mouth). The distances between pairs of these points can indicate the facial expression, such as a smile or a frown. We chose $p = 9$ of these distances as features for a classification algorithm. The features extracted from $n = 128$ face images (like the two shown above) are stored in the $n \times p$ matrix $\mathbf{X}$ in the Matlab file `face_emotion_data.mat`. This file also includes an $n \times 1$ binary vector $\mathbf{y}$; happy faces are labeled $+1$ and mad faces are labeled -1. The goal is to find a set of weights for the features in order to predict whether the emotion of a face image is happy or mad.

a) Use the training data $\mathbf{X}$ and $\mathbf{y}$ to find an good set of weights.

b) How would you use these weights to classify a new face image as happy or mad?

c) Which features seem to be most important? Justify your answer.

d) Can you design a classifier based on just 3 of the 9 features? Which 3 would you choose? How would you build a classifier?



e) A common method for estimating the performance of a classifier is cross-validation (CV). CV works like this. Divide the dataset into 8 equal sized subsets (e.g., examples $1-16$, $17-32$, etc). Use 7 sets of the data to chose your weights, then use the weights to predict the labels of the remaining "hold-out" set. Compute the number of mistakes made on this hold-out set and divide that number by 16 (the size of the set) to estimate the error rate. Repeat this process 8 times (for the 8 different choices of the hold-out set) and average the error rates to obtain a final estimate.

f) What is the estimated error rate using all 9 features? What is it using the 3 features you chose in (d) above?

5. Polynomial fitting. Suppose we observe pairs of scalar points (z_i, y_i), $i = 1, \ldots, n$. Imagine these points are measurements from a scientific experiment. The variables z_i are the experimental conditions and the y_i correspond to the measured response in each condition. Suppose we wish to fit a degree $d < n$ polynomial to these data. In other words, we want to find the coefficients of a degree d polynomial p so that $p(z_i) \approx y_i$ for $i = 1, 2, \ldots, n$. We will set this up as a least-squares problem.

a) Suppose p is a degree d polynomial. Write the general expression for $p(z) = y$.

b) Express the $i = 1, \ldots, n$ equations as a system in matrix form $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$. Specifically, what is the form/structure of $\boldsymbol{X}$ in terms of the given x_i.

c) Write a Matlab or Python script to find the least-squares fit to the $n = 30$ data points in
`polydata.mat`. Plot the points and the polynomial fits for $d = 1, 2, 3$.