COMPENG 4SL4 Assignment 1 Report

Raeed Hassan

hassam41

400188200

October 9, 2022

Linear Regression Modelling with Least Squares

Ten linear regression models of increasing capacity were trained for a training set of the relation

$$t = \underbrace{\sin(4\pi x)}_{f_{true}(x)} + \epsilon \tag{1}$$

where ϵ random noise with a Gaussian distribution with 0 mean and variance 0.09. The training set was constructed with 10 examples $\{(x^{(1)}, t^{(1)}), \ldots (x^{(10)}, t^{(10)})\}$, uniformly spaced in the internal [0,1], and the validation was similar constructed with 100 examples. The targets for each set were generated according to relation (1). The seed for the random numbers for the noise was formed with the last 4 digits of my student ID (8200).

Ten predictors for M from 0 to 9 were found using least squares, with the predictor coefficients founding using the relation

$$w = (X^T X)^{-1} X^T t \tag{2}$$

where X is the Vandermonde matrix of x. The predictor functions derived through the least squares method are shown in Table 1. The plot of each predictor function is shown in Figures 1–10. We can observe that as M increases, the predictor $f_M(x)$ becomes better at fitting the training points in the training set, and begins to resemble a sin function, but at $M = 9$ we can see that the predictor begins to overfit to the training data.

The training and validation errors for each predictor is shown in Table 2. As expected, as we increase M the training error decreases, however we can see that $M = 9$ the validation error begins to increase despite having the lowest training error, due to the overfitting discussed earlier. If we were to select a predictor function based on this data, we would select $M = 7$ as our predictor function as this predictor produces the lowest validation error when compared to our validation set.

M	Predictor Function
0	-0.01033359
1	0.43409653 - 0.88886025x
2	0.45604542 - 1.03701525x + 0.148155x^2
3	0.48031649 - 1.43662107x + 1.20134604x^2 - 0.70212736x^3
4	0.22706564 + 8.06028578x - 46.75803356x^2 + 76.22281814x^3 - 38.46247275x^4
5	1.52971950e-1 + 4.26817105e1x - 3.35349081e2x^2 + 8.91973513e2x^3 - 9.73497467e2x^4 + 3.74013998e2x^5
6	-1.49350573e-1 + 4.17153461e1x - 3.23426783e2x^2 + 8.40658825e2x^3 - 8.74333070e2x^4 + 2.85805534e2x^5 + 2.94028210e1x^6
7	-1.08187756e-1 - 2.79901302e0x + 4.11094984e2x^2 - 3.51513917e3x^3 + 1.14419516e4x^4 - 1.76749271e4x^5 + 1.30636216e4x^6 - 3.72406252e3x^7
8	-1.11244799e-1 + 1.90768110e1x - 3.00623774e1x^2 - 1.87674593e2x^3 - 1.11072465e3x^4 + 8.45853799e3x^5 - 1.74254206e4x^6 + 1.49426978e4x^7 - 4.66669003e3x^8
9	-1.1254719035468952e-1 + 1.3331309735774994e2x - 2.6686857490539551e3x^2 + 2.3579426277160645e4x^3 - 1.1276578366088867e5x^4 + 3.1264942968750000e5x^5 - 5.1552117065429688e5x^6 + 4.9806604736328125e5x^7 - 2.6027476269531250e5x^8 + 5.6801928695678711e4x^9

Table 1: Predictor Functions for $M = 0$–9

M	Training Error	Validation Error
0	0.5201122363344061	0.6110146395929855
1	0.43964188447651165	0.535424084259008
2	0.4394652414421542	0.536733362125242
3	0.4391787138011748	0.5319587930886094
4	0.3825648436890513	0.595438430761644
5	0.06963627001271834	0.2660364124955042
6	0.06954009795213771	0.26695432746209075
7	0.008517383401040548	0.16060941834056727
8	0.005845212432320224	0.17961847258518052
9	5.197575747112509e-8	0.4183415258492314

Table 2: Training and Validation Errors for $M = 0$–9

Linear Regression Modelling with Regularization

To control the overfitting seen in $M = 9$, we will train the model with regularization. This allows to limit overfitting in the model, but we must also make sure we select values for λ to produce the best results and also to avoid underfitting. The model was trained with several values of λ to determine the best fit, ranging from e^{-50} to 1. The coefficients for the regularized predictors was calculated with this relation:

$$w = (X^TX + \frac{N}{2}B)^{-1}X^Tt \tag{3}$$

The training and validation errors for these values of λ are shown in Table 3. To eliminate overfitting, we select the value of λ that produces the lowest validation error, λ_1, which is found to be e^{-25} where $\ln \lambda_1 = -25$. A plot of the predictor function for λ_1 can be found in Figure 11. We also select $\ln \lambda_2 = 0$ or $\lambda_2 = 1$ to demonstrate that values of λ that are too high can lead to underfitting.

$\ln \lambda$	Training Error	Validation Error
0	0.5630718561051139	0.6513961458453568
-5	0.4866095450868606	0.57564785030175
-10	0.8730640240006535	1.0322392754481602
-15	0.239989434383105	0.4122650499965516
-20	0.04611112257732629	0.21396326035682034
-25	0.00681937450789687	0.17768961157179558
-30	0.017055602801581055	0.23379480397505514
-40	3.008108918763905e-6	0.41834448198239266
-50	5.210996570901634e-8	0.41834152598343965

