Julio Soldevilla
EECS 545 Winter 2018 — Problem Set 1

Problem 1 THis problem has parts b,c,d,e that we present solutions in this document. The code for the solutions is in the zip file $jsolde\_hw1\_code.zip$ under the name prob1.py.

1. For problem 1b, we used Stochastic Gradient descent to train the regression model. For this problem, The weight for the stochastic grad desc is

$$\begin{bmatrix} 22.94022368 & -0.93943062 & 1.1925544 & 0.21191992 & 0.65780365 & -2.11200674 \\ 2.74910346 & 0.30622184 & -3.12390266 & 2.95424284 & -2.46344951 & -2.00963041 \\ 0.91701884 & -4.05450806 \end{bmatrix}$$

   where this last expression is just one 14×1 vector (but for the sake of space we wrote it as above.
   From this weight vector, we can see that the **bias term** is 22.94022368.
   The **training errors** (the MSE for training data) is 23.1929613454, while the **test errors** (the MSE for test data) is 10.9671896913.

   Finally, we can see the plot of the training error at the end of each epoch in figure 1 below.

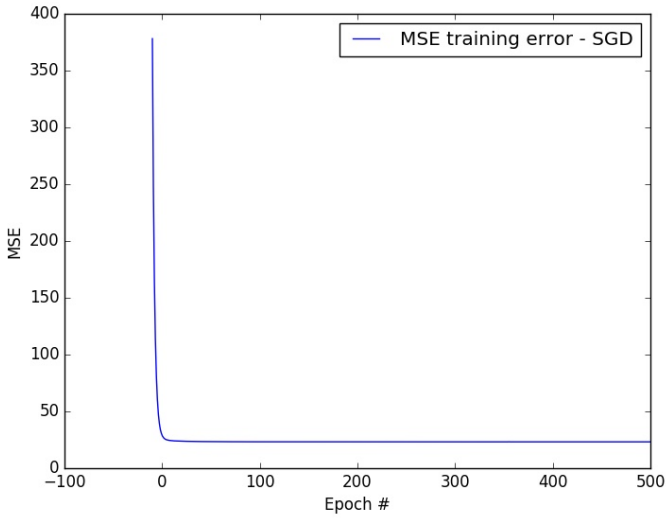

Figure 1: Graph showing the MSE training error when using SGD

2. For part c) we compute the batch gradient descent to train our model. The learned weight vector is

$$\begin{bmatrix} 22.94100877 & -0.936336221.18953456 & 0.21723452 & 0.66970526 & -2.10547565 \\ 2.75126807 & 0.30756426 & -3.12381189 & 2.95817231 - 2.4511536 \\ -2.00722057 & 0.90553088 & -4.05729245 \end{bmatrix}$$

where this last expression is just one 14×1 vector.
From this weight vector, we can see that the **bias term** is 22.94100877.

The **training errors** (the MSE for training data) is 23.191557895, while the **test errors** (the MSE for test data) is 10.9611791412.

Finally, we can see the plot of the training error at the end of each epoch in figure 2 below.

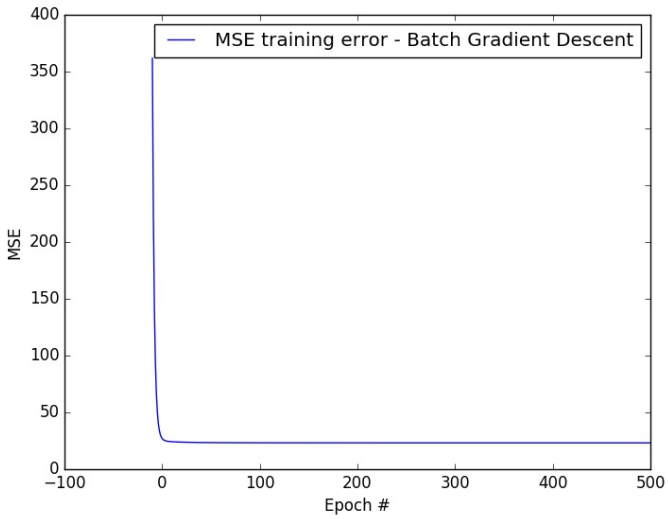

Figure 2: Graph showing the MSE training error when using BGD

3. In this part of the problem, we use the close solution to identify the weights and bias that minimize the MSE. The weight vector we obtain from using the closed form solution is

$$\begin{bmatrix} 22.94100877 & -0.936527281.18983479 & 0.2180906 & 0.66954197 & -2.10545149 \\ 2.75102471 & 0.30777503 & -3.12356704 & 2.96148512 & -2.45469868 \\ -2.00737039 & 0.90552685 & -4.05749492 & & \end{bmatrix}$$

