COMP47750/COMP47990 Tutorial

Regression

1. Assess the performance on the bike share dataset of Linear Regression models that use:
    1. all features except `'casual','registered', 'instant'` and `'dteday'.` (as set up in notebook 12 Regression).
    2. `'temp'` as the only input variable.
       `(X = bikes_df[['temp']].values)`
    3. `'hum'` as the only input.

    Use all the data for training and test.
    You may use LinearRegression or SGDRegressor.
    Score performance using R2, MAPE and mean_absolute_error.

2. For the bike sharing dataset, the weather features are normalized but the calendar features are not. Find the regression coefficients for the following models:
    1. Partially normalized: i.e. original format with only the weather features normalized. (provided in notebook 12 Regression Exercise).
    2. Fully normalized: (also provided) - what happens to the `'mnth'` coefficient?
    3. With a model that uses only the normlaized `'temp'` and `'mnth'` features - what has happened to the `'mnth'` coefficient?

3. Calculate error scores for the following weather predictions:

```
temp =   [12,13,15,12,11,11,17,13,12,14]
t_pred = [12,12,14,13,12,11,15,12,12,13]

rain =   [0,0,5,7,1,1,0,8,0,4]
r_pred = [0,1,4,7,1,0,0,0,1,4]
```

    `t_pred` and `r_pred` are the predicted values.

    Calculate, R^2 MAE, MAPE and RMSE. What do we learn from the different scores? There is code in the notebook **12 Regression Exercise** to get you started.