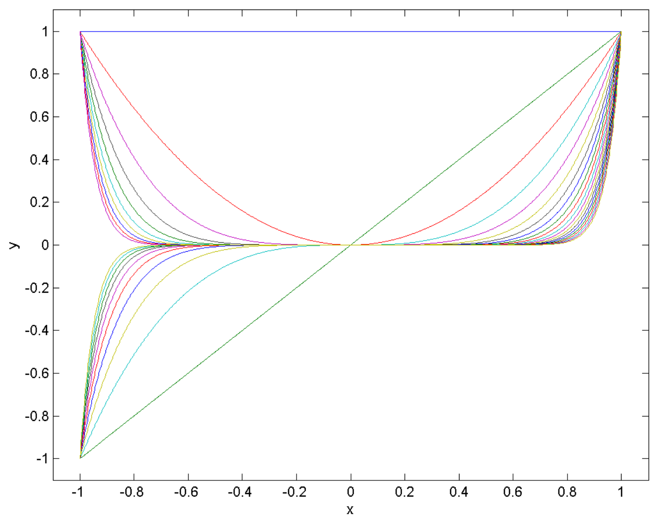
Q2

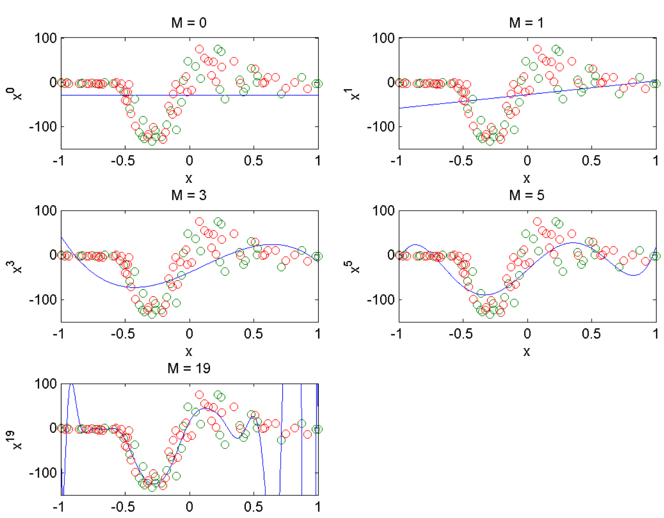
2.1

The function is in getPowers.m, the plots are generated by q2_1.m

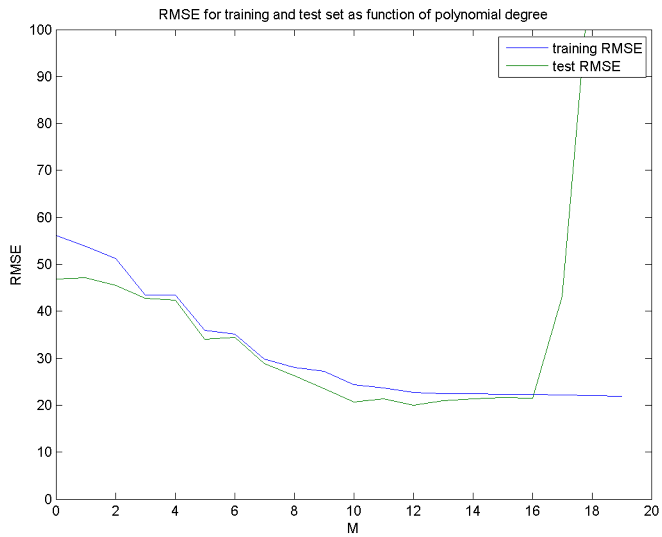


2.2

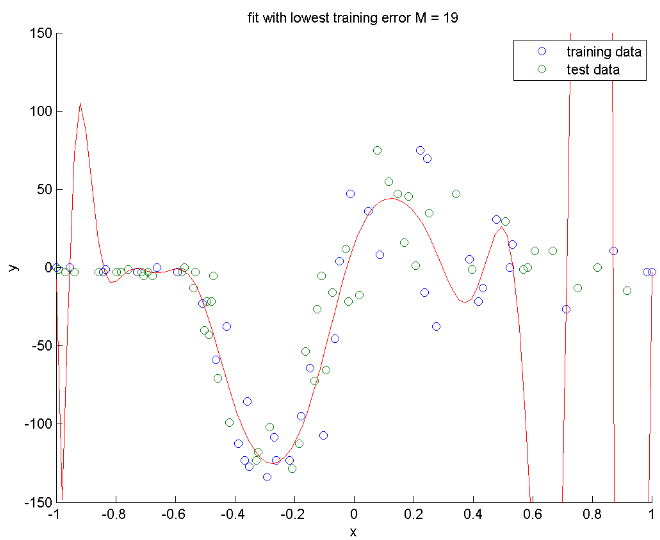
The figures are generated by q2_2.m



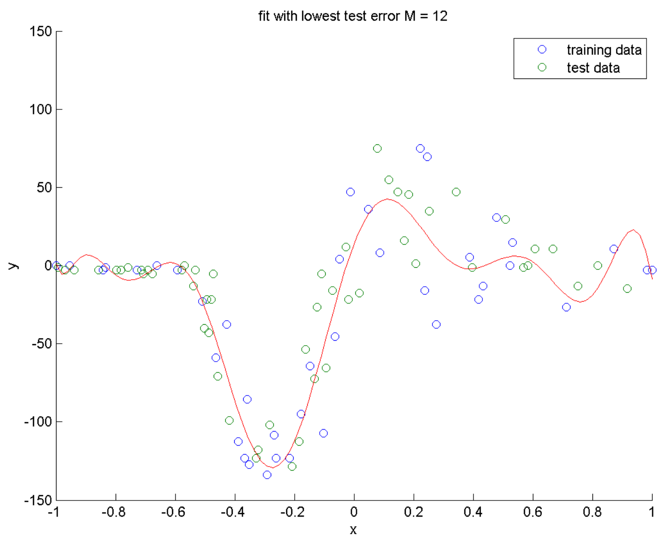
2.3