(which again is a 14×1 vector but for the sake of space we wrote it as above) and from this expression we can see that the **bias that minimzes** the MSE is 22.94100877. Furthermore, from running the code, we see that The **training errors** (the MSE for training data) is 23.1915564692, while the **test errors** (the MSE for test data) is 10.9665431668.
From tese computations, we can see that the closed form parameters agree very closely with the ones we comptued in the previous part.

4. After computing the random train and test splits we obtained the following mean training error 21.8527415202 and the mean test error 23.3165328596

Problem 2 1. In the first of the problem, we get the following RMSE errors:

- 0th training RMSE: 9.48269466477
- 0th test RMSE: 5.99400813094
- 1-th training RMSE: 4.81576125542
- 1-th test RMSE: 3.3115771419
- 2-th training RMSE: 3.76629417811
- 2-th test RMSE: 7.14746413377
- 3-th training RMSE: 8.18681202323
- 3-th test RMSE: 12.1042413766
- 4-th training RMSE:8.50491214305
- 4-th test RMSE:141.55345743

Finally, we can also see the plot of these values in the followin graph (Figure 3):

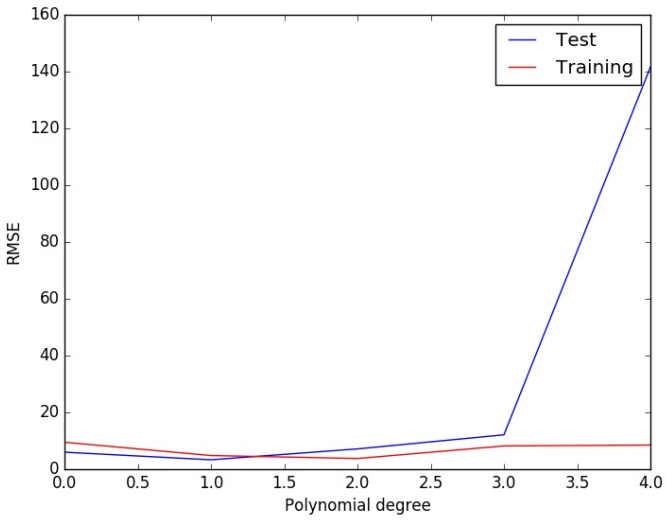

Figure 3: Graph showing the MSE error when doing polynomial features of different degrees

Since the last part of the graph is not very logical, to appreciate better the graph, we can just graph all elements except the last one, which we can see in Figure 4:

2. Finally, we get the following results for when we restrict the training set and start growing:

   - This is the RMSE for training with 20% data 11.4601626769
   - This is the RMSE for test with 20% data 38.918275945
   - This is the RMSE for training with 40% data 2.96802006698
   - This is the RMSE for test with 40% data 9.32135308145

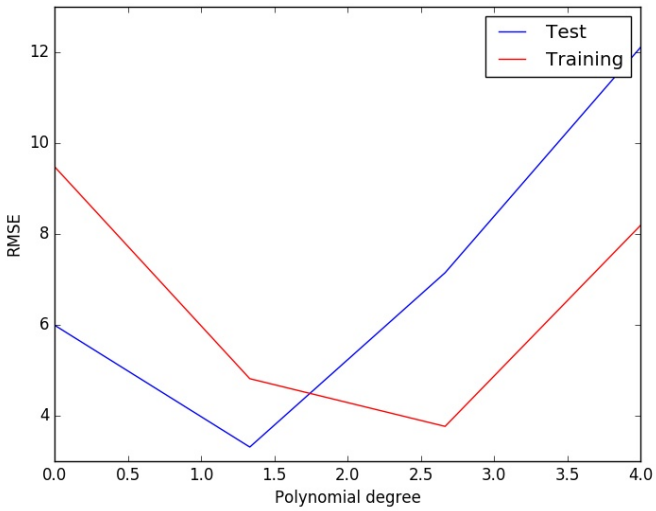

Figure 4: Graph showing the MSE error when doing polynomial features of different degrees, just removing the last value

