Part I (40 points)

(1) The function
\[ f(x) = 9x^4 - 56x^3 + 102x^2 - 72x \]
may have maxima or minima at \( x = \frac{2}{3}, 1, \) and 3.

(a) Determine if the function has a local minimum or local maximum at each point.

(b) Can you say that one of the local minima is the global minimum? If so, what is the minimum and the argmin?

(2) Consider the linear equation \( y = \frac{3}{2}x + 2. \)

(a) Write this equation in Hesse normal form \( (\mathbf{a} \cdot \mathbf{x} = d). \)

(b) Plot this line and draw (and label) \( \mathbf{a}, \hat{\mathbf{a}}, \) and \( d. \)

(c) Find the intersection between this line and the line \( -3x - 2y = -1 \) using Gaussian elimination.

(3) Write a quadratic program (QP) for SVM that calculates a classifying line to separate the points \((1, 1)\) and \((-1, 2)\) from the points \((-3, -4)\) and \((2, -1)\). You do not need to solve the resulting QP.

(4) Decompose the vector \( \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} \) over the vectors \( \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\} \).

Part II: Machine Problem (60 points)

A team of researchers used DNA microarrays to measure gene expression in a large set of breast cancer cell lines (Kao, et. al, PLOS ONE 4(7): e6146. doi:10.1371/journal.pone.0006146). In this exercise, you will use gene expression profiles from this study to build a classifier that differentiates between invasive and regular ductal carcinoma (IDC and DC).

(1) Load the mat file `HW3_data.mat`, which contains the following variables:
• **training_lines** is a Matlab table containing gene expression data for the IDC and DC cell lines. Each of the 8750 rows corresponds to a gene with variable expression across the cell lines. Each of the 28 columns represents a cell line. The following cell lines were classified as invasive (IDC) by a pathologist: BT474, BT483, BT549, EFM19, MDA134, MDA175, SUM102, T47D, UACC812, UACC893, ZR75.1, and ZR75.30. The remaining cell lines are noninvasive ductal carcinoma (DC).

• **patient_samples** is a Matlab table containing gene expression values for the same 8750 genes from the training data. Each column corresponds to a different patient biopsy.

(2) Build an SVM classifier that separates IDC from DC samples.

• The Matlab command `fitcsvm` accepts numerical arrays, not tables, so convert your table with the function `table2array`.

• Pay attention to the dimensions of your inputs, especially what rows and columns correspond to in your data and for `fitcsvm`.

(3) Perform both $k$-fold (with 4 folds) and leave-one-out cross validations using the command `crossval`. Using the function `kfoldLoss`, report the accuracy of your model using each validation method.

(4) Repeat the cross validation five times for both the $k$-fold and leave-one-out methods. Does the accuracy change for either method? Why or why not?

(5) Using the Matlab `predict` function, determine if each biopsy in the patient data set is invasive (IDC) or regular (DC) ductal carcinoma.

Remember to submit all code, outputs, and explanations for these problems.