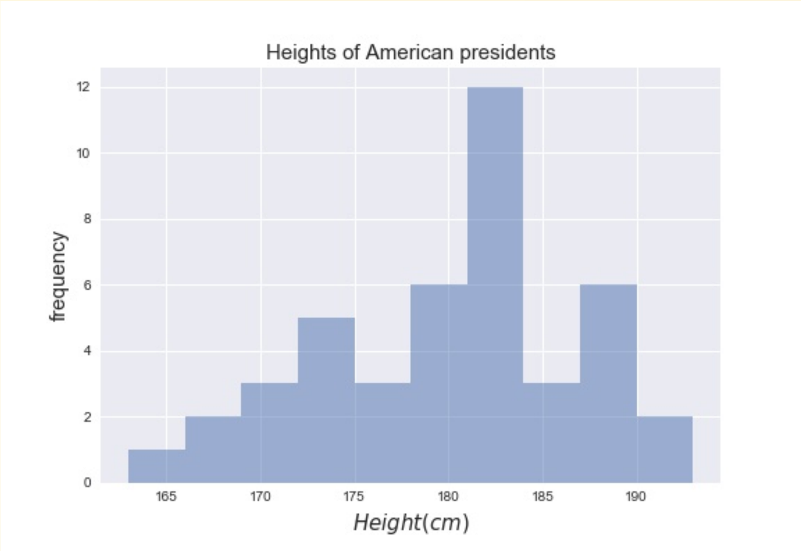
Fitting a Gaussian distribution to data

Objectives:

- Be able to recognise applications of statistical estimation
- Get the basics of maximum likelihood

The data

- The heights of American presidents (43 of them) (from Kaggle):

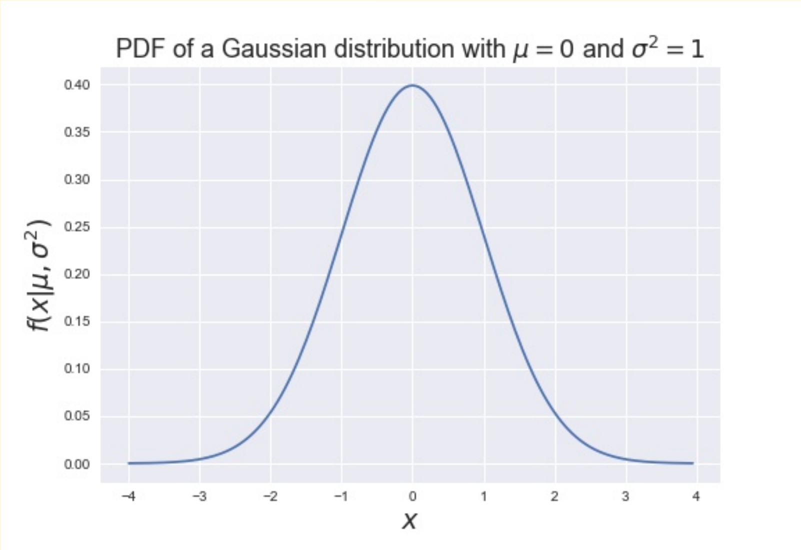


- Consider the height to be realisations of a random variable X_i
- Assume data was generated from a Gaussian (Normal) distribution

The Gaussian distribution

- The probability density function (pdf):

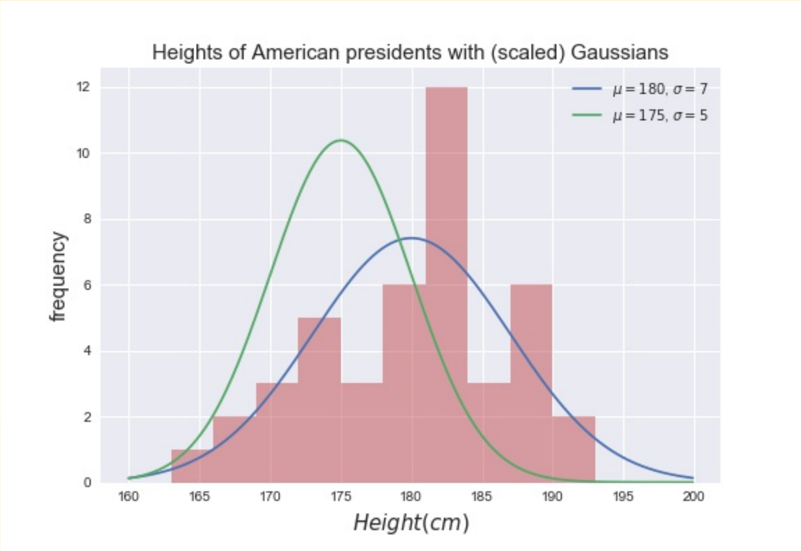
$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$



- The area under the curve gives the probability of the random variable being in an interval

Fitting the Gaussian to data

- There are infinitely many values of mean μ and variance σ^2 to choose from



- Which one fits our data the best ?

