1. Consider a linear machine with discriminant $g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$.

   (a) Show that the distance from the hyperplane $g(\mathbf{x}) = 0$ to a point $\mathbf{x}_a$ is $|g(\mathbf{x}_a)|/\|\mathbf{w}\|$, where $\|\cdot\|$ is the Euclidean norm, by minimizing $\|\mathbf{x} - \mathbf{x}_a\|^2$ subject to the constraint $g(\mathbf{x}) = 0$.

   (b) Show that the projection of $\mathbf{x}_a$ onto the hyperplane is given by

   $$\mathbf{x}_p = \mathbf{x}_a - \frac{g(\mathbf{x}_a)}{\|\mathbf{w}\|^2}\mathbf{w}.$$

2. For this problem, you will use the data provided in 'DiabetesTraining.csv'. The file contains the values of 8 features (in columns 1 through 8) and the label in the last column (0 - no diabetes, 1 - diabetes). The feature vector contains both categorical and numerical features.

   (a) Using the numerical features 'age', 'BMI,' 'HbA1c level', and 'blood glucose level', find the vector $\mathbf{w}$ to be used for linear discriminant analysis for binary classification. You should standardize the data first by dividing the feature values by respective standard deviation.

   (b) Suppose that we are interested in constructing a decision tree using all 8 features. Compute the Gini impurity and information gain for attributes 'hypertension' and 'heart disease' at the beginning.