- This is the RMSE for training with 60% data 3.14261785825
- This is the RMSE for test with 60% data 7.57749760854
- This is the RMSE for training with 80% data 3.25304667758
- This is the RMSE for test with 80% data 4.52585085547
- This is the RMSE for training with 100% data 4.81576125542
- This is the RMSE for test with 100% data 3.3115771419

Again, we can visualize this information as shown in Figure 5:

Problem 3 1. For this first problem, we show the work done in FIgure 6:

2. FOr this problem we repport all the possible RMSE, namely validation, training and test set for the possible lambdas:

   - The RMSE error of training set with $\lambda = 0$ is 4.75899856011
   - The RMSE error of validation set with $\lambda = 0$ is 7.84564027897
   - The RMSE error of test set with $\lambda = 0$ is 4.22655365439
   - The RMSE error of training set with $\lambda = 0.1$ is 5.31569487929
   - The RMSE error of validation set with $\lambda = 0.1$ is 5.49507024934
   - The RMSE error of test set with $\lambda = 0.1$ is 3.18380156819
   - The RMSE error of training set with $\lambda = 0.2$ is 6.37532816836

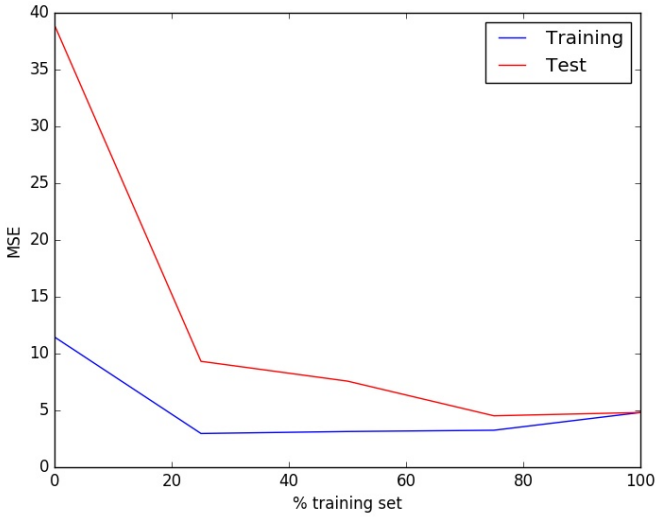

Figure 5: Graph showing the MSE error for training and test data when changing the available percentage of data.

- The RMSE error of validation set with $\lambda = 0.2$ is 4.54395472308
- The RMSE error of test set with $\lambda = 0.2$ is 3.92160042074
- The RMSE error of training set with $\lambda = 0.3$ is 7.51722511924
- The RMSE error of validation set with $\lambda = 0.3$ is 4.48444830133
- The RMSE error of test set with $\lambda = 0.3$ is 5.11613124943
- The RMSE error of training set with $\lambda = 0.4$ is 8.60568912038
- The RMSE error of validation set with $\lambda = 0.4$ is 4.93033974684
- The RMSE error of test set with $\lambda = 0.4$ is 6.29716462016
- The RMSE error of training set with $\lambda = 0.5$ is 9.60402790097
- The RMSE error of validation set with $\lambda = 0.5$ is 5.57508121459
- The RMSE error of test set with $\lambda = 0.5$ is 7.36843458243

Furthermore, we see that the lowest RMSE error on the validation set is 4.48444830133 happening on $\lambda = 0.3$. The corresponding test error for this $\lambda = 0.3$ is 5.11613124943.

Finally, we can see a picture of how the test, validation and training MSE behave for different λ in the table below, this picture corresponds to figure 7.

