CSE 446 A3

Johan J How

**0. Policies**

0.1: List of Collaborators

0.2 List of Acknowledgements

0.3 I have read and understood these policies

**1. Binary Classification with Linear Regression on MNIST**

**1.1: Linear Regression, using the Closed Form Estimator**

1.1.1

Python spat out an error, saying that there was a singular matrix. This was because taking the inverse of 1/N XTX does not work, as the columns are not linearly independent. We must add lambda \* Identity Matrix to avoid this.

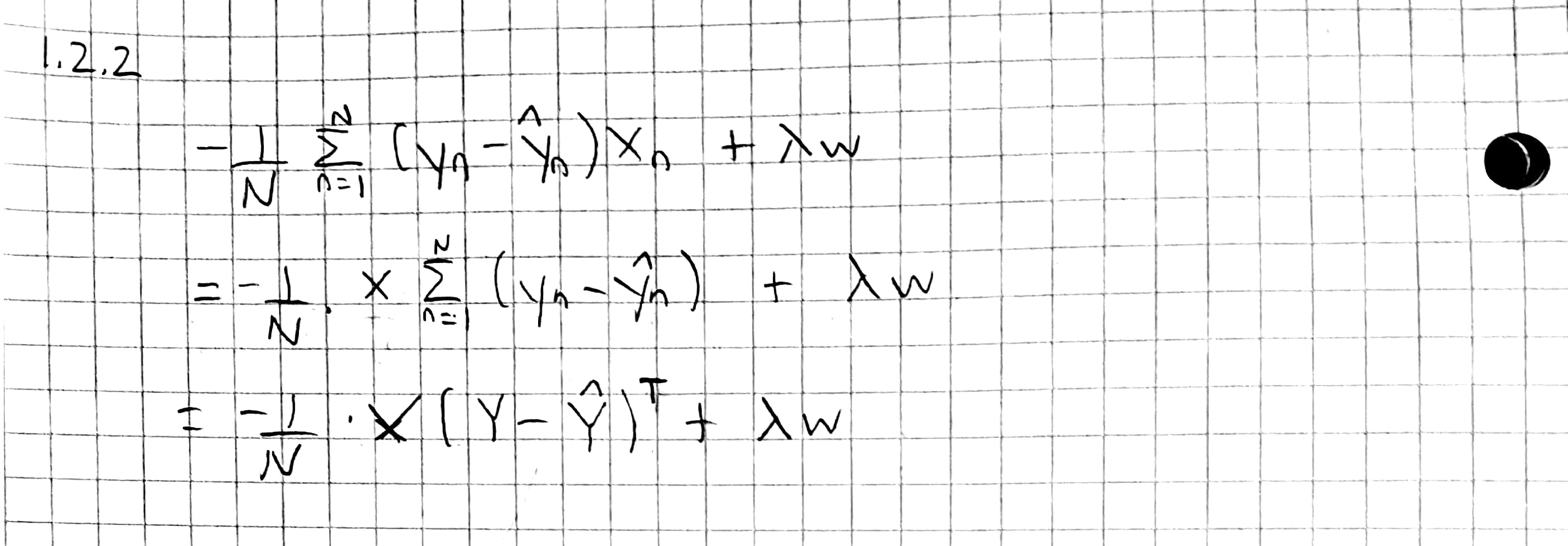
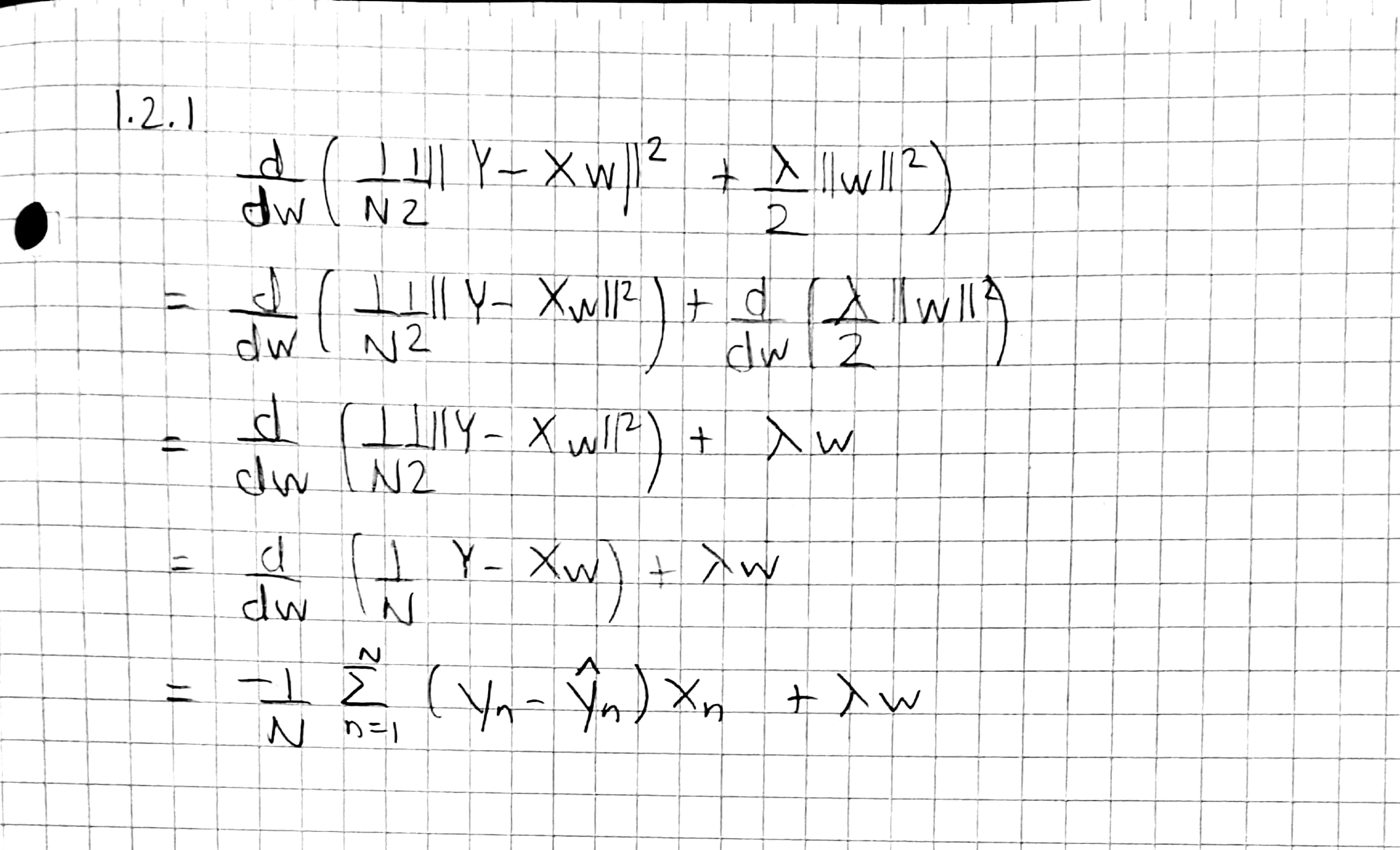
1.1.2

