Politecnico di Bari

FOUNDATION OF MACHINE LEARNING
Exam Session
Prof. Tommaso Di Noia
16/01/2024

This dataset encompasses details about various workers and their corresponding employment levels, featuring a diverse set of attributes ranging from categorical to continuous.
Initiate the data loading process using the suitable Pandas function, and meticulously inspect for any instances of null or duplicated data.

Specifically focusing on the '*salary_in_usd*' feature, identify and eliminate outliers while devising a strategy to address any missing values.

Employ the Pandas ***get_dummies*** method to encode categorical features, namely '*work_year*', '*experience_level*', '*employment_type*', '*job_title*', '*employee_residence*', '*remote_ratio*', '*company_location*', and '*company_size*'.

Following the preprocessing steps, normalize the dataset utilizing the z-score technique to ensure consistent scaling across features. Subsequently, construct a neural network using PyTorch, incorporating 2 hidden layers with 5 and 3 neurons, respectively.
Carefully select an appropriate learning rate and normalization value for optimal model training.

Finally, assess the model's performance using a relevant evaluation metric, ensuring a comprehensive understanding of its effectiveness in handling the given employment dataset.