Table 3: Training and Validation Errors for Various λ

Training and Validation Errors

The average error between the validation set and $f_{true}(x)$ was calculated and determined to be 0.0.23043795821629443. A plot of the training and validation errors versus M is shown in Figure 13. For $M = 9$, the errors for the model with no regularization and for regularization with λ_1 and λ_2 are shown. The average error between the validation set and $f_{true}(x)$ is also plotted on this graph.
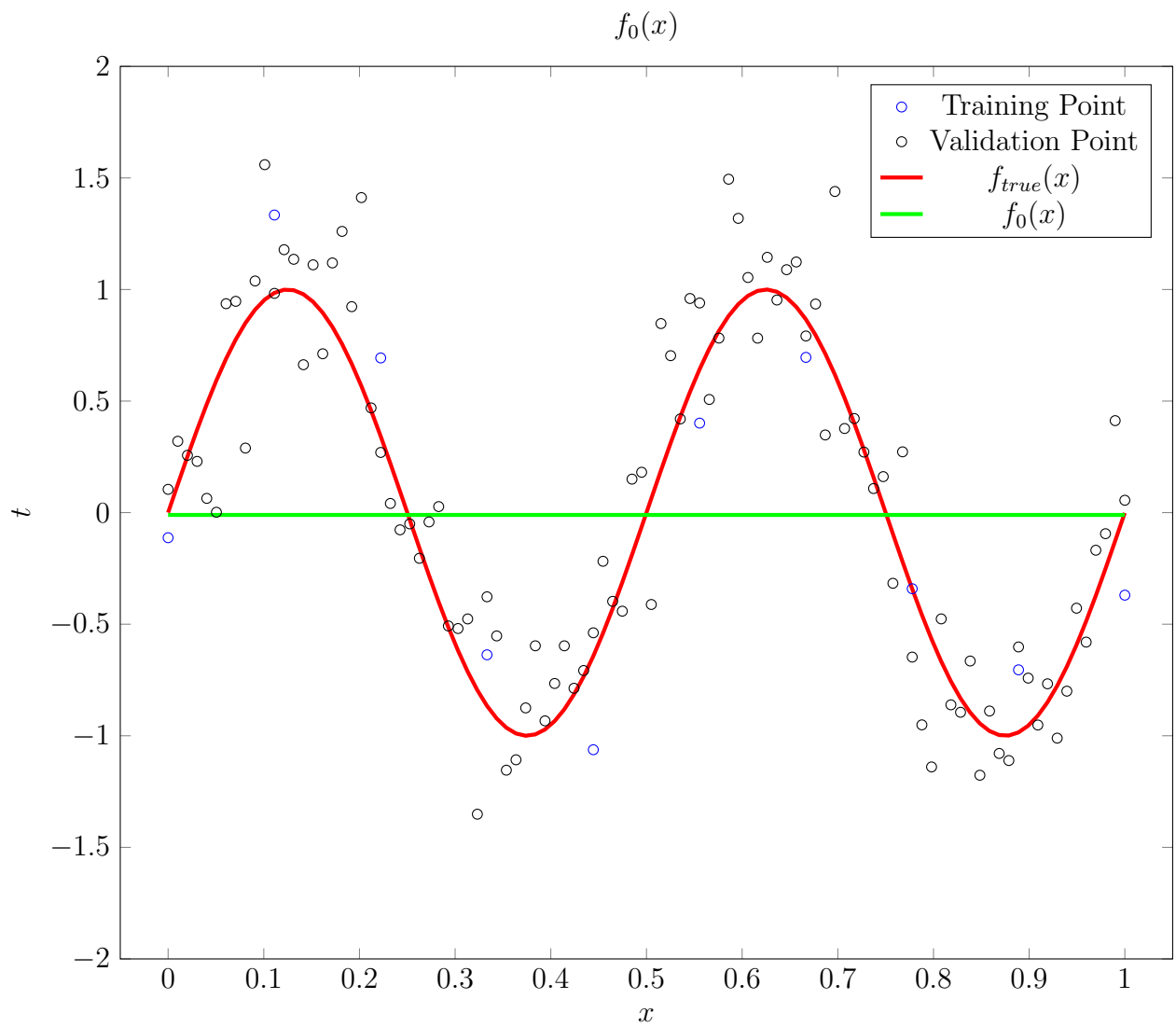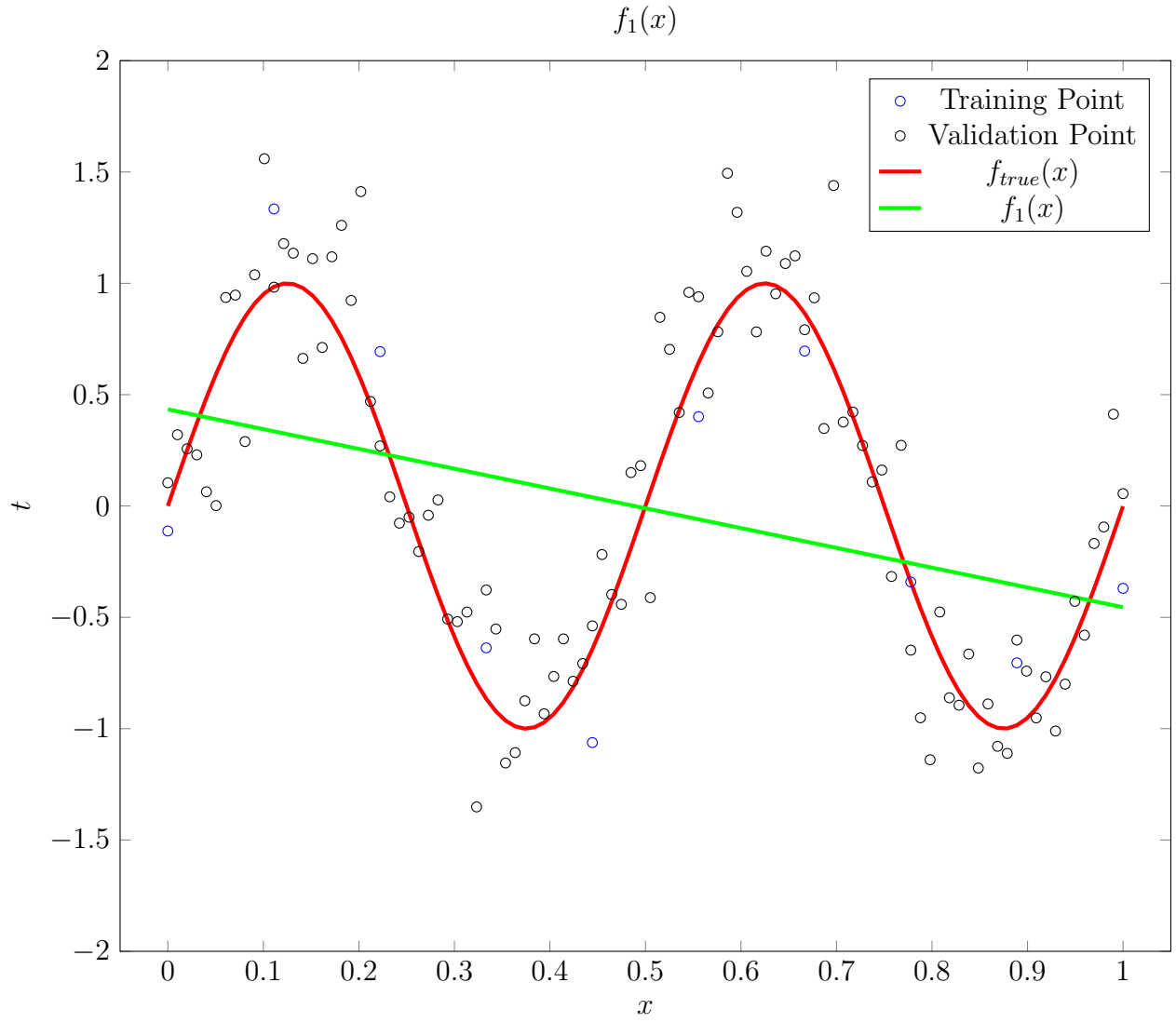
Figure 1: $f_M(x)$ for $M = 0$
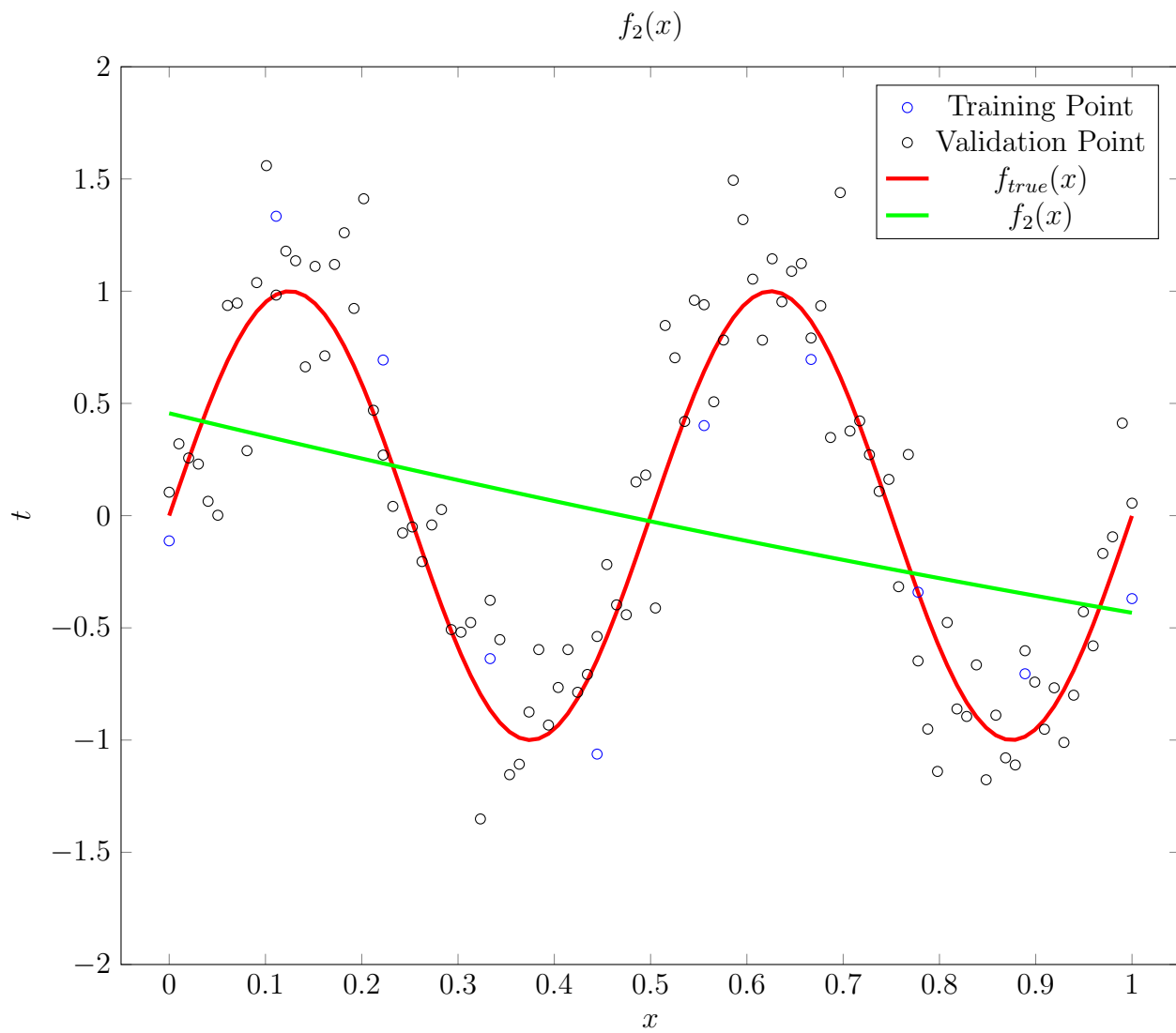
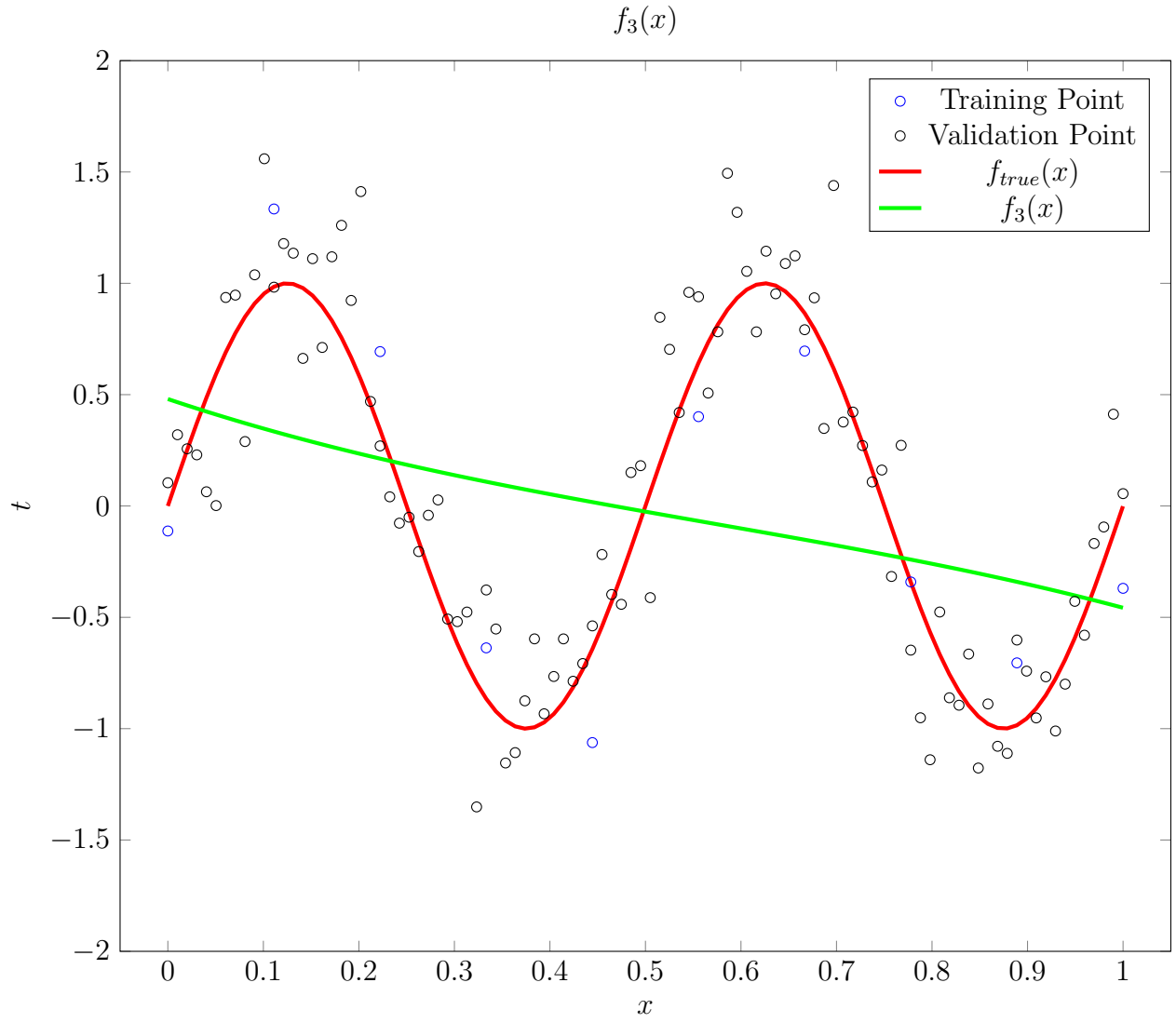Figure 2: $f_M(x)$ for $M = 1$

