

THỰC HÀNH 1: HỒI QUY TUYẾN TÍNH

Hướng dẫn nộp bài:

Các bạn submit code lên github. Sử dụng 01 file txt để lưu link của github đến bài labc thực hành, đặt tên file txt theo cú pháp <MSSV>.txt và submit lên courses.

Bài 1 (3 điểm): Sử dụng Numpy xây dựng phương pháp hồi quy tuyến tính và tiến hành huấn luyện trên bộ dữ liệu Forest Fires (link: <https://archive.ics.uci.edu/static/public/162/forest+fires.zip>).

Gợi ý: tham số tối ưu $\hat{\theta}$ được xác định bởi $\hat{\theta} = (X^T X)^{-1} X^T y$.

Bài 2 (3 điểm): Chuẩn hoá dữ liệu về phân phối có kỳ vọng là 0 và phương sai là 1. So sánh kết quả của mô hình sau khi được huấn luyện với dữ liệu đã được chuẩn hoá so với kết quả trong bài 1.

Bài 3 (2 điểm): Tiến hành kỹ thuật Feature Engineering, tìm ra các thuộc tính có tính chất colinearity (linearly dependent). Huấn luyện lại phương pháp hồi quy tuyến tính với bộ dữ liệu chỉ giữ lại các thuộc tính uncorrelated.

Bài 4 (2 điểm): Sử dụng thư viện Machine Learning (Sklearn hoặc Skorch) thực thi lại phương pháp Linear Regression.