Problem 4 Finally, we show the theoretical in the following scanned pictures, corresponding to Figure 8, 9, 10 and 11.

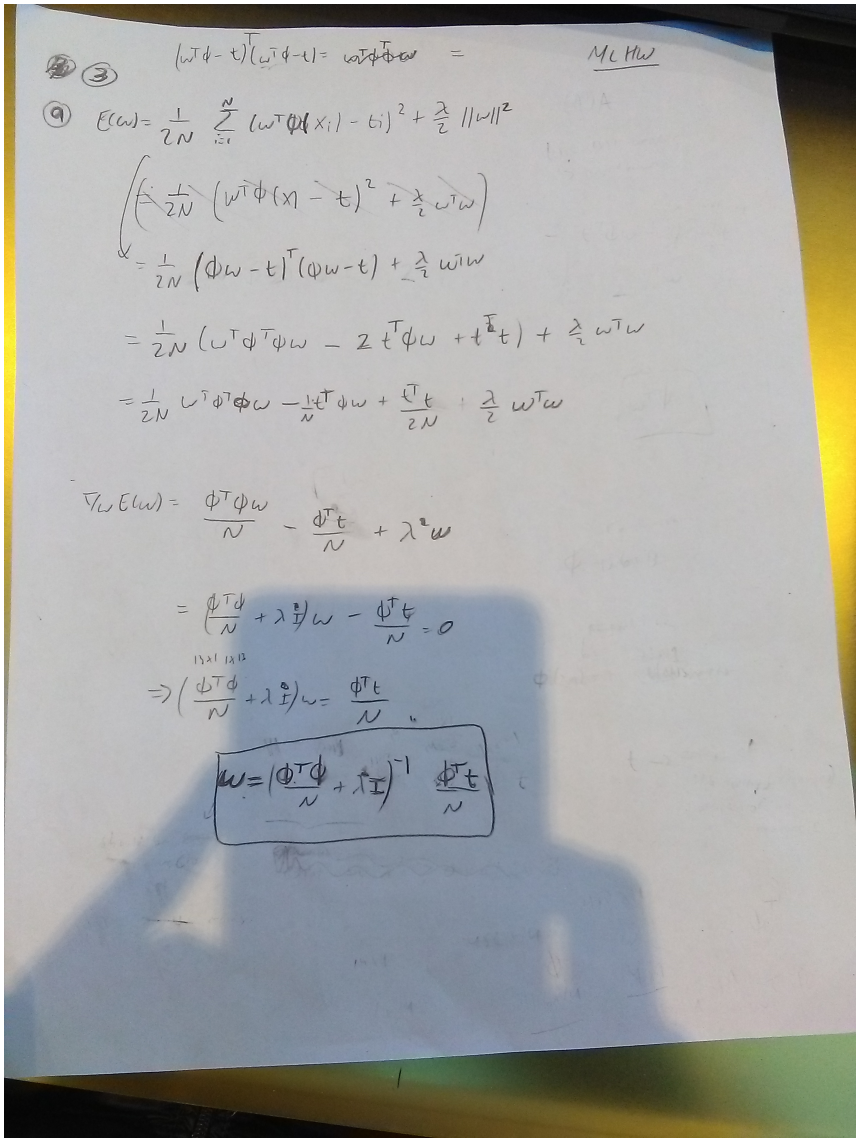

Figure 6: Image showing how we computed the closed form solution of the regularization problem in problem 3a.

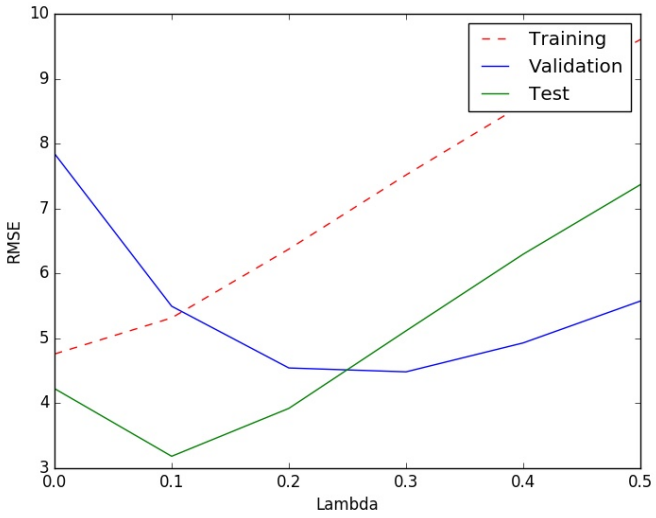

Figure 7: Image showing the behavior of RMSE when we have different regularization parameters.

