

# Machine learning for Neuroscience (Tel Aviv University, 2020)

## Exercise 2

---

### Paper and dataset

---

The below questions are based on the following [paper](#):

Sager R, Güsewell S, Rühli F, Bender N, Staub K, "Multiple measures derived from 3D photonic body scans improve predictions of fat and muscle mass in young Swiss men". PLoS ONE, 2020.

In the study, the researchers analysed 104 men aged 19–23. They performed three types of measurements:

- a. Bioelectrical Impedance Analysis (BIA).
- b. Body shape measurements using 3D laser-based photonic scanners.
- c. Manual anthropometric measurements (height, weight, waist circumference).

The researchers have shared the dataset [here](#).

### Questions

---

In the below "Visc" stands for Visceral fat mass, "WC" stands for Waist circumference and "WHtR" stands for Waist to Height Ratio.

Split the data to 80% training set and 20% test set.

1. (25 points) **Linear regression with continuous predictors**

- a. Train a Linear Regression model which predicts "Visc" from "BMI", "Waist circumference" and WHtR ("Waist circumference"/"Height\_value-1").
- b. Report the root mean squared prediction error (RMSE) on the test set.
- c. Report the estimated coefficient's values.
- d. Plot a scatter plot of the predicted target variable v.s. the actual target variable based on the test set.

2. (25 points) **Linear regression with categorical predictors**

Repeat question 1 - where the predictors are transformed to categorical variables based on their quartiles (with levels: 0-25th percentile, 25-50th percentile, 50-75th percentile, 75-100th percentile).

*Hint:* use the function `qcut` (Pandas) and `dmatrix` (patsy). It is recommended to split the dataset after transforming to categorical variables.

3. (25 points) **Piecewise linear regression model**

- a. Plot a scatter plot of Visc v.s. WC
- b. Define two new variables

$$WC1 = WC - 0.78$$

$$WC2 = \begin{cases} 0 & WC < 0.78 \\ 1 & WC \geq 0.78 \end{cases}$$

and train the following linear regression model

$$y = \beta_0 + \beta_1 \cdot WC + \beta_2 \cdot WC1 \cdot WC2$$

- c. Plot the a scatter plot of the predictions on the test set.

4. (25 points) **Logistic regression**

- a. Transform the "Visc" variable to a binary variable (with the value "1" if greater than the median and "0" otherwise).
- b. Train a Logistic regression model using BMI, WC and WHtR as predictors.
- c. Report the AUC on the test set.
- d. Based on a 0.5 threshold, print the confusion matrix, the sensitivity and specificity on the test set.