Assignment 2 - AIT 512 Maths for ML

Due on **October 04, 2025**

September 27, 2025

The Iris Dataset (`https://archive.ics.uci.edu/dataset/53/iris`) is a well-known dataset used to test classifiers; it describes 4 attributes, per plant, of 3 different plants, namely `Iris-setosa`, `Iris-versicolor`, and `Iris-virginica`. It has 150 records with 50 records of each plant; compose the matrix of this dataset as $\mathbf{X} \in \Re^{4\times 150}$.

1. **Least-squares classifier**: Build a least-squares classifier to classify `Iris-setosa` from the other two plants; this is a binary classification problem. Present the confusion matrix (see the section on binary confusion matrix here: `https://en.wikipedia.org/wiki/Confusion_matrix`) for this classifier and comment on it. **Hints**:

   - Assign Label $y^i = +1$ to the data points corresponding to `Iris-setosa` and $y^i = -1$ for the rest.
   - Build a regression model $\overrightarrow{y} = \mathbf{X}^T\overrightarrow{\beta} + \nu\overrightarrow{1}$, where $\overrightarrow{y} \in \Re^{150\times 1}$ is the vector of labels and $\overrightarrow{\beta} \in \Re^{4\times 1}$, $\nu \in \Re$ are the regression parameters. These can be found by *minimizing* $\left\|\overrightarrow{y} - \mathbf{X}^T\overrightarrow{\beta} - \nu\overrightarrow{1}\right\|^2$; in turn, this can be solved by minimizing a quadratic expression, say using `https://pypi.org/project/quadprog/`.
   - The least-squares classifier, which yields the predicted labels, $\hat{y}_i$, itself is given by $\hat{y}_i = \operatorname{sign}\left(\overrightarrow{x}_i^T\overrightarrow{\beta} + \nu\right)$, where $\overrightarrow{x}_i$ is the column describing the attributes for datapoint i from the matrix $\mathbf{X}$.

2. **Least-squares classifier for a reduced-dimension dataset**: Redesign the least-squares classifier, if possible, by reducing the dimension of the Iris dataset. Present the confusion matrix and compare performance with that obtained using the full-order dataset. Did reducing the dimensions change the performance drastically? If not, why?

   Hints:

   - Perform an SVD for the matrix $\mathbf{X} \in \Re^{4\times 150}$ and judiciously choose the largest singular values.
   - Suppose the first 2 singular values are selected, so that $\hat{\mathbf{X}} = \sum_{i=1}^{2} \sigma_i \overrightarrow{u}_i \overrightarrow{v}_i^T$ is the rank-2 approximation of this dataset; note that $\overrightarrow{u}_i, \overrightarrow{v}_i$ are the corresponding singular vectors. Note that the dimension of $\hat{\mathbf{X}}$ is still 4×150.
   - A reduced dimension dataset (of dimension 2×150) can be determined from the product $\mathbf{U}^T\hat{\mathbf{X}}$, where $\mathbf{U}$ is the matrix of singular vectors $\overrightarrow{u}_i$. (Observe the values of the elements of this product). The reduced dimension dataset should be similar to the plot shown in Fig. 1. Comment on why the operation $\mathbf{U}^T\hat{\mathbf{X}}$ reduces the dimension.

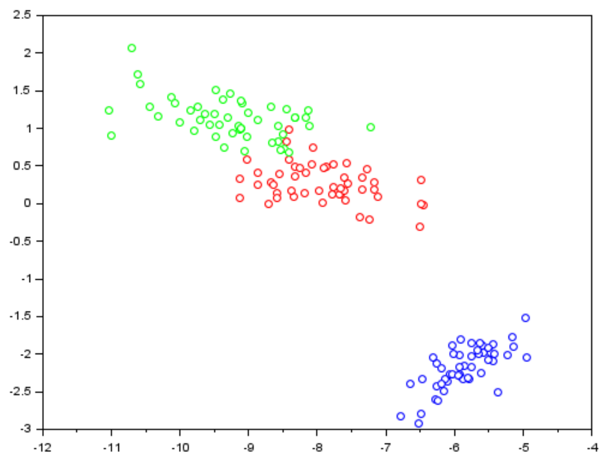

Figure 1: Iris Dataset with reduced dimensions