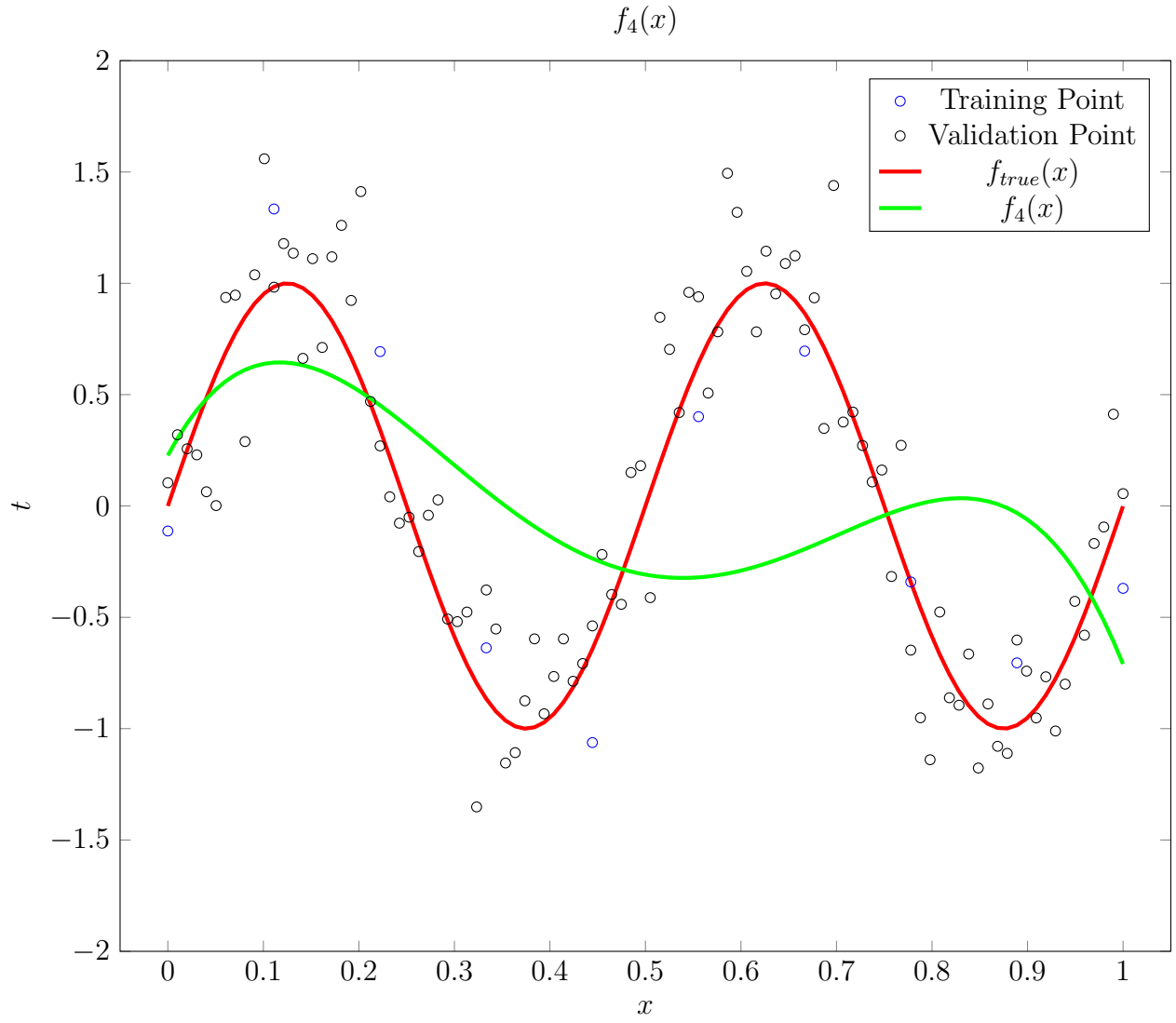Figure 3: $f_M(x)$ for $M = 2$

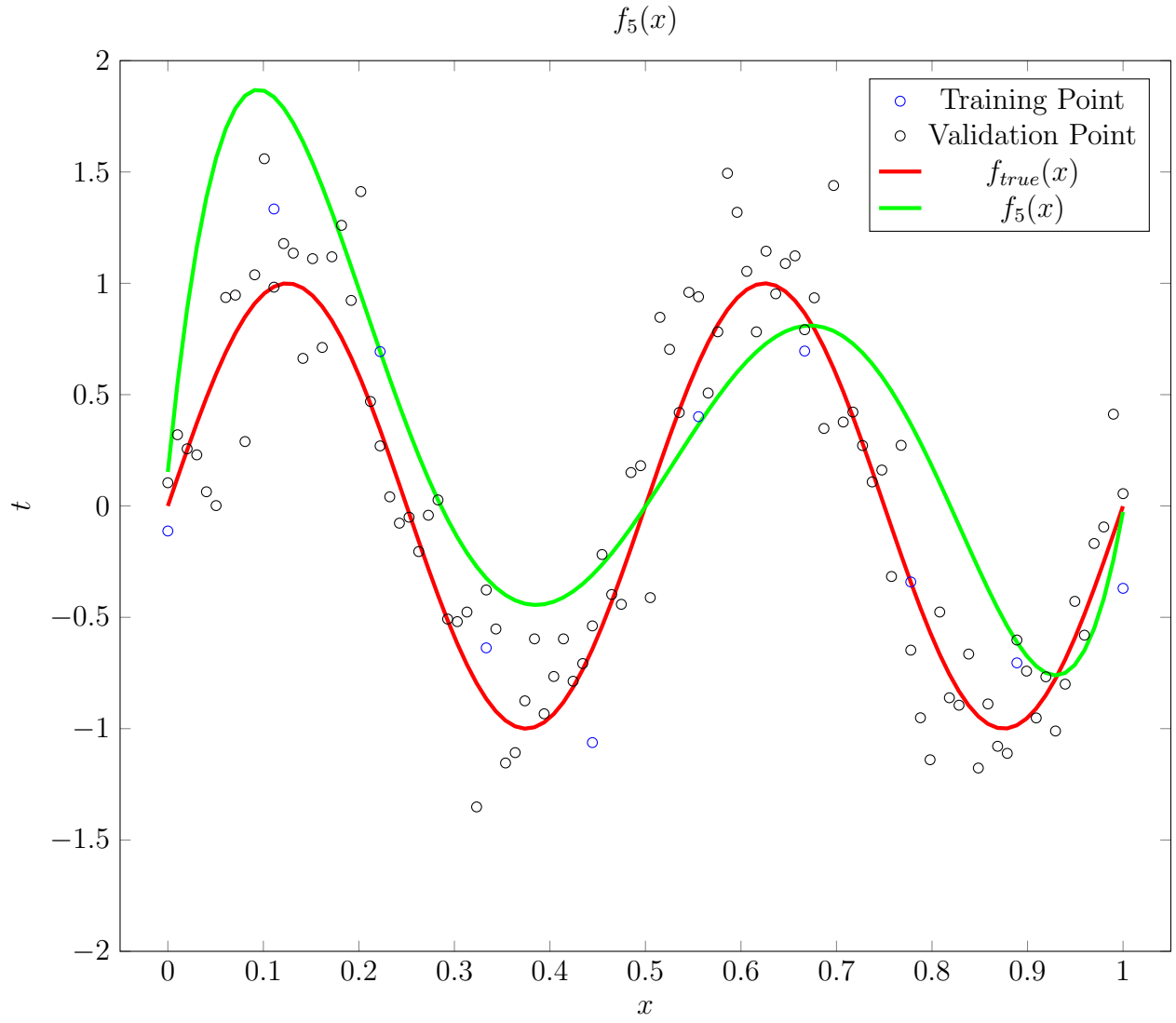Figure 4: $f_M(x)$ for $M = 3$

