

Fitting the distribution of heights data

December 4, 2019

1 Fitting the distribution of heights data

1.1 Instructions

In this assessment you will write code to perform a steepest descent to fit a Gaussian model to the distribution of heights data that was first introduced in *Mathematics for Machine Learning: Linear Algebra*.

The algorithm is the same as you encountered in *Gradient descent in a sandpit* but this time instead of descending a pre-defined function, we shall descend the χ^2 (chi squared) function which is both a function of the parameters that we are to optimise, but also the data that the model is to fit to.

1.2 How to submit

Complete all the tasks you are asked for in the worksheet. When you have finished and are happy with your code, press the **Submit Assingment** button at the top of this notebook.

1.3 Get started

Run the cell below to load dependancies and generate the first figure in this worksheet.

```
In [6]: # Run this cell first to load the dependancies for this assessment,  
        # and generate the first figure.  
        from readonly.HeightsModule import *
```

1.4 Background

If we have data for the heights of people in a population, it can be plotted as a histogram, i.e., a bar chart where each bar has a width representing a range of heights, and an area which is the probability of finding a person with a height in that range. We can look to model that data with a function, such as a Gaussian, which we can specify with two parameters, rather than holding all the data in the histogram.

The Gaussian function is given as,

$$f(\mathbf{x}; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\mathbf{x} - \mu)^2}{2\sigma^2}\right)$$

The figure above shows the data in orange, the model in magenta, and where they overlap in green. This particular model has not been fit well - there is not a strong overlap.