Lab Assignment 1

Steven Truong

Task 1

First note that

$$\mathbf{z}_n^\top \mathbf{w} = \sum_{i=0}^{M-1} w_i x_n^i$$

which is linear in each w_i. Then

$$\begin{aligned}
\frac{\partial J(\mathbf{w})}{\partial w_i} &= \frac{\partial}{\partial w_i}\left(\frac{1}{2}\sum_{n=1}^{N}(\mathbf{z}_n^\top \mathbf{w} - t_n)^2\right) \\
&= \sum_{n=1}^{N}(\mathbf{z}_n^\top \mathbf{w} - t_n) x_n^i.
\end{aligned}$$

This gives us the following gradient:

$$\nabla J(\mathbf{w}) = \sum_{n=1}^{N}(\mathbf{z}_n^\top \mathbf{w} - t_n)\mathbf{z}_n$$

Then

$$\begin{aligned}
\nabla J(\mathbf{w}) = \mathbf{0} &\iff \sum_{n=1}^{N}(\mathbf{z}_n^\top \mathbf{w} - t_n)\mathbf{z}_n = \mathbf{0} \\
&\iff \sum_{n=1}^{N}\mathbf{z}_n \mathbf{z}_n^\top \mathbf{w} = \sum_{n=1}^{N} t_n \mathbf{z}_n \\
&\iff A\mathbf{w} = \mathbf{b},
\end{aligned}$$

where

$$A = \sum_{n=1}^{N}\mathbf{z}_n \mathbf{z}_n^\top \qquad \text{and} \qquad \mathbf{b} = \sum_{n=1}^{N} t_n \mathbf{z}_n.$$

Task 2

Note that

$$\begin{pmatrix} \mathbf{z}_1^\top \\ \mathbf{z}_2^\top \\ \vdots \\ \mathbf{z}_N^\top \end{pmatrix}$$

is the Vandermonde matrix, which we'll call V. Also, it is known that V is always full rank as long as all the x_n are different. This means that the $\mathbf{z}_n$ are linearly independent if $N < M$.

There will always be a solution to the system $A\mathbf{w} = \mathbf{b}$, and uniqueness depends on M and N. There are two cases to consider.

Case 1: $N < M$
In this case, solutions are not unique. For $N < M$, the $\mathbf{z}_n$ are linearly independent, so we have

$$J(\mathbf{w}) = \mathbf{0} \iff \mathbf{z}_n^\top \mathbf{w} - t_n = 0 \ \forall n$$

which, written in matrix form, is $V\mathbf{w} = \mathbf{t}$, where $\mathbf{t} = \begin{pmatrix} t_1 & t_2 & \cdots & t_N \end{pmatrix}^\top$, and V is full row rank. V is thus onto, but not one-to-one, which means that we have non-unique solutions for all $\mathbf{t}$.

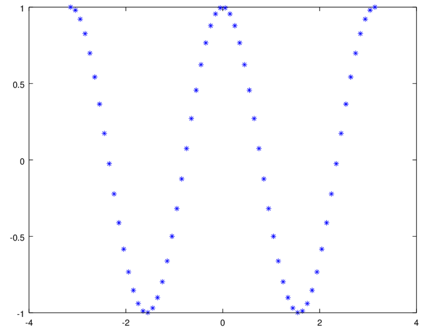
Case 2: $N \geq M$
We'll show that A is positive definite, so it's bijective. In other words, for every $\mathbf{b}$, there exists a unique $\mathbf{w}$ satisfying $A\mathbf{w} = \mathbf{b}$.

A is positive semidefinite since it is the sum of positive semidefinite matrices. Suppose there exists some $\mathbf{z} \neq \mathbf{0}$ such that $\mathbf{z}^\top A\mathbf{z} = 0$. Then since each $\mathbf{z}_n\mathbf{z}_n^\top$ is positive semidefinite,

$$\mathbf{z}^\top A\mathbf{z} = \sum_{n=1}^{N} \mathbf{z}^\top \left(\mathbf{z}_n\mathbf{z}_n^\top\right)\mathbf{z} = 0 \iff \mathbf{z}^\top \left(\mathbf{z}_n\mathbf{z}_n^\top\right)\mathbf{z} = 0 \ \forall n.$$

Since $\mathbf{z}$ is not the zero vector, $\mathbf{z}$ must be orthogonal to $\mathbf{z}_n$ for all n. But since $N \geq M$ and V has full rank, $\{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ spans $\mathbb{R}^M$. This implies that $\mathbf{z}$ must be $\mathbf{0}$ since it will be orthogonal to every vector in $\mathbb{R}^M$ since any vector can be written as a linear combination of $\{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$, which is a contradiction. Thus, A is positive definite, so $A\mathbf{w} = \mathbf{b}$ has a unique solution for any $\mathbf{b}$.