1. 1. I chose a lambda value of 1.
   2. Training Average Squared Error: 0.0276063621216%
   3. Test Average Squared Error: 0.0267821022407%
   4. Dev Average Squared Error: 0.0247244576109%
2. 1. Training Misclassification Error: 4.82121333869%
   2. Test Misclassification Error: 4.94855463008%
   3. Dev Misclassification Error: 4.1004613019%

1.1.3

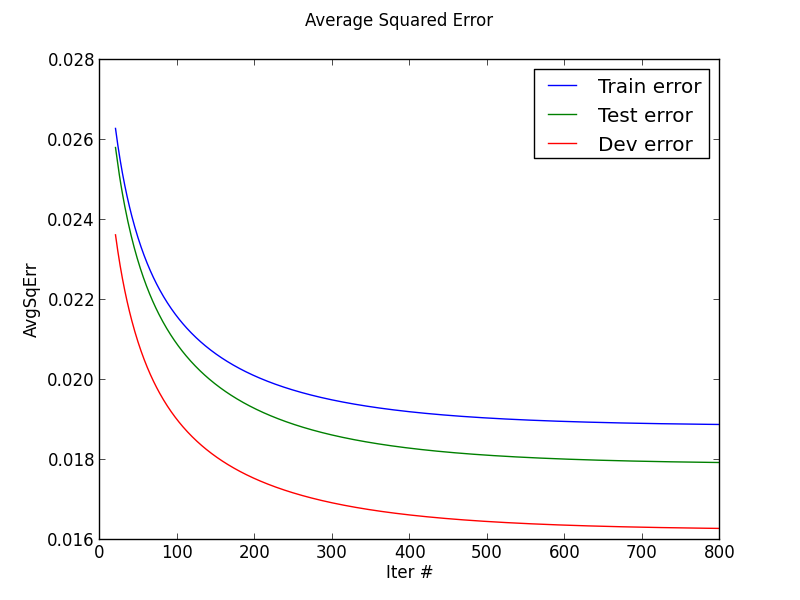
Because linear regression gives labels, rather than probabilities. We would rather use logistic regression because it gives probabilities.

