
Exercise - Linear Models

Data Modeling of Objects in Free Fall

Exercises in Data Modeling (190.021)

¹*dm@ai-lab.science* February 8, 2026

This cheat sheet is supposed to give some additional information on the Jupyter notebook on linear models.

1 Data Exploration and Model Requirement Identification

To visualize the data, it will be plotted in the form of a so-called scatter plot. The x-axis represents the fall time, whereas the y-axis may represent the weight of the different objects in grams or the height in meters. Therefore, the following values are possible for the parameter *y_axis*:

- "Weight (g)"
- "Height (m)"

Additionally, the data is also plotted as a three-dimensional plot containing all three features of the dataset.

2 Limitations of Linear Models

To decide how good the model fits the data, the mean squared error can be used. By this metric, the error is calculated by the average of the squared distance between the predicted and the actual values. This provides a measure of the accuracy of a model, where smaller values indicate a better fit.

To train a model on our data, we will use a Linear regression, trying to fit a straight line to our data and evaluate it by calculating the mean squared error of that model.

3 Explore a Non-Linear Mixture of Expert Models

Because fitting a straight line to a quadratic function does not yield the best results, we will try a mixture of different linear regression models to model the overall relationship between fall time and height. In this Section you have to decide how many splits the data may be divided into to lead to the best result possible. To solve this task, you can define the parameter *num_splits*, that defines the number of subsets in which the data is divided. You can use integer values ranging between 1 and 34.

4 Explore Locally Weighted Regression

As the final part of this notebook, we will make predictions using a locally weighted regression model. This model is based on the non-linear mixture models defined above. Based on the *num_splits* that can be redefined within this Section, we are trying to predict one point of each subset using the locally weighted regression. A central parameter of the locally weighted regression is the *bandwidth*, which defines the influence of the nearby data points on the prediction. Change the bandwidth parameter to see changes in the prediction. You are allowed to enter values between 0 and 100.