RMSE for training and test set as function of polynomial degree
training RMSE
test RMSE
RMSE
M



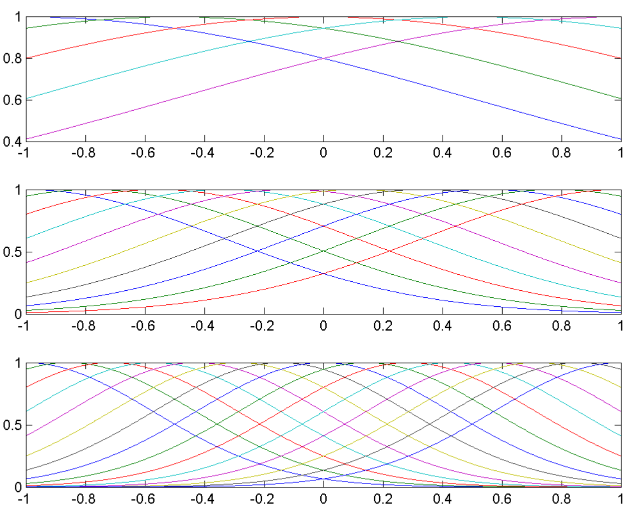
fit with lowest training error M = 19
training data
test data
y
x



The RMSE is the square root of the ML estimate of σ^2.
Minimum training error for M = 19
Minimum test error for M = 12
The figures are generated by q2_3.m