**1.2: Linear regression using gradient descent**

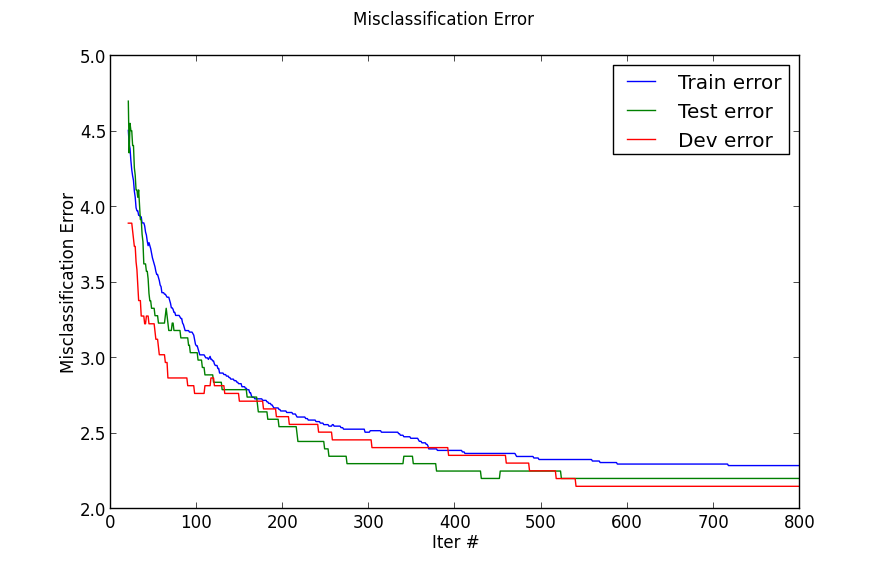


1.2.3

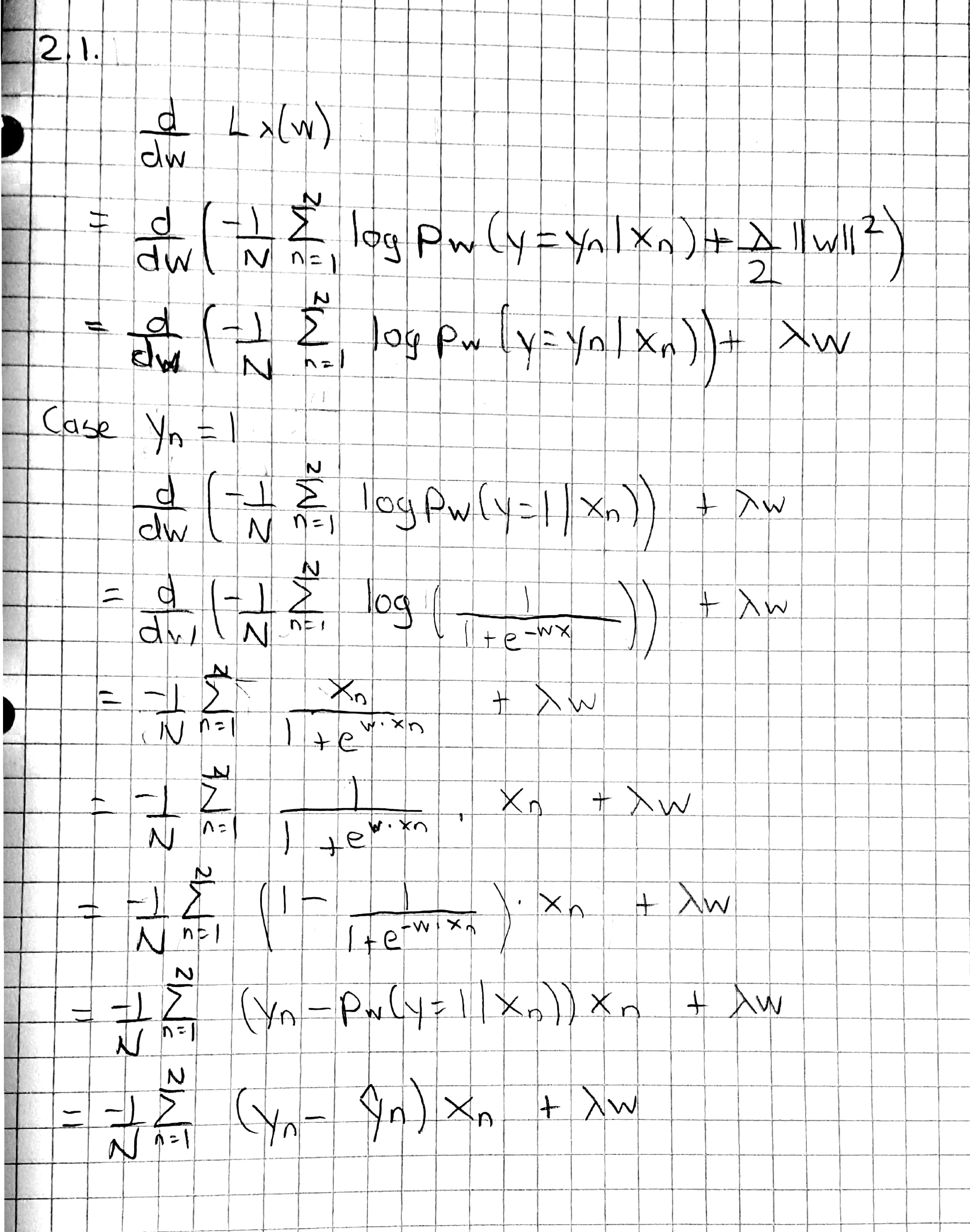
1. I chose a step size of 0.04, as this was one of the largest step sizes I could use.
2. I used a lambda value of 0.1