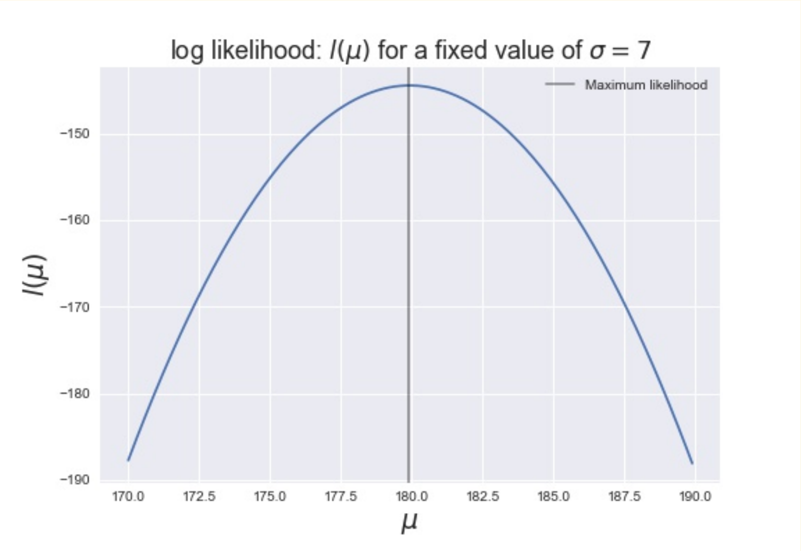
Maximum likelihood

- Consider all possible values of mean and variance
- For each pairs of values, how likely is it that the data has been generated by the Gaussian ?
- Choose the mean and variance that fit the best.

Maximum likelihood

- **independent** random variables:
  - dice roll: independent
  - eye colour in a family: dependent
- Consider data to be *independent and identically distributed* (iid): $X = (X_1, X_2, ..., X_N)$
- multiply probabilities to get the likelihood: $L(\mu, \sigma^2) = \prod_1^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2} = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{\frac{N}{2}} e^{-\frac{1}{2\sigma^2}\sum_1^N (x_i - \mu)^2}$

- We want the maximum, so where the gradient is zero.
- Take the logarithm of the likelihood: easier to differentiate.



$$l(\mu, \sigma) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_1^N (x_i - \mu)^2$$
$$\frac{\partial}{\partial\mu} l(\mu, \sigma) = -\frac{1}{2\sigma^2}\sum_1^N \frac{\partial}{\partial\mu}(x_i - \mu)^2$$

- after differentiation we obtain:
  $\frac{\partial}{\partial\mu}l(\mu,\sigma) = \frac{\sum_1^N x_i}{\sigma^2} - \frac{N\mu}{\sigma^2}$
- We set to zero, and get best fit for the mean μ:
  $\frac{\sum_1^N x_i}{\sigma^2} - \frac{N\hat{\mu}}{\sigma^2} = 0$
  $\hat{\mu} = \frac{\sum_1^N x_i}{N}$
- So the sample mean is value the gives the best fit for a Gaussian distribution!

Recap

- We found data in the real world and wanted to see which parametrisation of a Gaussian fit best
- For the height data, sample mean: $\hat{\mu} = 179.9\text{cm}$
- For the variance σ^2, do the same (try it!):
  - differentiate the log-likelihood and set to zero: $\frac{\partial}{\partial \sigma^2} l(\mu, \sigma)\big|_{\hat{\sigma}^2, \hat{\mu}} = 0.$
  - Obtain the sample variance: $\hat{\sigma}^2 = \frac{1}{N} \sum_1^N (x_i - \hat{\mu})^2 \approx 49$

Possible extensions

We could have a deterministic model for the mean. Then the gaussian represents the error of the model: plot linear regression with a gaussian at each point on the line.