

## Exercise 1

This is an example for predicting the missed data.

From data frame `df_wind` select only two columns `time` and `velocity` and remove the velocity data for  $4 \text{ s} < \text{time} < 6 \text{ s}$ .

Fit the linear model.

Plot the data and the fitted line.

## Exercise 2

This is an example for training set manipulation.

From data frame `df_wind` select only two columns `time` and `velocity`.

Replace the velocity data for  $4 \text{ s} < \text{time} < 6 \text{ s}$  by the mean velocity computed from the other lines.

Plot the velocity vs. time.

Make a linear fit to these columns.

Plot the data and the fitted line.

What do you conclude ?

## Exercise 3

This is an example of a training set with low statistics (over fitting).

Consider `carF1.csv` and take only 10% of the data for training.

Make the fit and compute the score for both the training and the test sets.

Plot the data and the fitted line together.

What do you conclude ?

## Exercise 4

This is an example for a bad model usage where the features are not well selected (under fitting).

Consider the data `carF1_wind.csv`. Take only these columns : `time` and `velocity_wind`.

Fit a straight line to these columns and compute the score and compare it to the polynomial fit score.

Plot the data and the fitted line together.

What do you conclude ?