Assignment

One of the most difficult problems in data science is choosing a feature set for your models. Once a feature set is chosen, it is possible that some features are redundant (e.g. feature X_k is a linear combination of X_i and X_j) whereas some features are more important than others. For example, in our setting we use two features (μ, σ) to describe each week. Do we need them both? How much accuracy can be attributed to each?

In this assignment, we will use SHapley Additive exPlanations (SHAP) to attribute accuracy in your models to each feature. The basic idea is a concept from cooperative game theory named in honor of Lloyd Shapley. Think of each feature as a player in a game. The accuracy of your model is the "payout". To attribute the contribution to the payout from each player i (feature), you remove that player and let the remaining players play the game. If you have n features $X_1, X_2, \ldots, X_n$, consider reduced feature sets:

- set $Y_1 = (\text{absent}, X_2, \ldots, X_n)$: feature X_1 is removed
- set $Y_2 = (X_1, \text{absent}, \ldots, X_n)$: feature X_2 is removed
- set $Y_n = (X_1, X_2, \ldots, \text{absent})$: feature X_n is removed

Let Accuracy(X) denote the accuracy of your predictive model when all features are used. For each "reduced" set of features Y_i, run your model with these features and compute the accuracy Accuracy(Y_i). For each feature (player) X_i compute its "marginal" contribution Δ_i to accuracy (payout) as follows:

$$\Delta_i = \text{Accuracy}(X) - \text{Accuracy}(Y_i)$$

As an example, consider predicting week labels using logistic regression. Your feature set $X = (\mu, \sigma)$. You trained your model in year 1 and computed its accuracy Accuracy(X) in year 2. Now we do the following: remove μ from the feature set and use $Y_1 = (\sigma)$. Using only σ to train your model in year 1, predict your labels in year 2 and compute its accuracy Accuracy(Y_1). Next, take your original X and remove the second feature σ. Use the reduced feature set $Y_2 = (\mu)$ to train your model in year 1, predict your labels for year 2 and compute its accuracy Accuracy(Y_1).

Compute the marginal contributions of features μ and σ to accuracy:

$$\begin{aligned}
\Delta_1 &= \text{Accuracy}(X) - \text{Accuracy}(Y_1) \\
\Delta_2 &= \text{Accuracy}(X) - \text{Accuracy}(Y_2)
\end{aligned}$$

Questions:

1. compute the contributions of μ and σ for logistic regression, Euclidean kNN and (degree 1) linear model. Summarize them in a table and discuss your findings.

2. take IRIS dataset. In this set we have 3 flowers (Iris Versicolor, Iris Setosa and Iris Virginica) and 4 features (sepal length, sepal width, petal length, petal width). We can assign multiple labels by training 3 distinct binary classifiers using "one-vs-all" classification method. In this method, you take labels of one class as positive and the rest of the classesusing as negative. For example, take Iris Versicolor as one class and the remaining two (Iris Setosa and Iris Virginica) as the second class. We then use a binary classifier using binary classification. (Making decisions means applying all classifiers to an unseen sample x and predicting the label k for which the corresponding classifier reports confidence score). For each of the 3 flower types, construct a one-vs-all logistic regression classifier. For each of the of the 4 features (sepal length, sepal width, petal length, petal width) compute its marginal contributions Δ to accuracy (split the dataset 50/50 into training and testing parts). Summarize your findings in a table (shown below) and discuss them

Flower	Versicolor	Setosa	Virginica
sepal length Δ			
sepal width Δ			
petal length Δ			
petal width Δ			