Task 4



Task 6

$M = 4$ gives a poor approximation of our data since it's unable to capture the oscillation of f at all.

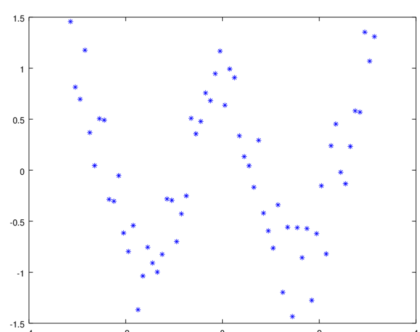
$M = 8$ gives a decent approximation of our data. It's able to capture the oscillation of f on the interval of interest, and it matches the overall shape of our sample. However, it's not able to capture the peaks very well, especially at the origin.

$M = 16$ gives, by far, the best approximation of the data. It pretty much fits the data set exactly.

$M = 32$ is also a good approximation of the data, but not as well as $M = 16$. It fits the data extremely well except at the ends of the interval, where it begins to explode because of its large degree.

$M = 64$ is mostly the same as $M = 32$, but it explodes much faster at the end points than $M = 32$, and in the opposite direction.

Task 7



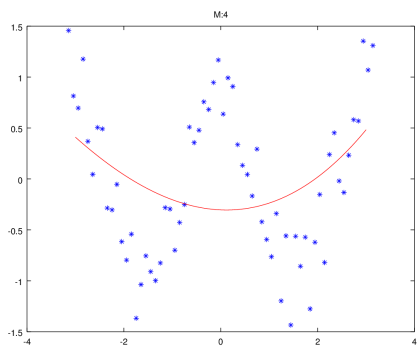
Task 8

$M = 4$ doesn't give a very good approximation of the data since it's unable to capture the oscillation of f.

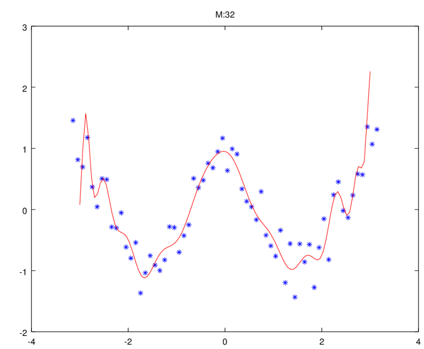
$M = 8$ and $M = 16$ are the best approximations for the data sample from our tested polynomials. However, $M = 16$ seems to be the best overall since their key difference is that $M = 16$ captures the peak at the origin better than $M = 8$.

When M is close to N, we get graphs that oscillate and explode close to the ends of our sample. This is true for $M = 32$ and $M = 64$, which both oscillate and have sharp jumps in the graph. For example, close to $x = -3$, $M = 64$ has a huge jump right before it explodes.

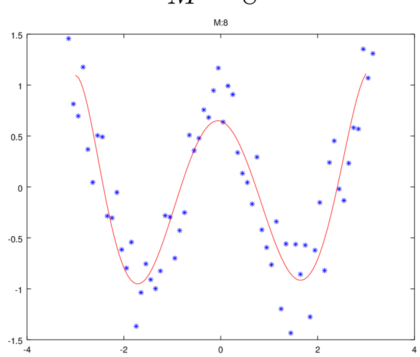
$M = 4$



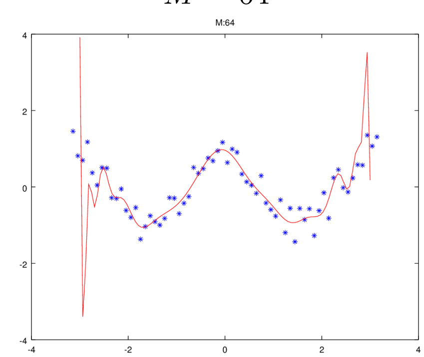
$M = 32$



$M = 8$



$M = 64$



$M = 16$