

MA289 Intro to Statistical Learning

Project 2: Binary Classification with Image Data

MAJ Patrick Kuiper

DUE No Later Than 17 March 2026 at Midnight

Submit a **PDF write-up in Canvas** and share your **Google Colab notebook (edit access)** with the instructor.

Generative AI Usage Policy

Generative AI tools may be used for coding or drafting, but all usage must be documented in your notebook. Include: (1) prompts used, (2) a brief summary of the output (link if available), and (3) how you verified or modified the output. Undocumented use will result in a deduction.

Assignment Overview

You will work with the **Fashion-MNIST** dataset (28×28 grayscale images) and restrict the data to a **binary classification task: Sneaker (Class 0) vs. Ankle Boot (Class 1)**. Starter code in the Colab notebook loads, filters, preprocesses, and visualizes the data; you are expected to run and inspect it before modeling.

Objective

Develop and evaluate **two classifiers**: one **linear** and one **non-linear or distance-based**. For each model, perform proper preprocessing, cross-validation for hyperparameter tuning, and full performance evaluation. Use Chapters 2 and 3 of Géron as guidance.

Modeling Requirements (For Each Classifier)

For **each** classifier, you must complete the following:

1. **Confusion Matrix (Test Set)** Provide a confusion matrix on the held-out test set and clearly explain:
 - True Positives
 - True Negatives
 - False Positives
 - False Negatives

2. **Cross-Validated Performance Metrics** Compute cross-validated:

Precision, Recall, F_1 -score

3. **Precision–Recall Curve** Plot a properly labeled Precision–Recall curve.

4. **ROC Curve** Plot a properly labeled ROC curve.

5. **Hyperparameter Tuning** Use cross-validation on the training set to tune relevant hyperparameters. Clearly report:

- Parameters tuned
- Search method used
- Best parameter values

6. **Sensitivity to Standardization** Standardize pixel values and compare model performance before and after scaling. Discuss which classifier is more sensitive to feature scaling and why.

Written Analysis

At the end of your notebook, include a clear and well-developed written analysis (minimum two paragraphs) that:

- Compares the two classifiers
- Interprets precision, recall, F_1 , and ROC curves
- Explains hyperparameter choices
- Discusses why one model outperformed the other
- Evaluates sensitivity to standardization

Your discussion should reference specific numerical results from your analysis.

Grading Rubric (300 Points)

Criterion	Points	Score
1. Model Implementation (Both Classifiers)	100	
– Correct preprocessing and train/test split	25	_____
– Proper cross-validation and hyperparameter tuning	35	_____
– Correct confusion matrices and evaluation metrics	40	_____
2. Performance Evaluation and Visualization	100	
– Precision, Recall, F_1 correctly computed	30	_____
– Properly labeled PR and ROC curves	40	_____
– Analysis of scaling sensitivity	30	_____
3. Written Analysis and Interpretation	100	
– Clear comparison of classifiers	40	_____
– Insightful interpretation of metrics	30	_____
– Thoughtful discussion of hyperparameters and results	30	_____
Total	300	_____