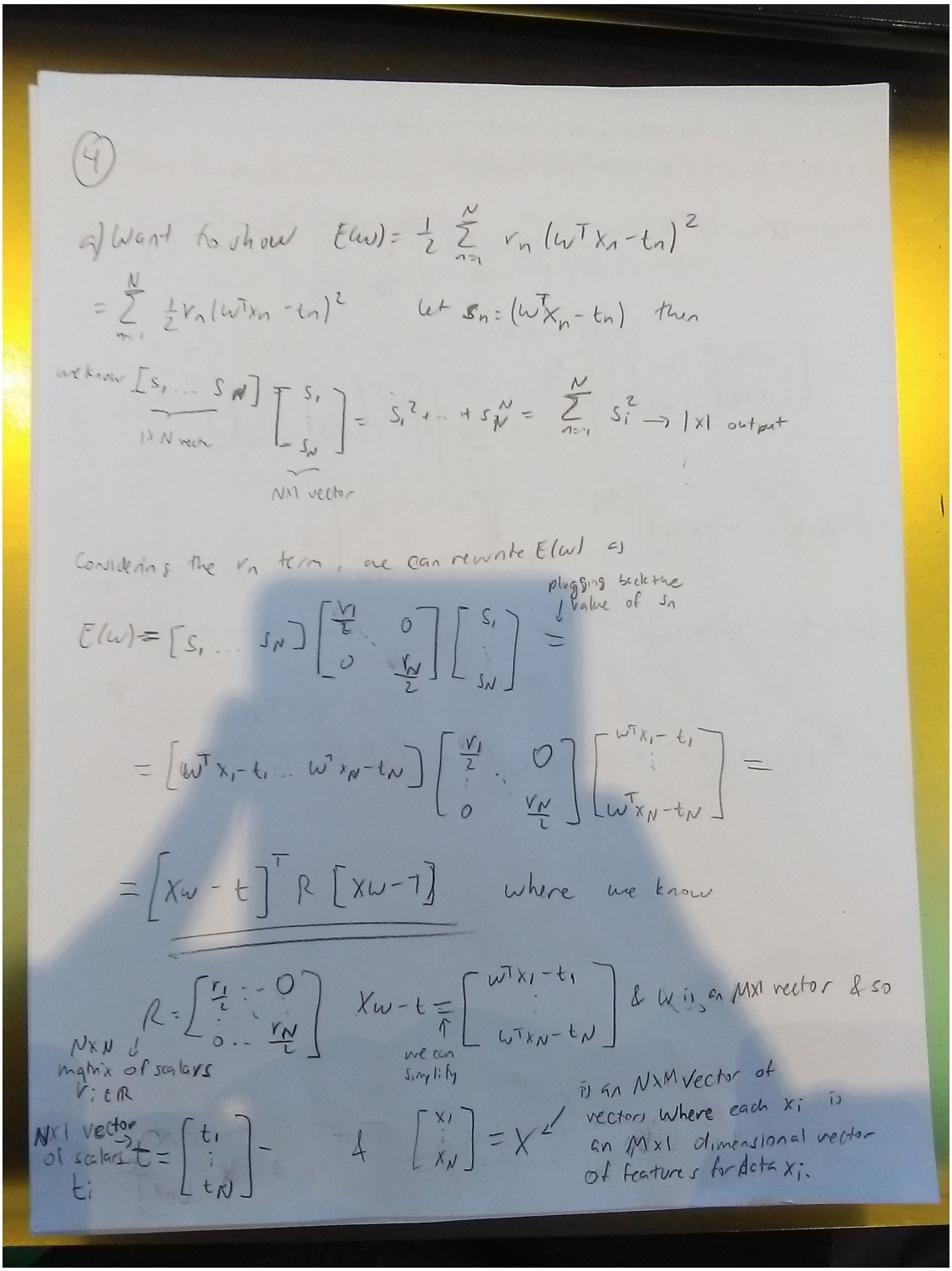

Figure 8: Image showing part a of problem 4.