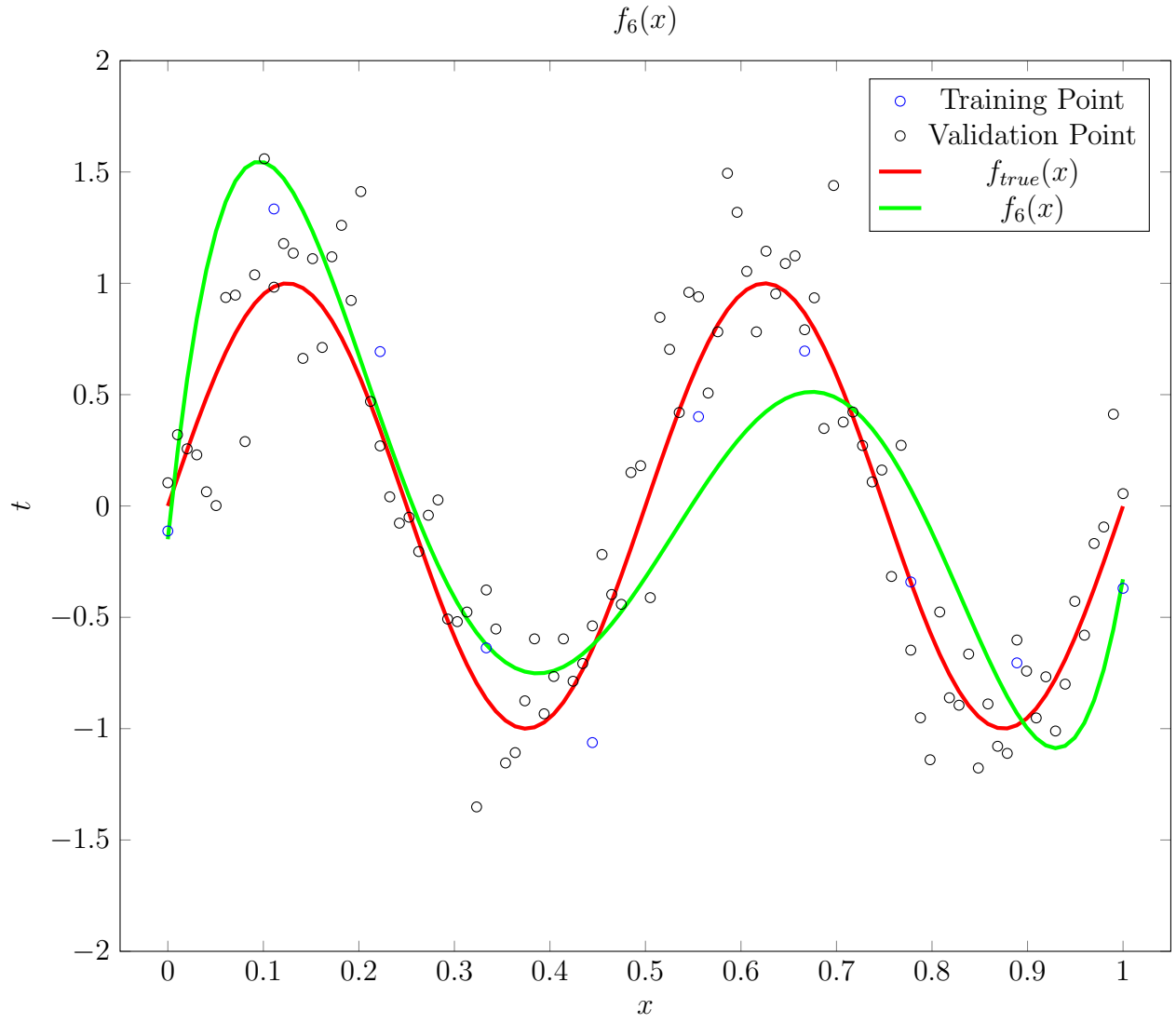Figure 5: $f_M(x)$ for $M = 4$

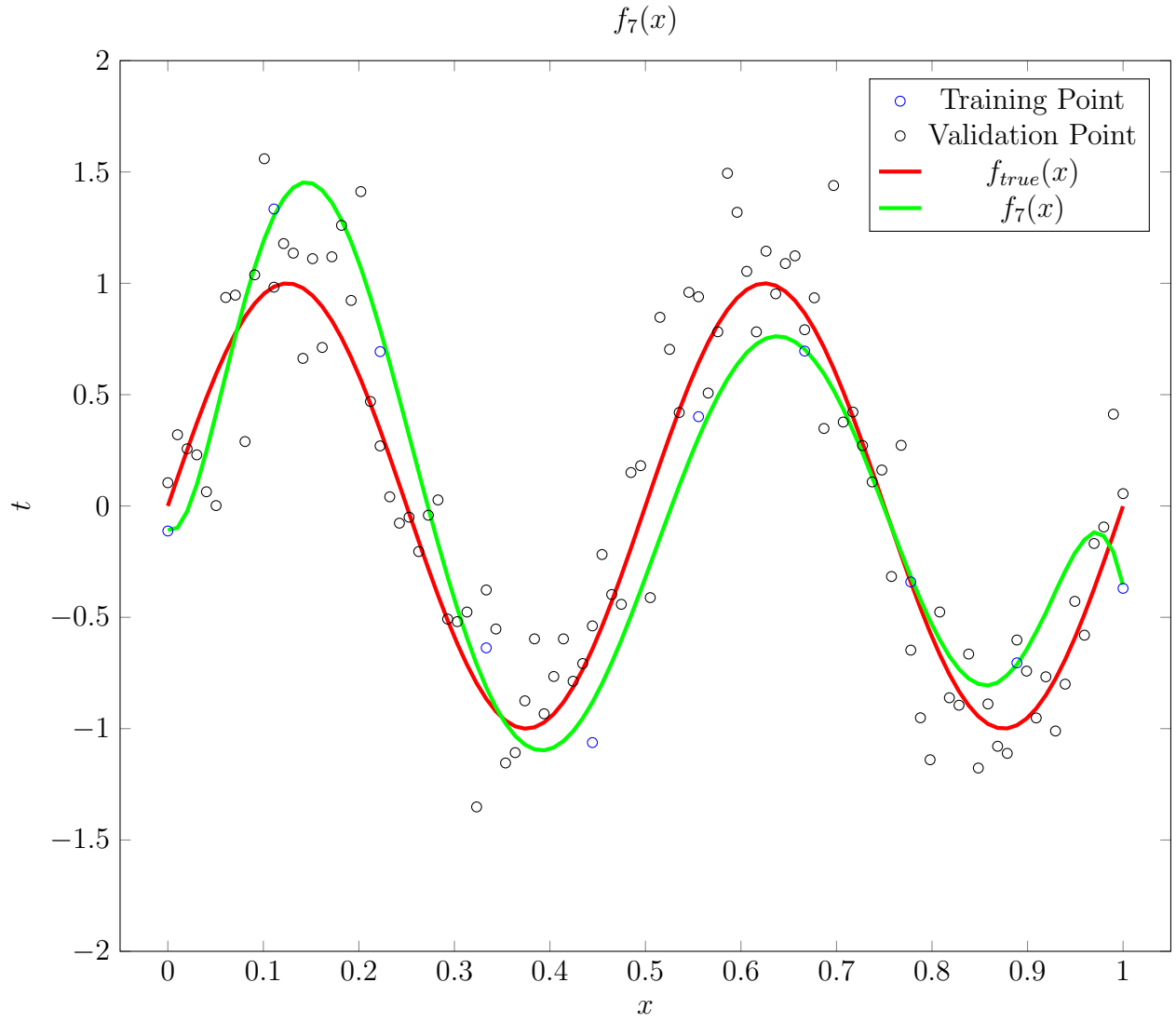Figure 6: $f_M(x)$ for $M = 5$