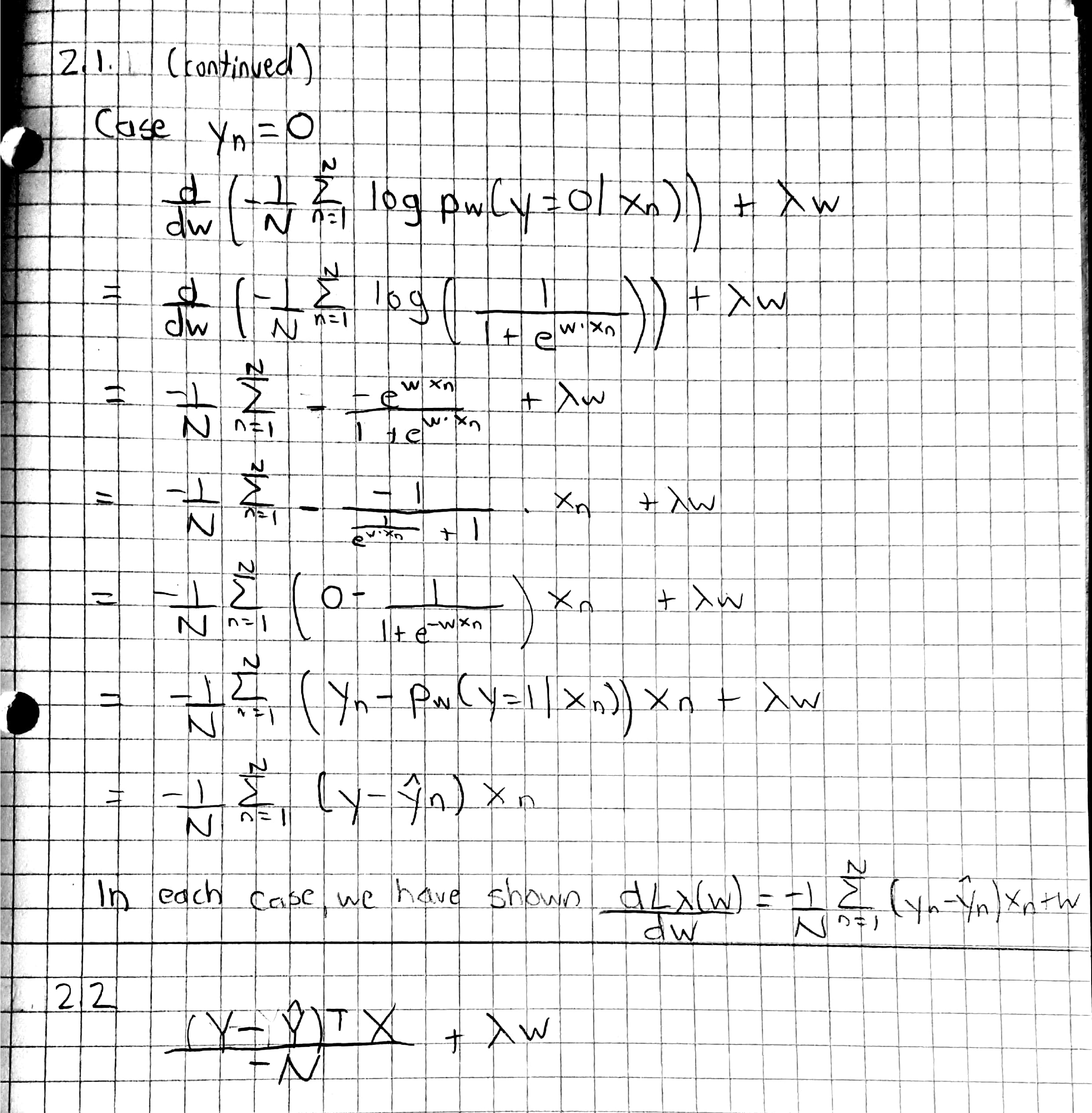
1. Lowest misclassification error: Train: 2.28%, Test: 2.156%, Dev: 2.153%



