University of the Philippines Cebu

College of Science
Department of Computer Science
032 232 8187 dcs.upcebu@up.edu.ph

COURSE GUIDE

Course Number	CMSC 173
Course Title	Machine Learning
Number of Units	3 units (3 hours lecture)
Prerequisites	CMSC 170
Instructor	Noel Jeffrey Pinton

COURSE OUTLINE

I. **Introduction to Machine Learning**
   a. What is Machine Learning?
   b. Supervised Learning
   c. Unsupervised Learning
   d. Semi-supervised Learning
   e. Reinforcement Learning
   f. Machine Learning Workflow
   g. Applications and Ethics

II. **Parameter Estimation**
   a. Method of Moments
   b. Maximum Likelihood Estimation
   c. Maximum A Posteriori Estimation

III. **Linear Regression**
   a. Simple Linear Regression
   b. Multiple Linear Regression
   c. Least Squares Method
   d. Gradient Descent
   e. Model Evaluation Metrics

IV. **Regularization**
   a. Overfitting and Underfitting
   b. Ridge Regression (L2)
   c. Lasso Regression (L1)
   d. Elastic Net
   e. Regularization Parameter Selection

V. **Exploratory Data Analysis**
   a. Data Preprocessing
   b. Feature Engineering
   c. Data Visualization
   d. Handling Missing Values
   e. Outlier Detection

VI. **Model Selection and Evaluation**
   a. Bias-Variance Tradeoff
   b. Cross-Validation (k-fold, Leave-One-Out)
   c. Evaluation Metrics (RMSE, MAE, R^2)
   d. Hyperparameter Tuning
   e. Model Comparison Strategies

VII. **Dimensionality Reduction**

a. Curse of Dimensionality
b. Principal Component Analysis (PCA)
c. Kernel PCA
d. Feature Selection vs. Feature Extraction
e. Applications (Visualization, Compression, Noise Filtering)

VIII. **Classification**

a. Logistic Regression
b. Naïve Bayes Classifier
c. K-Nearest Neighbors (KNN)
d. Decision Trees
e. Support Vector Machines (SVM)
f. Kernel Methods
g. Classification Evaluation Metrics

IX. **Clustering**

a. Partitional Clustering (K-means, K-medoids)
b. Hierarchical Clustering (Agglomerative, Divisive)
c. Density-Based Clustering (DBSCAN)
d. Clustering Evaluation Metrics
e. Applications of Dimensionality Reduction in Clustering

X. **Neural Networks**

a. Perceptron and Multilayer Perceptron
b. Activation Functions
c. Backpropagation Algorithm
d. Optimization Methods (SGD, Adam, RMSprop)
e. Regularization in Neural Networks (Dropout, Batch Normalization)

XI. **Advanced Neural Network Architectures**

a. Convolutional Neural Networks (CNNs)
b. Recurrent Neural Networks (RNNs)
c. Transformers and Attention Mechanisms
d. Generative Adversarial Networks (GANs)
e. Variational Autoencoders (VAEs)
f. Diffusion Models

Grading Scale

96 – 100	– 1.0
91 – 95	– 1.25
86 – 90	– 1.5
81 – 85	– 1.75
76 – 80	– 2.0
72 – 75	– 2.25
68 – 71	– 2.5
64 – 67	– 2.75
60 – 63	– 3.0
50 – 59	– 4.0
<50	– 5.0

Grading System

Machine Problems and Exams	50%
Group Project	50%
Total 100%	

Group Project Guidelines and Process

Group Formation

- Groups may have up to **3 members**.
- Groups of **2 members** or even **solo projects** are allowed, but the project scope should be scaled accordingly.
- Each group must register by **Midterms** with names, emails, and a tentative project title.

1. Project Proposal

Deliverable: 2–3 page write-up.

- Title – concise and descriptive.
- Problem Statement – define the machine learning task (classification, regression, clustering, etc.).
- Motivation and Importance – why the problem matters (social, industry, academic relevance).
- Objectives – specific aims of the project.
- Proposed Dataset – source and characteristics of the data.
- Planned Methods – initial choice of algorithms or models.

Instructor Checkpoints: feasibility, ethical concerns, scope appropriateness.

2. Data Gathering

- Use open datasets (Kaggle, UCI, government portals, etc.) or collect your own (e.g., scraping, surveys).
- Ensure dataset size is sufficient ($n \geq 500$ preferred).
- Document the source, license, and collection methodology.

Checkpoint: Dataset must be approved by the end of Week 5.

3. Data Cleaning & Preprocessing

- Handle missing values (imputation, removal).
- Outlier detection (z-score, IQR, visualization).
- Feature engineering (encoding, normalization, dimensionality reduction).
- Data splitting (e.g., 70–15–15 train/validation/test).

Deliverable: Report + Jupyter notebook of preprocessing steps.

4. Model Training

- Establish baselines: regression, decision trees, Naïve Bayes, KNN.
- Explore advanced models: SVM, Random Forest, Gradient Boosting, Neural Networks.
- Apply regularization (Lasso, Ridge) to prevent overfitting.
- Hyperparameter tuning via grid/random search.

5. Model Evaluation

Metrics:

- Regression: RMSE, MAE, R^2.
- Classification: Accuracy, Precision, Recall, F1, ROC-AUC.
- Clustering: Silhouette score, Davies-Bouldin index.

Deliverables: Evaluation report with metrics, baseline vs. advanced model comparisons, and visualizations (confusion matrix, ROC curves, error plots).

6. Final Presentation & Report

Presentation (10 min per group):

1. Introduction and Motivation
2. Data and Preprocessing
3. Methods and Models
4. Results and Evaluation
5. Discussion and Limitations

Final Report:

- Must document the full pipeline: proposal, preprocessing, training, evaluation, conclusions.
- Include mathematical formulations of at least one chosen model.
- Submit reproducible code (well-documented Jupyter notebook).

Grading Breakdown

Component	Weight
Proposal	10%
Data & Preprocessing	15%
Model Training	20%
Evaluation	20%
Final Report	20%
Presentation	15%

Rubric

Criteria	Excellent (4)	Good (3)	Fair (2)	Poor (1)
Proposal	Clear, original, feasible, well-motivated	Clear but lacks originality or depth	Somewhat unclear, weak justification	Incomplete, vague, or irrelevant
Data & Preprocessing	Well-documented, rigorous cleaning, justified methods	Adequate cleaning, some justification	Minimal preprocessing, limited documentation	Missing or inappropriate methods
Model Training	Comprehensive, compares multiple models, well-tuned	Uses several models, some tuning	Limited models, basic training	Single model, poorly trained
Evaluation	Uses correct metrics, strong analysis, insightful comparisons	Appropriate metrics, some analysis	Limited metrics, shallow analysis	Missing or incorrect evaluation
Final Report	Professional, detailed, reproducible, includes math	Clear, complete, minor gaps	Some missing sections, limited detail	Poorly written, incomplete
Presentation	Engaging, clear visuals, excellent delivery	Clear, good visuals, adequate delivery	Somewhat unclear, weak visuals	Unclear, disorganized, poor delivery

Report Grading Process

To ensure fairness and real-world alignment, the final project reports will be graded not only by the instructor but also by invited professors from related fields (e.g., Computer Science, Statistics, Engineering) and practitioners from the data science and machine learning industry. These external evaluators will use the same rubric above when assigning scores. The final grade will be computed as an aggregate of the evaluations.

Suggested References

1. Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: Springer.
2. Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT Press.
3. Alpaydin, E. (2021). *Machine learning*. MIT Press.