Figure 7: $f_M(x)$ for $M = 6$

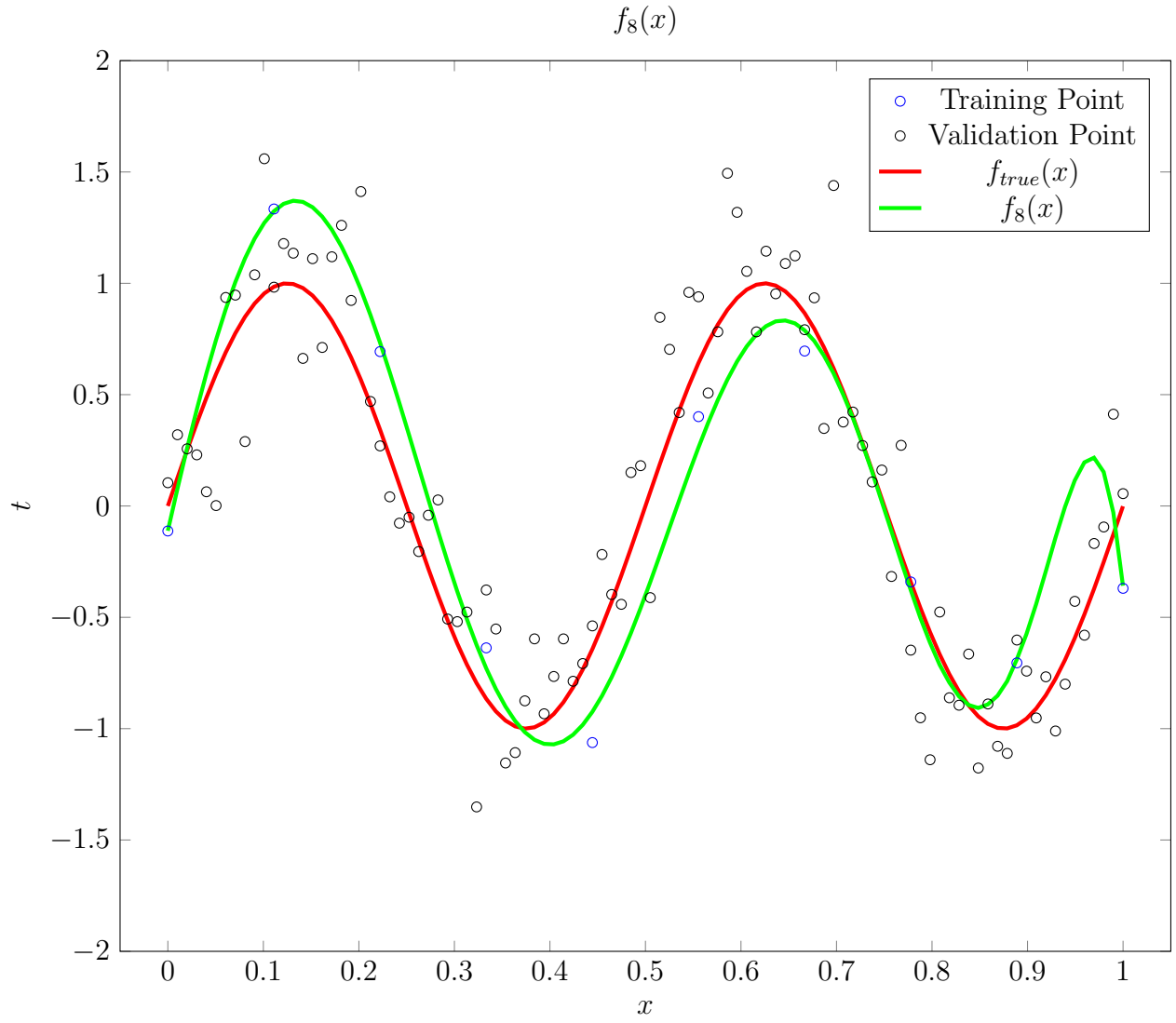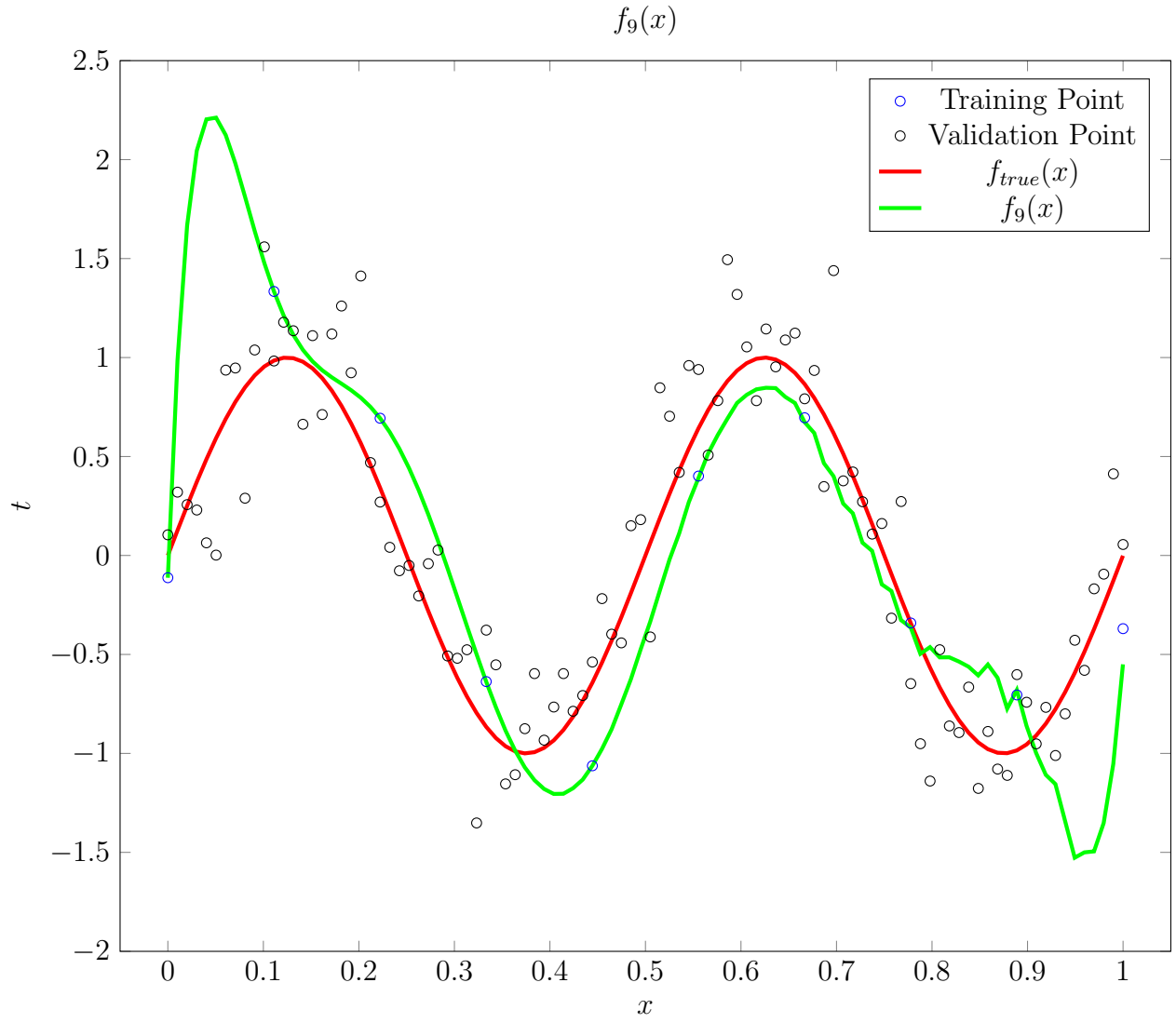Figure 8: $f_M(x)$ for $M = 7$