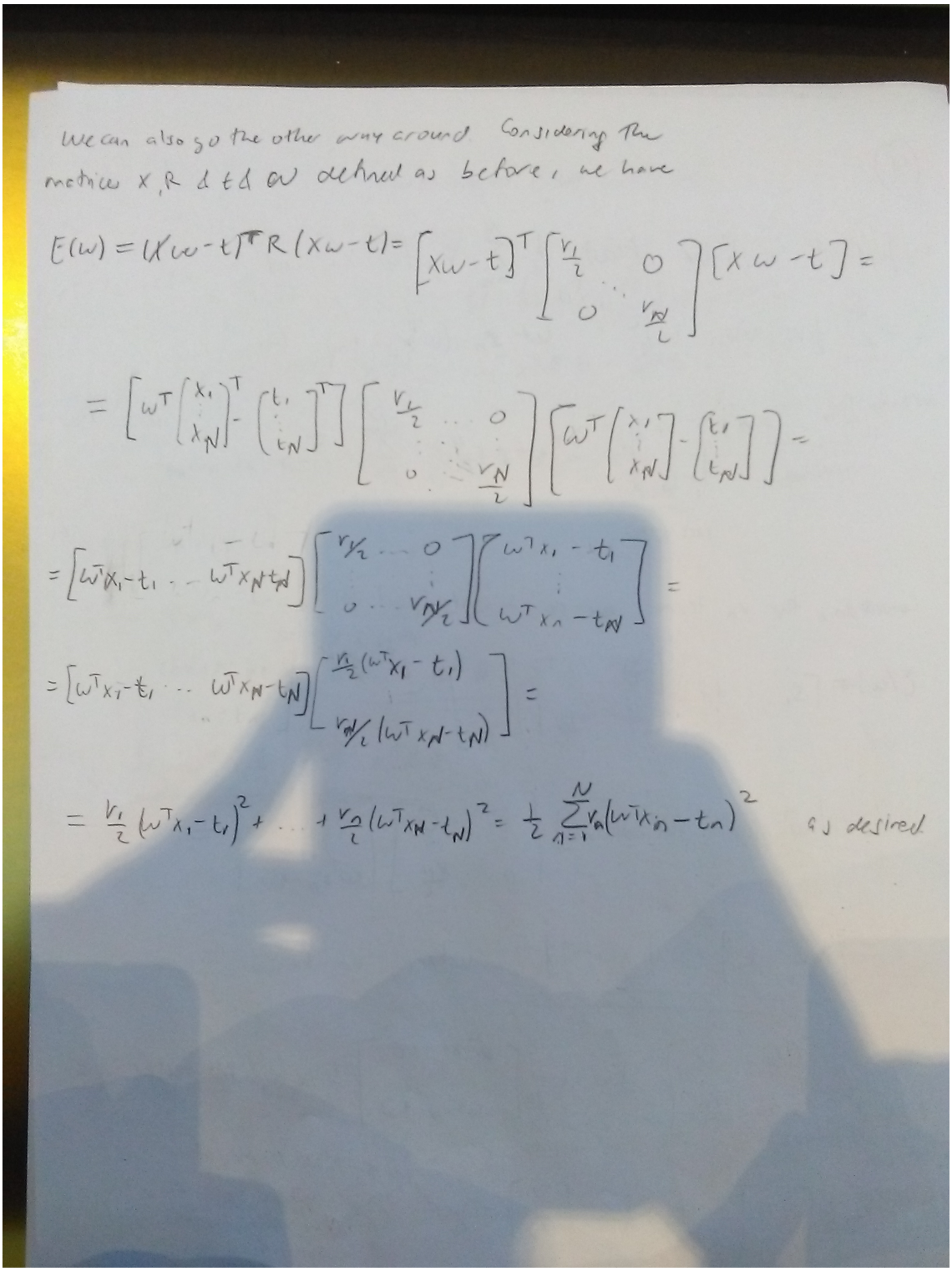

Figure 9: Image showing a continuitation to part a of problem 4a.

