

	VIETTEL AI RACE HƯỚNG DẪN THỰC HÀNH TRÍ TUỆ NHÂN TẠO VÀ HỌC MÁY BẰNG PYTHON	Public 476 Lần ban hành: 1
---	--	-----------------------------------

1. GIỚI THIỆU VỀ PYTHON VÀ AI

Python là ngôn ngữ lập trình phổ biến cho nghiên cứu và ứng dụng trí tuệ nhân tạo (AI) và học máy (Machine Learning). Trong tài liệu này, chúng ta sẽ tìm hiểu cách mô phỏng và thực hành các mô hình AI cơ bản bằng Python, với mục tiêu giúp học viên hiểu và áp dụng lý thuyết vào thực tế.

2. DỮ LIỆU VÀ MÔ HÌNH HỌC MÁY

2.1 Yêu cầu trước khi làm thí nghiệm

- Có kiến thức cơ bản về Python.
- Hiểu các khái niệm thống kê: trung bình, phương sai, độ lệch chuẩn.
 - Cài đặt thư viện: numpy, pandas, scikit-learn, matplotlib.

2.2 Mục đích của phần thí nghiệm

Dùng MATLAB mô phỏng các nội dung sau:

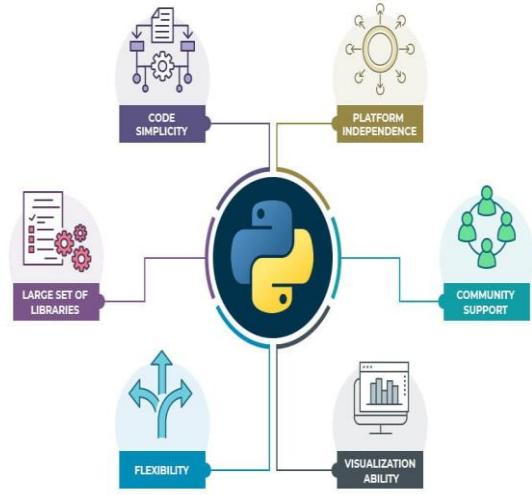
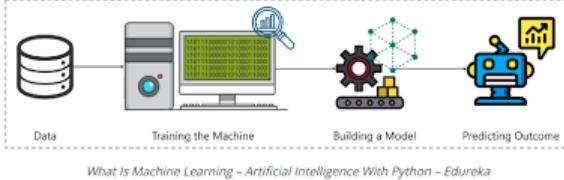
- Làm quen với quy trình xử lý dữ liệu và huấn luyện mô hình.
- Hiểu cách áp dụng các thuật toán cơ bản và đánh giá kết quả

2.3 Tóm tắt lý thuyết

2.3.1 Định nghĩa cơ bản

Định nghĩa	Công thức
Hồi quy tuyến tính (Linear Regression): tìm tham số w, b để tối thiểu hóa MSE	$\hat{y} = w^T x + b$ $MSE = (1/n) \sum (y_i - \hat{y}_i)^2$

	VIETTEL AI RACE HƯỚNG DẪN THỰC HÀNH TRÍ TUỆ NHÂN TẠO VÀ HỌC MÁY BẰNG PYTHON	Public 476
		Lần ban hành: 1

Hồi quy logistic (Logistic Regression): phân loại nhị phân với hàm sigmoid	$\hat{y} = \sigma(z) = 1 / (1 + e^{-z}), z = w^T x + b$ $Loss = -(1/n) \sum [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$
Mạng nơ-ron (Neural Networks): nhiều tầng tuyến tính kết hợp hàm kích hoạt	$a^l(l) = f(W^l a^{l-1} + b^l)$ 
Chuẩn hóa dữ liệu (Standardization)	$z = (x - \mu) / \sigma$  What Is Machine Learning – Artificial Intelligence With Python – Edureka
Dánh giá mô hình	$Accuracy = (TP + TN) / (TP + TN + FP + FN)$ $Precision = TP / (TP + FP)$ $Recall = TP / (TP + FN)$ $F1-score = 2 * (Precision * Recall) / (Precision + Recall)$

	VIETTEL AI RACE	Public 476
	HƯỚNG DẪN THỰC HÀNH TRÍ TUỆ NHÂN TẠO VÀ HỌC MÁY BẰNG PYTHON	Lần ban hành: 1

2.3.2 Một số định nghĩa khác

Gradient Descent: thuật toán tối ưu để cập nhật tham số	$\theta := \theta - \alpha \nabla J(\theta)$
Chuẩn hóa L2 (Ridge Regression)	$\theta := \theta - \alpha \nabla J(\theta)$
Chuẩn hóa L1 (Lasso Regression)	$J(w) = (1/n) \sum (y_i - \hat{y}_i)^2 + \lambda w _1$
Phân tích thành phần chính (Principal Component Analysis - PCA)	$\max_{\{w\}} \text{Var}(w^T X), \text{với ràng buộc } w = 1$
K-Means Clustering: gom cụm dữ liệu	$J = \sum \sum x_i - \mu_k ^2, \text{với } \mu_k \text{ là tâm cụm}$
Naive Bayes Classifier: áp dụng định lý Bayes với giả định độc lập	$P(y x) \propto P(y) \prod P(x_i y)$
Support Vector Machine (SVM): tìm siêu phẳng tối ưu	$\min_{\{w, b\}} (1/2) w ^2, \text{với ràng buộc } y_i(w^T x_i + b) \geq 1$

2.3.3 Hệ thống

2.3.3.1. Cross-Entropy Loss (Entropy chéo cho phân loại đa lớp)

[CT1]	$L = - \sum y_i \log(\hat{y}_i)$
-------	----------------------------------

	VIETTEL AI RACE	Public 476
	HƯỚNG DẪN THỰC HÀNH TRÍ TUỆ NHÂN TẠO VÀ HỌC MÁY BẰNG PYTHON	Lần ban hành: 1

2.3.3.2. *Softmax Function: chuẩn hóa xác suất cho phân loại đa lớp*

[CT2]	$\sigma(z)_j = e^{\{z_j\}} / \sum e^{\{z_k\}}$
-------	--