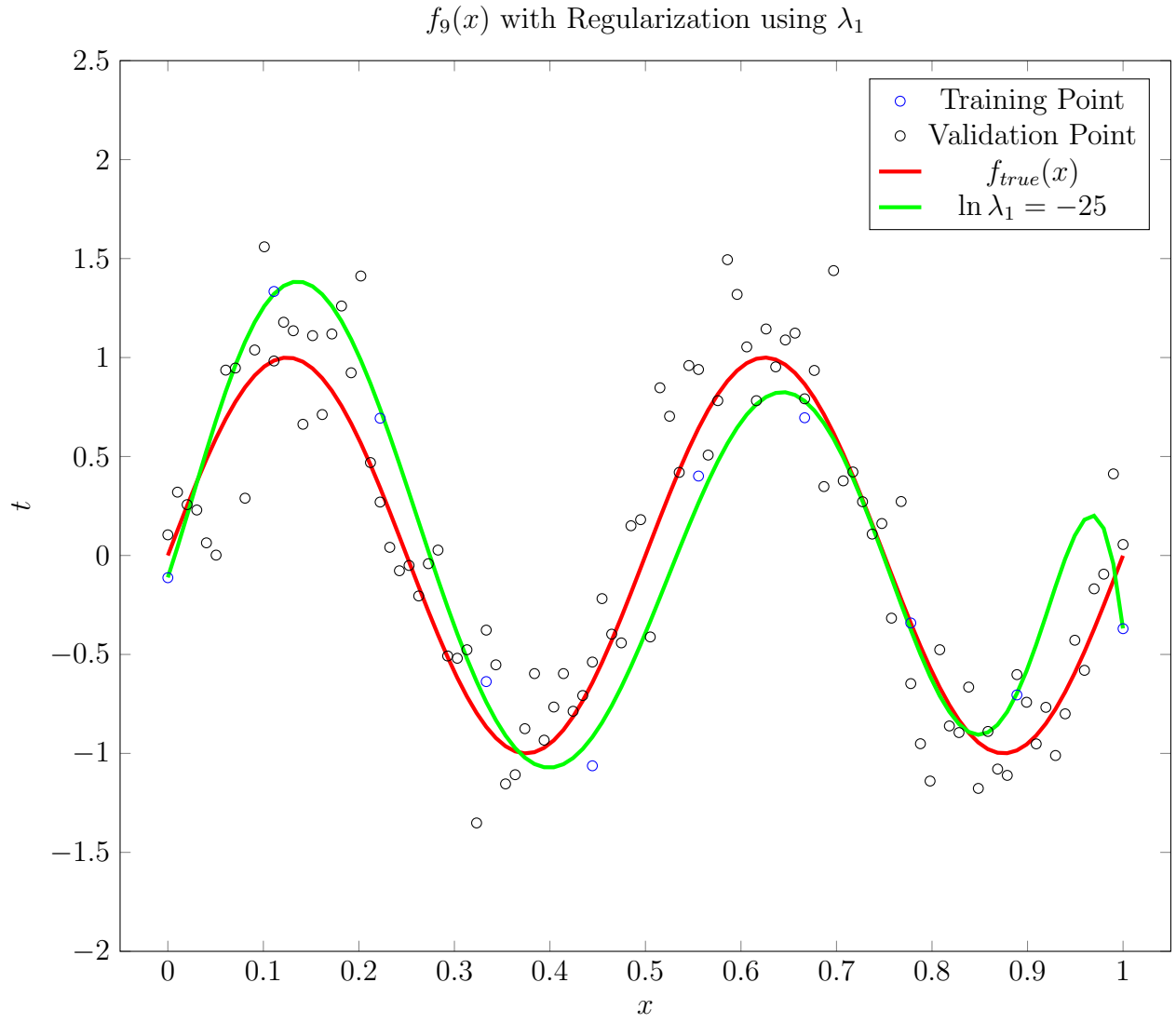Figure 9: $f_M(x)$ for $M = 8$

Figure 10: $f_M(x)$ for $M = 9$

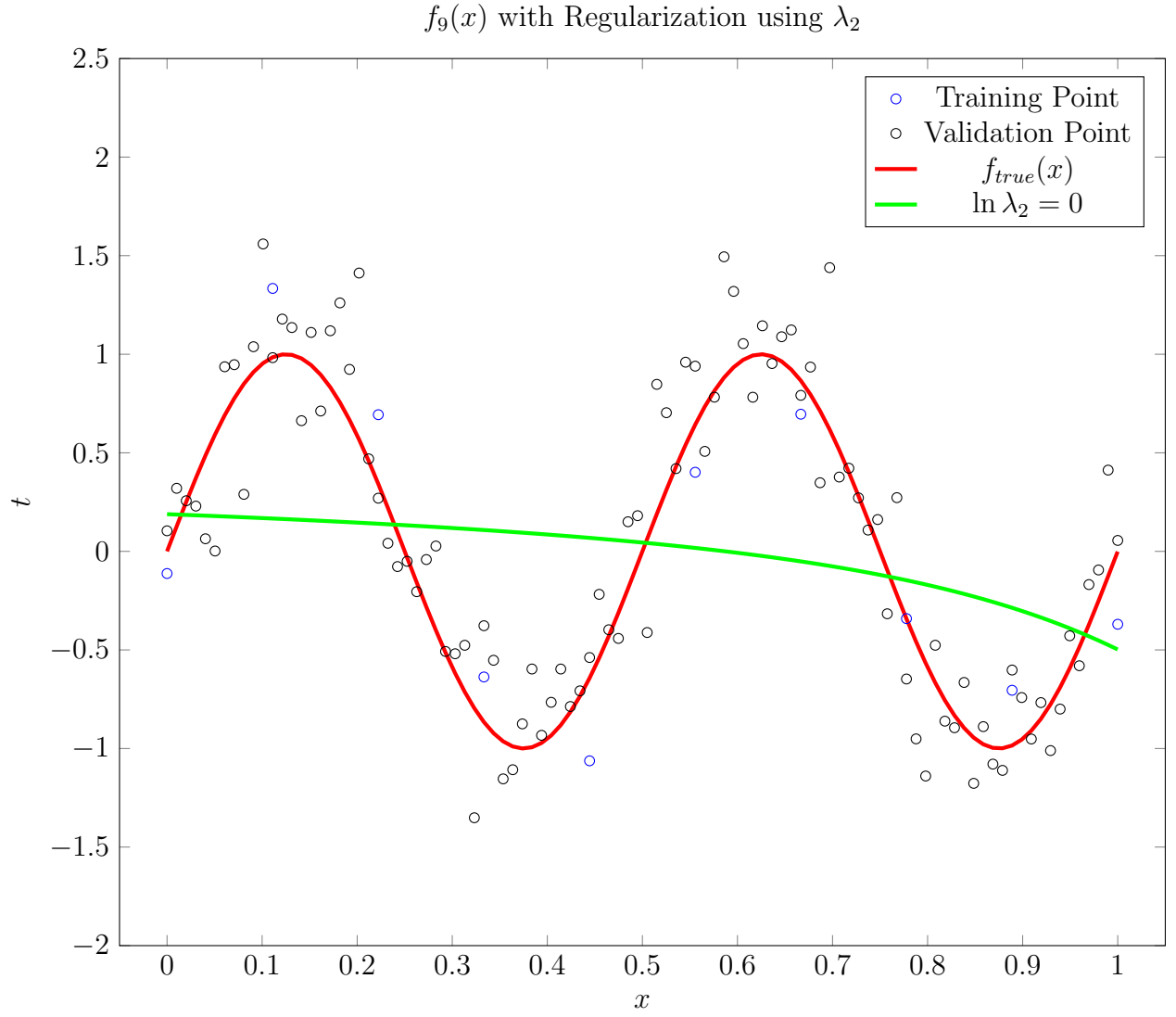Figure 11: $f_M(x)$ for $M = 