1. Explain the k-nearest neighbour classification algorithm (<https://scikit-learn.org/stable/modules/neighbors.html>).

Explain hyperparameters k and weight.

Explain the K-D tree algorithm and how it is used in the k-nearest neighbour classification algorithm.

Evaluate the k-nearest neighbour classification algorithm in the Fashion-MNIST dataset (compare with logistic regression).

1. Explain multiclass classification strategies:

one-vs-the-rest

one-vs-one

error correcting output codes

(<https://scikit-learn.org/stable/modules/classes.html#module-sklearn.multiclass>)

Compare these strategies for logistic regression on the Fashion-MNIST dataset.

1. Explain L1, L2 and elastic-net regularization terms in logistic regression (https://scikit-learn.org/stable/auto\_examples/linear\_model/plot\_logistic\_l1\_l2\_sparsity.html). Compare these in the Fashion-MNIST dataset. Report accuracy and sparsity results.
2. Explain what self-training is in the context of semi-supervised learning. Apply it on the breast cancer dataset (<https://scikit-learn.org/stable/auto_examples/semi_supervised/plot_self_training_varying_threshold.html#sphx-glr-auto-examples-semi-supervised-plot-self-training-varying-threshold-py>)
3. Explain permutation-based feature selection (https://scikit-learn.org/stable/modules/permutation\_importance.html#permutation-importance).

Apply it to the logistic regression and Random Forest algorithms on the breast cancer dataset (<https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html>).

1. Explain recursive feature elimination with cross-validation (<https://scikit-learn.org/stable/auto_examples/feature_selection/plot_rfe_with_cross_validation.html#sphx-glr-auto-examples-feature-selection-plot-rfe-with-cross-validation-py>). Apply it to the logistic regression and Random Forest algorithms on the breast cancer dataset (<https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html>).
2. Explain out-of-bag errors in the Random Forest algorithm (https://scikit-learn.org/stable/auto\_examples/ensemble/plot\_ensemble\_oob.html). Compare it to cross-validation on the breast cancer dataset ( <https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html>).
3. Explain k-nearest neighbour missing value imputation (https://scikit-learn.org/stable/auto\_examples/impute/plot\_missing\_values.html#sphx-glr-auto-examples-impute-plot-missing-values-py). Compare it to mean value imputation on the diabetes dataset (https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\_diabetes.html#sklearn.datasets.load\_diabetes).

x²