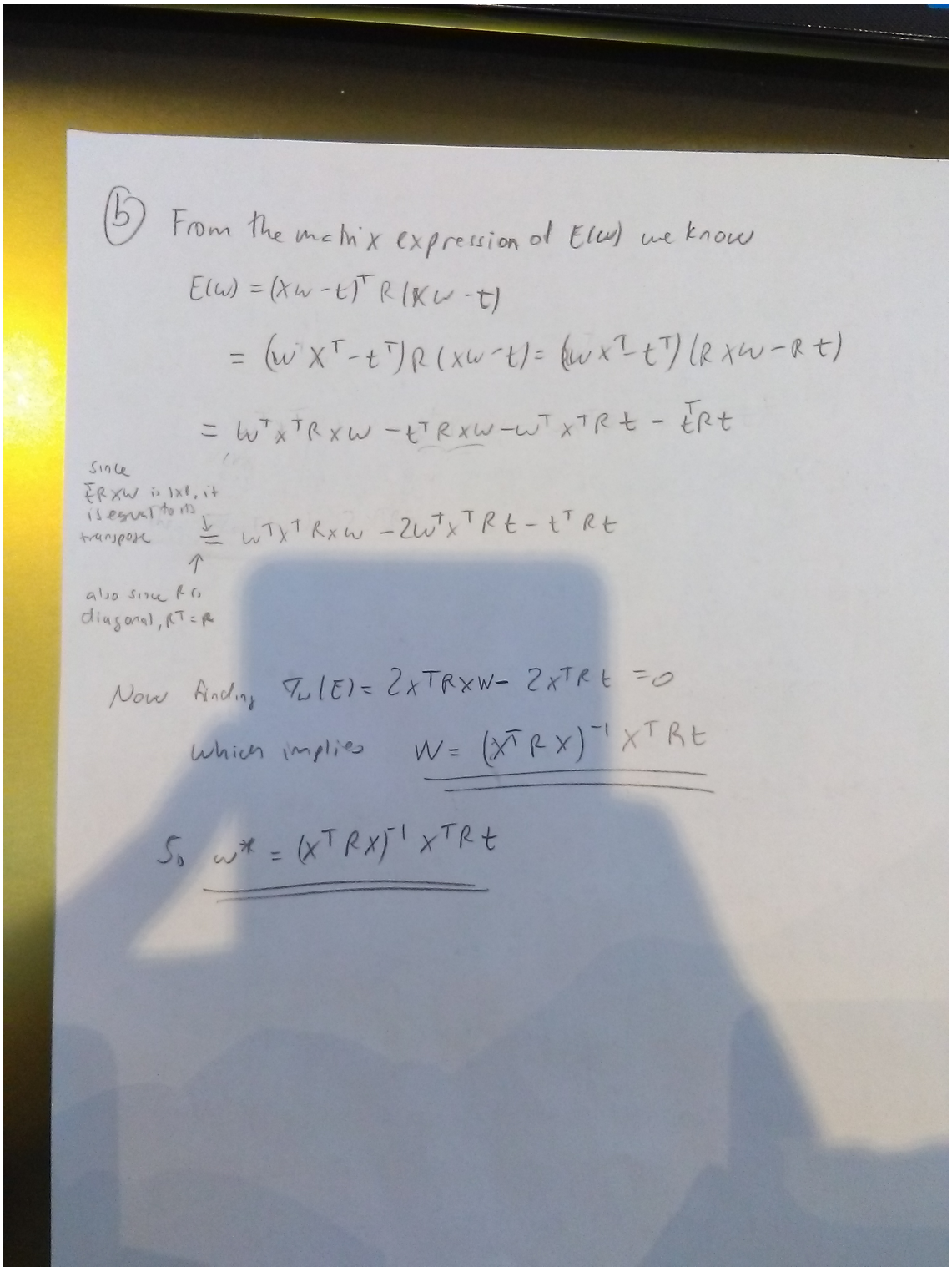

Figure 10: Image showing the work done to solve problem 4b.