9$ and $\ln \lambda_1 = -25$
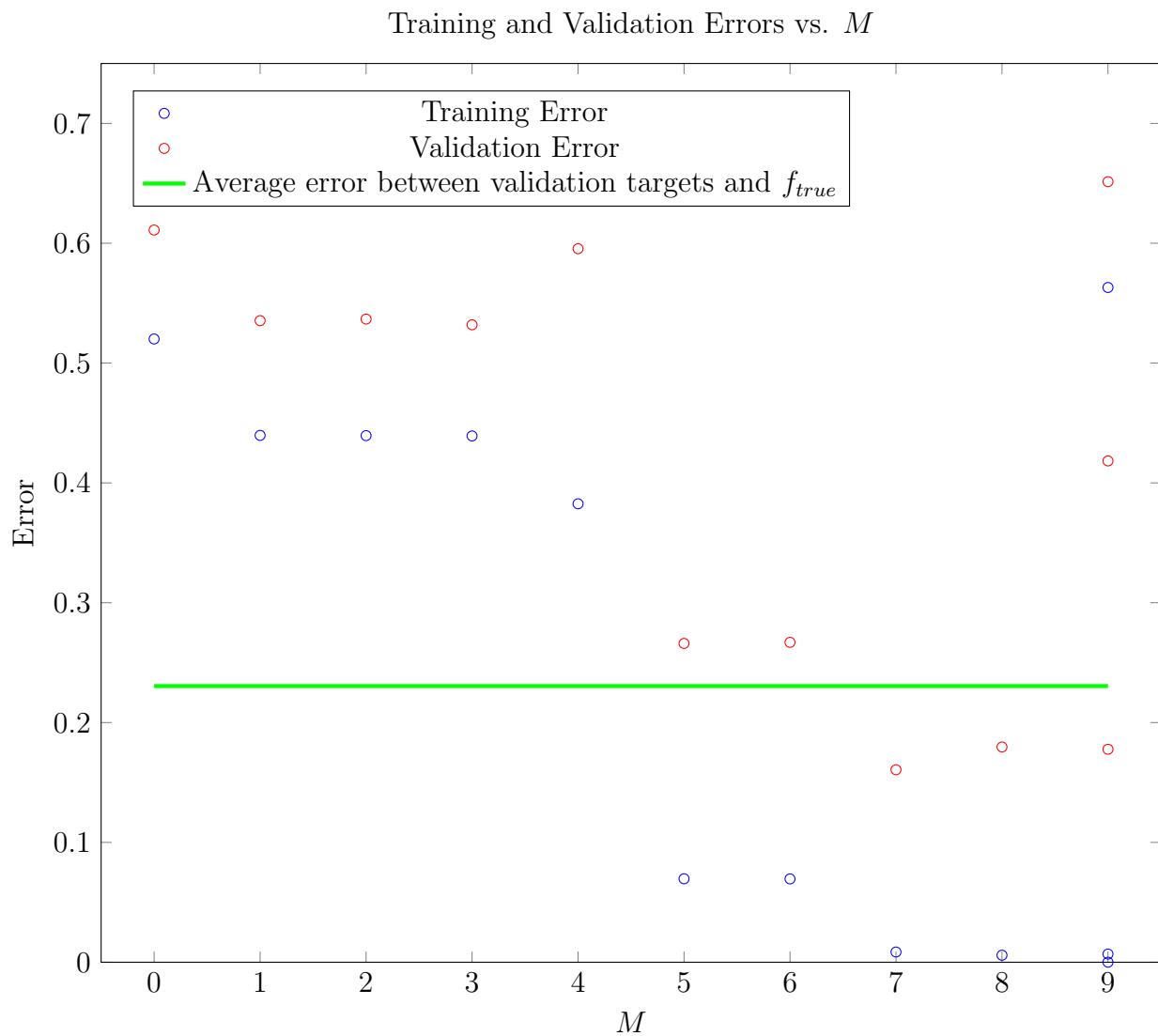
Figure 12: $f_M(x)$ for $M = 9$ and $\ln \lambda_2 = 0$

Figure 13: Training and Validation Errors vs. M