The American University in Cairo
Onsi Sawiris School of Business

AACSB ACCREDITED · ASSOCIATION OF MBAS ACCREDITED · EQUIS ACCREDITED

Labs Guidelines

Machine Learning and Artificial Intelligence in Finance

FINC 5322

Fall 2025

The American University in Cairo

Onsi Sawiris School of Business

Heikal Department of Management

KNN and Linear Regression – Assignment 2

In this assignment, you will explore two core supervised learning algorithms: K-Nearest Neighbors (KNN) and Linear Regression, using simple datasets.

The goal is not to build complex models but to understand model behavior: how parameter choices, noise, and outliers affect results and model stability. You will visualize results, interpret findings, and reflect on key machine-learning concepts such as overfitting, bias-variance trade-off, and model robustness.

Accessing the Assignment

- All assignments are hosted on our course Code Repository (GitHub): https://github.com/kelkess43/AUC-Material
- The assignment will also be shared weekly through Canvas modules for easy access.

Structure of the Assignment

Your notebook is divided into nine Sections.

You must complete the tasks and reflections in the notebook and keep all work within the. ipynb file.

All questions and grades are mentioned in the notebook.

Question 1 (10 points) – Visualize the Dataset

Create a scatter plot of sepal length vs sepal width, color-coded by species.

Label axes and add a legend and title.

Deliverable: Scatter plot with labels and captions.

Question 2 (10 points) – Split the Data

Split the data into training (80 %) and testing (20 %) sets using a fixed random state.

Print the number of samples in each set.

Deliverable: Code output showing training/test sizes.

Question 4 (15 points) – Train a Baseline KNN Model

Train a KNN classifier with k = 5 using the default minkowski distance.

Report model accuracy, confusion matrix, and classification report.

Deliverable: Evaluation metrics and code used to generate them.

Question 5 (15 points) – Effect of Different k Values
Train multiple KNN models for k = 1 to 20.
Plot training and testing accuracy vs k and explain which k performs best.
Discuss how small and large k values relate to **overfitting** and **underfitting**.
Deliverable: Accuracy plot and short written explanation.

Question 6 (10 points) – Visualize Decision Boundaries
Visualize how KNN separates classes for different distance metrics.
Use a meshgrid to create regions and overlay the test points.
Compare how boundaries differ among Euclidean, Manhattan, Chebyshev, and Minkowski metrics.
Deliverable: Decision-boundary plots for each metric and brief comparison.

Question 7 (15 points) – Generate Synthetic Data
Generate two synthetic regression datasets using make_regression:

- Scenario 1: Low Noise (noise=10)
- Scenario 2: High Noise (noise=50)

Deliverable: Two labeled plots and short interpretation of differences.

Question 8 (20 points) – Fit Linear Regression Models
Fit a Linear Regression model to both datasets (low and high noise).
Plot the fitted regression line on top of each scatter plot.
Compare how the noise level affects model fit and line stability.
Deliverable: Two plots showing fitted regression lines and a short-written explanation.

Question 9 (15 points) – Analyze Model Parameters
Compare the slope (coefficient) and bias (intercept) of both models.
Discuss how higher noise levels affect the model parameters and stability.
Deliverable: Printed slope/intercept values and a short-written analysis (3–5 sentences).

Deliverables and Submission

- Submit one Jupyter Notebook (.ipynb) file with all your work.
- Make sure all cells are executed, and outputs are visible.
- Add short comments/markdown where required
- Deadline: Wednseday, 15th October at 11:59 PM
- Submission: Upload your completed. ipynb notebook on Canvas