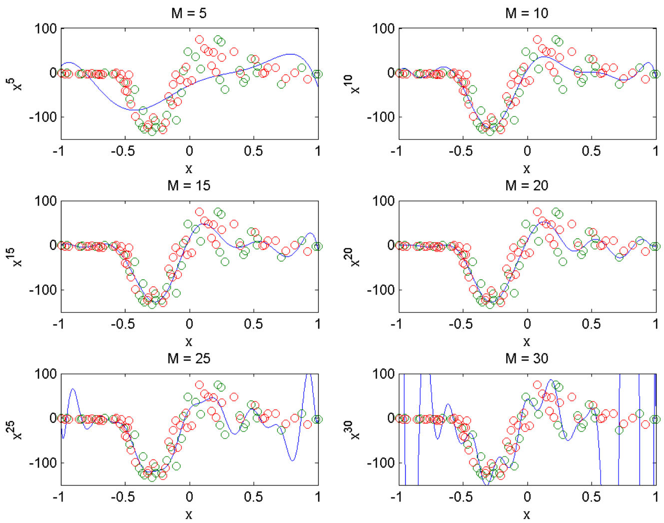
2.4

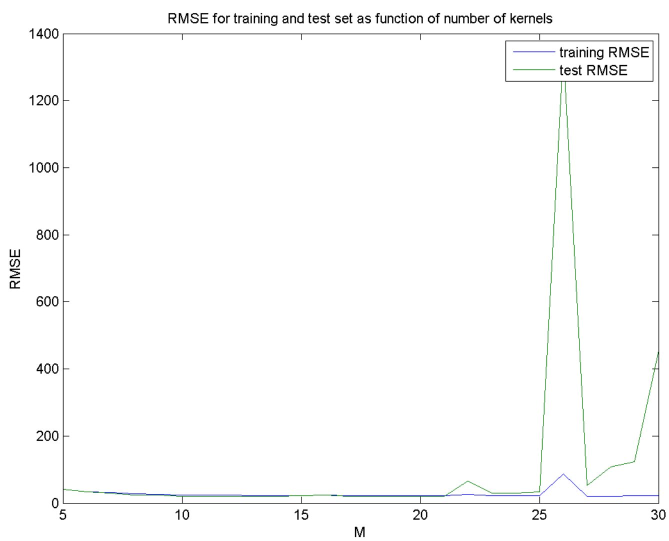
The figures are generated by q2_4.m



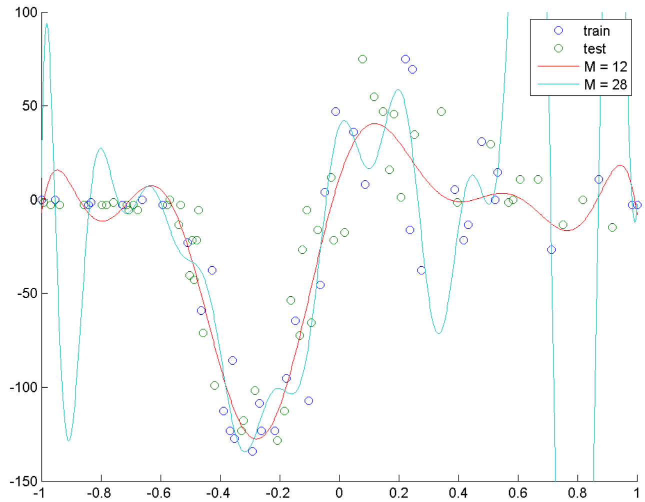
2.5

The figures are generated by q2_5.m





The lowest test error is for M = 8 The lowest train error is for M = 24



2.6

We have the following distributions:

$$p(w_k) = \frac{1}{\sqrt{2\pi\alpha^{-1}}} \exp(\frac{1}{2\alpha^{-1}}(w_k)^2)$$

$$p(y|x, w) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{1}{2\sigma^2}(xw - y)^2)$$

And we get

$$p(w) = \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi\alpha^{-1}}} \exp(\frac{1}{2\alpha^{-1}}(w_k)^2)$$

$$p(Y|X, w) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{1}{2\sigma^2}(x_i w - y_i)^2)$$

and

$$\begin{aligned}
\ln(p(w)) &= d\ln(\frac{1}{\sqrt{2\pi\alpha^{-1}}}) - \frac{1}{2\alpha^{-1}}\sum_{k=1}^{d} w_k^2 \\
&= d\ln(\frac{1}{\sqrt{2\pi\alpha^{-1}}}) - \frac{1}{2\alpha^{-1}}||w||^2 \\
&= C_1 - \frac{1}{2\alpha^{-1}} w^T w
\end{aligned}$$

$$\begin{aligned}
\ln(p(Y|X,w)) &= n\ln(\frac{1}{\sqrt{2\pi\sigma^2}}) - \frac{1}{2\sigma^2}\sum_{k=1}^{d}(x_i w - y_i)^2 \\
&= n\ln(\frac{1}{\sqrt{2\pi\sigma^2}}) - \frac{1}{2\sigma^2}||Xw - Y||^2 \\
&= C_2 - \frac{1}{2\sigma^2}(XW - Y)^T(XW - Y)
\end{aligned}$$

$$\begin{aligned}
p(w|X,Y) &= \frac{p(w,X,Y)}{p(X,Y)} \\
&\propto p(w,X,Y) \\
&= p(X,Y|w)p(w) \\
&= p(w)p(Y|w,X)p(X|w) \\
&= p(w)p(Y|w,X)p(X) \\
&\propto p(w)p(Y|w,X)
\end{aligned}$$

We wish to solve the following optimization problem:

