## Polynomial regression

10) The 2nd order regression model is given by

To do this, the feature vector is squared and concatenated to the feature matrix . Additionally, a column of 1’s is concatenated to include the intercept term. Later, refers to the column of 1’s, corresponds to in the formulation above and corresponds to .

The model can then be characterized by the following equation (in matrix notation):

The cost function is given by:

In matrix notation:

Note that the term is squared. This is allowed because the result of the product is a scalar.

Exploiting the matrix notation, we can derive the partial derivatives with respect to :

Using the expression above we can obtain the partial derivatives with respect to the components of :

The update rule for gradient descent is given by:

where is the learning rate.

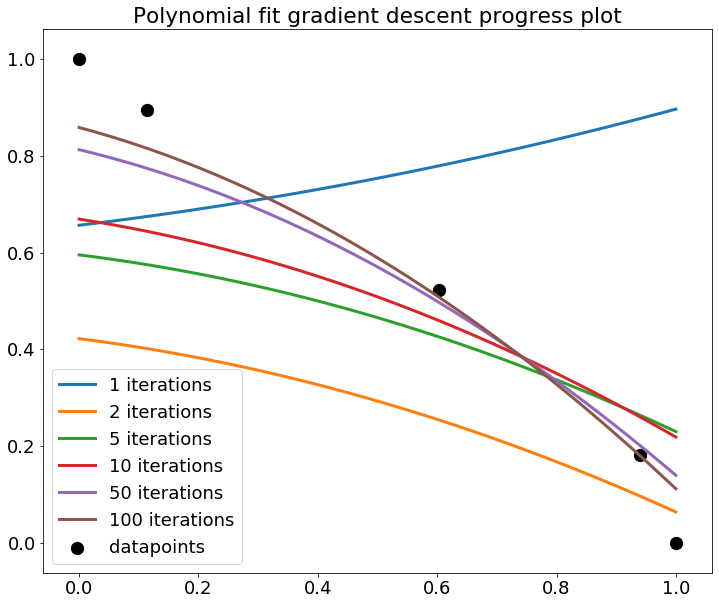
Writing down the components yields:

The gradient descent procedure stops if either of the following two conditions are met:

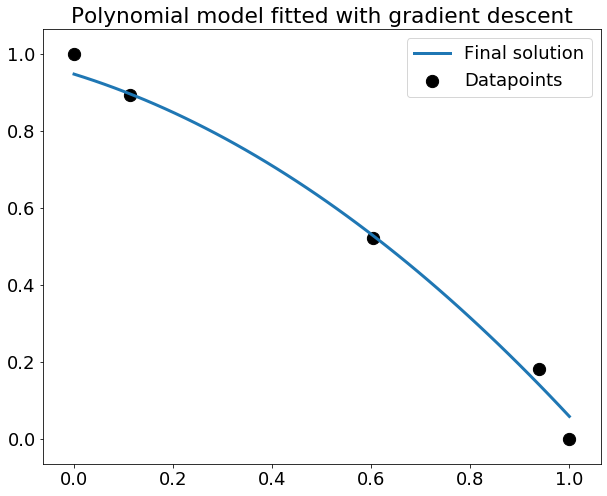
* The maximum amount of iterations has been reached.
* The absolute change in gradient magnitude is less than specified tolerance.

Considering the fact that the dataset contains values of x around 20, that the model consists of a second order polynomial and that the cost function includes a fourth order term, it would be beneficial to use the scaled dataset from subproblem 2. Without scaling the gradients (and the cost function) could take on large values which may lead to numerical stability problems and increased computation time.

The model was fitted whilst the estimated parameters and the cost are saved for every iteration. The initial parameter guess vector was [0, 0, 0], the learning rate was set to 0.6, the maximum number of iterations to 2000 and the gradient tolerance to 10-7. The model fit was plotted at 1, 2, 5, 10, 50 and 100 iterations. These model fits are shown below.



Gradient descent converged after 948 iterations and this resulted in the fit below.



Although a non-linear mapping is fitted by polynomial regression, estimation of the parameters is still a linear problem because the model equation is characterized by a *linear* combination of the input features with the model parameters. Due to the higher order terms, a polynomial regression model is more subject to overfitting to the data and is very outlier sensitive.