Ordinary Least Squares is an analytical method to find the optimal weight vector $\boldsymbol{w}$ for linear regression. This method is based on the assumption that the response of a linear function is $y(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{\varepsilon}$ where the residual error between model predictions and the true response $\varepsilon_i = y_i - \boldsymbol{w}^T\boldsymbol{x}_i$ is normally distributed with mean μ and variance σ^2.

The density function $p(y|x, \mu, \sigma^2)$ is also normally distributed with mean μ and variance σ^2. μ is a linear function of x: $\mu = \boldsymbol{w}^T\boldsymbol{x}_i$. It is reasonable to assume that the noise component, σ^2, is the same for the entire dataset.

Assuming each training example is independent and identically distributed (*iid*), the likelihood function for the dataset is

$$\mathcal{L}(\mu, \sigma^2) = \prod_{i=1}^{m}\left[\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}}.\,exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right)\right]$$

The optimal $\boldsymbol{w}$ can be determined by applying MLE to RSS, that is, by setting $\frac{\partial RSS}{\partial \boldsymbol{w}} = 0$ and solving for $\boldsymbol{w}$. In matrix notation $\boldsymbol{w}_{OLS} = \left(\mathrm{X}^\mathrm{T}\mathrm{X}\right)^{-1}\mathrm{X}^\mathrm{T}\boldsymbol{y}$. This is also known as the normal equation.

Complete the start-up code to perform the following tasks:

- Apply the normal equation to find $\boldsymbol{w}_{OLS} = [w_0, w_1\,]^T$.
- Print the predicted price of a 5000 square foot house. Remember to "normalize" the square footage first.
- Plot the regression line $\hat{y}_{OLS}$ over the training examples. The line is defined by $\boldsymbol{w}_{OLS}$.

Submit the python code (*.ipynb* file). Your code must run on Google Colab.