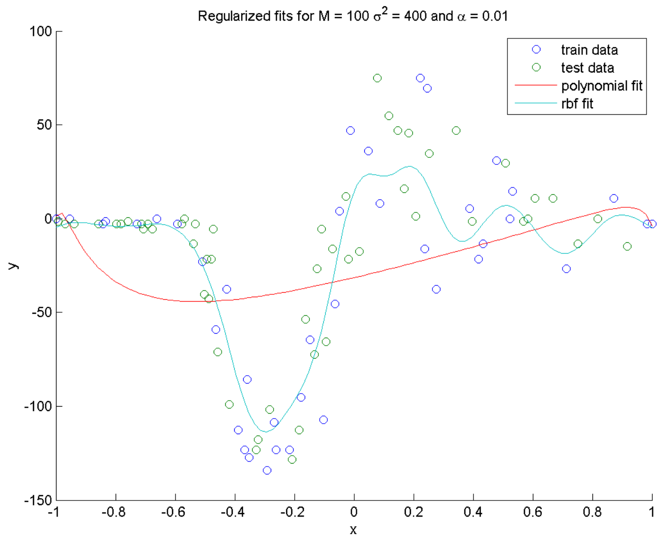
$$\begin{aligned}
& \max_w p(w|X,Y) \\
\iff & \max_w \quad -\ln(p(w|X,Y)) \\
\iff & \max_w \quad -\ln(p(w) - \ln(p(Y|X,Y)) \\
\iff & \max_w \quad \frac{1}{2\alpha^{-1}}w^T w + \frac{1}{2\sigma^2}(Xw - Y)^T(Xw - Y) - C_1 - C_2 \\
\iff & \max_w \quad \frac{\alpha}{2}w^T w + \frac{1}{2\sigma^2}(Xw - Y)^T(Xw - Y) \\
\iff & \max_w \quad \alpha\sigma^2 w^T w + (Xw - Y)^T(Xw - Y) \\
\iff & \max_w \quad \alpha\sigma^2 w^T w + w^t X^T X w - 2wX^T Y + Y^T Y \\
\iff & \max_w \quad w^T(\alpha\sigma^2 I + X^T X)w - 2wX^T Y
\end{aligned}$$

Now, we solve for critical points

$$\nabla_w(w^T(\alpha\sigma^2 I + X^T X)w - 2wX^T Y) = 0 \iff 2(\alpha\sigma^2 I + X^T X)w - X^t Y = 0$$