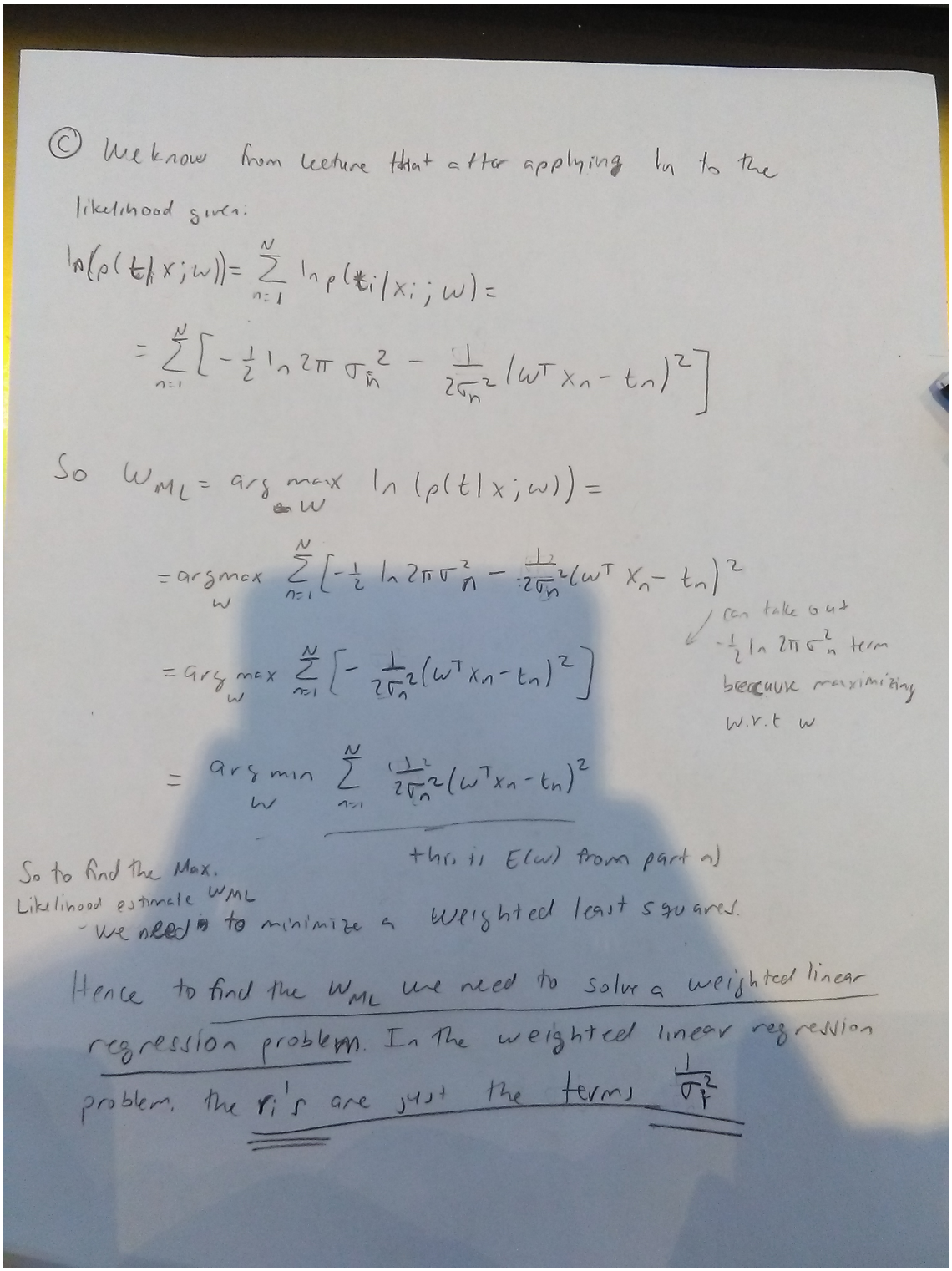

Figure 11: Image showing the work done to find the solution to problem 4c.