$$w = (\alpha\sigma^2 I + X^T X)^{-1}(X^T)Y$$

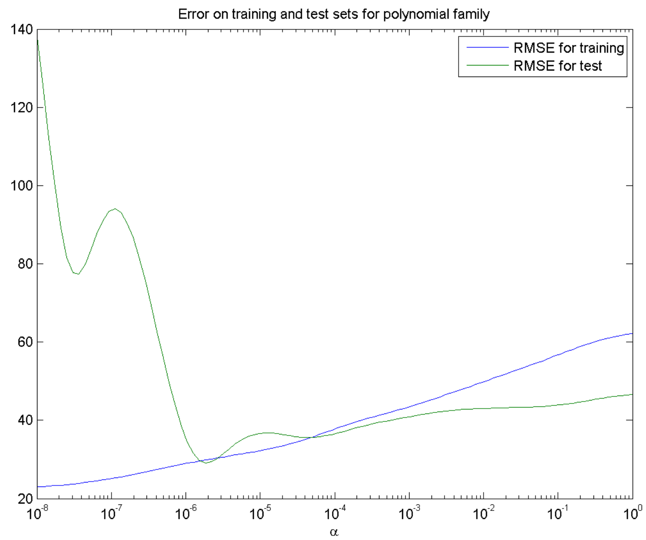
The figures are generated by q2_6.m

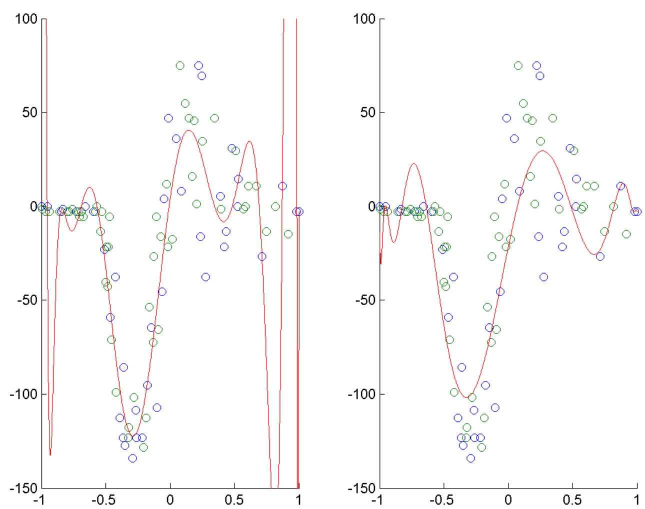


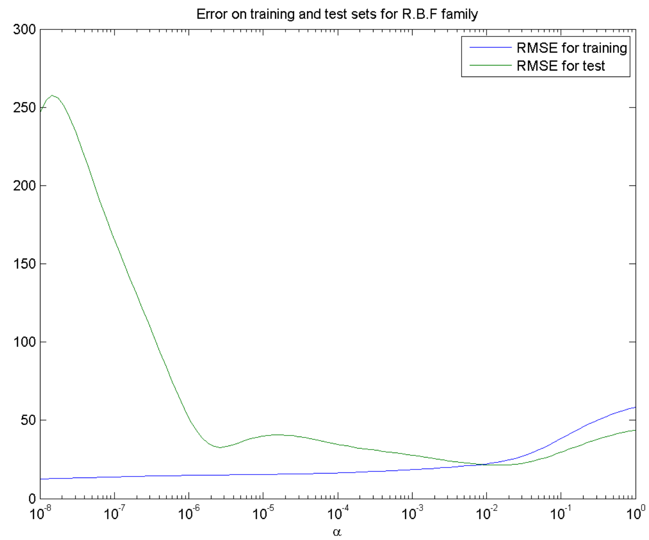
The usual methods for getting M.L. estimates will fail because the matrix $X^T X$ will be singular. Even if an alternate solution was found, such as finding a minimum norm solution or using a QR decomposition, the matrix will still be extrmely ill-behaved. In short, getting reasonable ML estimates is not possible for the case M = 100.

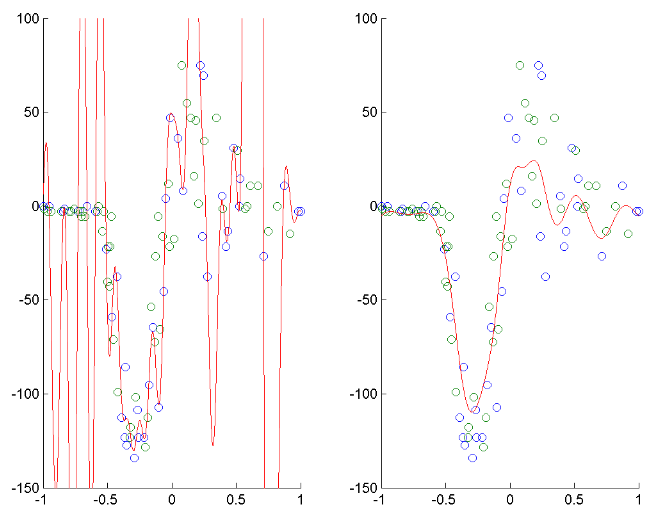
Q2.7

- The smallest training error for the polynomial case is obtained for $\alpha = 0.00000001$
- The smallest test error for the polynomial case is obtained for $\alpha = 0.000001831$
- The smallest training error for the rbf case is obtained for $\alpha = 0.00000001$
- The smallest test error for the rbf case is obtained for$\alpha = 0.013848864$

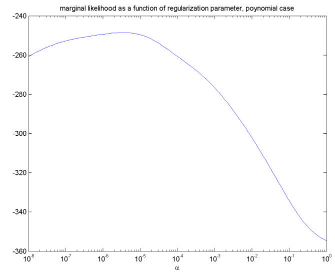

Error on training and test sets for polynomial family
RMSE for training
RMSE for test
α

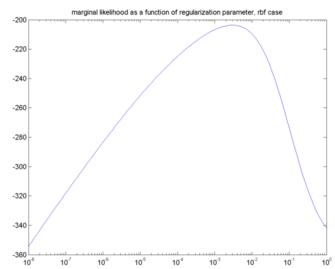



Error on training and test sets for R.B.F family
RMSE for training
RMSE for test
α



Q2.8





Q2.9

For both the polynomial and the R.B.F. case, the values of alpha which minimized the test errors and maximized the marginal likelihood are close. This suggests a way to bypass both cross validation and estimate regularization parameters and detecting overfitting without access to extra data.