**2. Binary Classification with Logistic Regression**

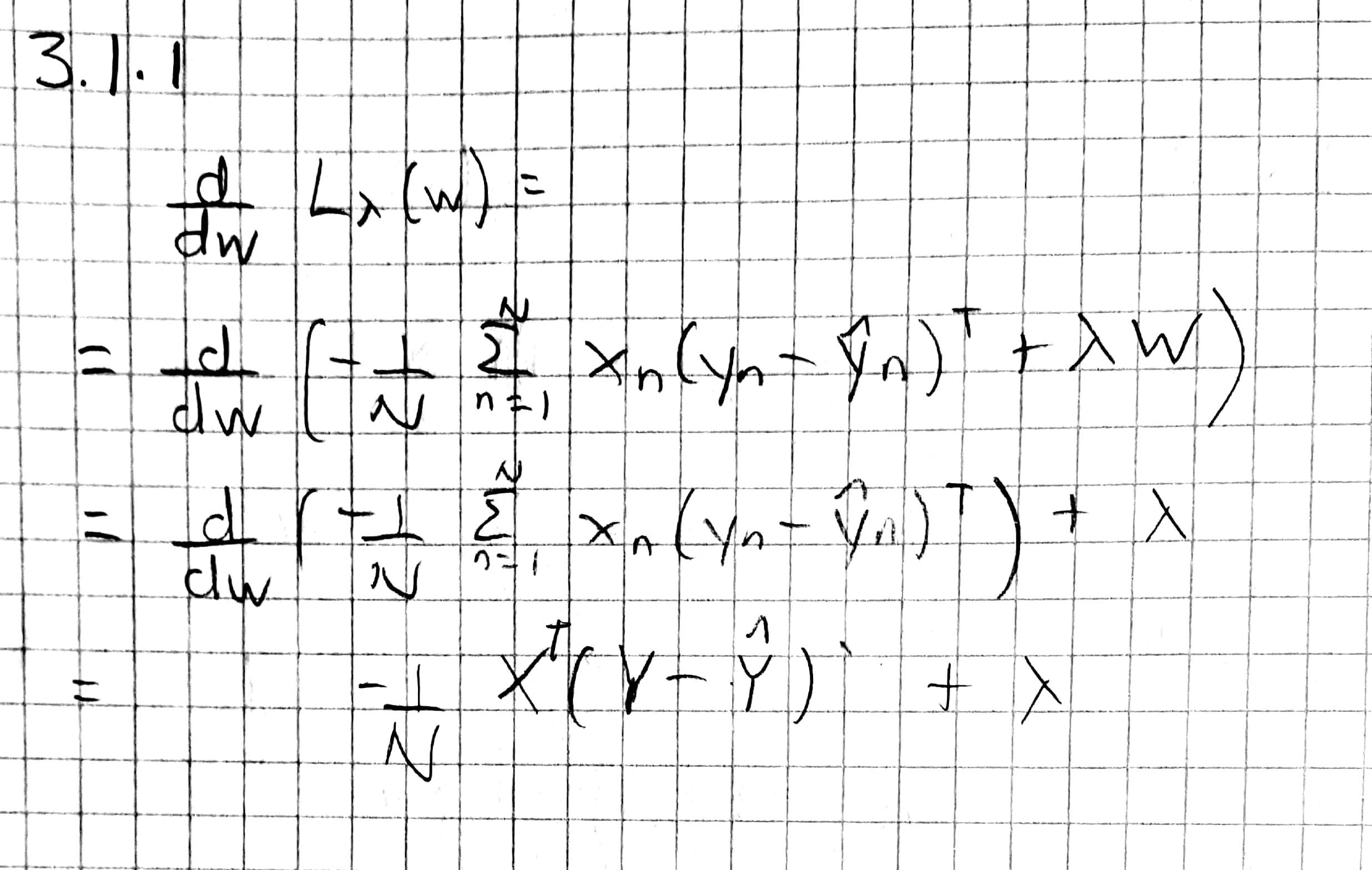


2.3



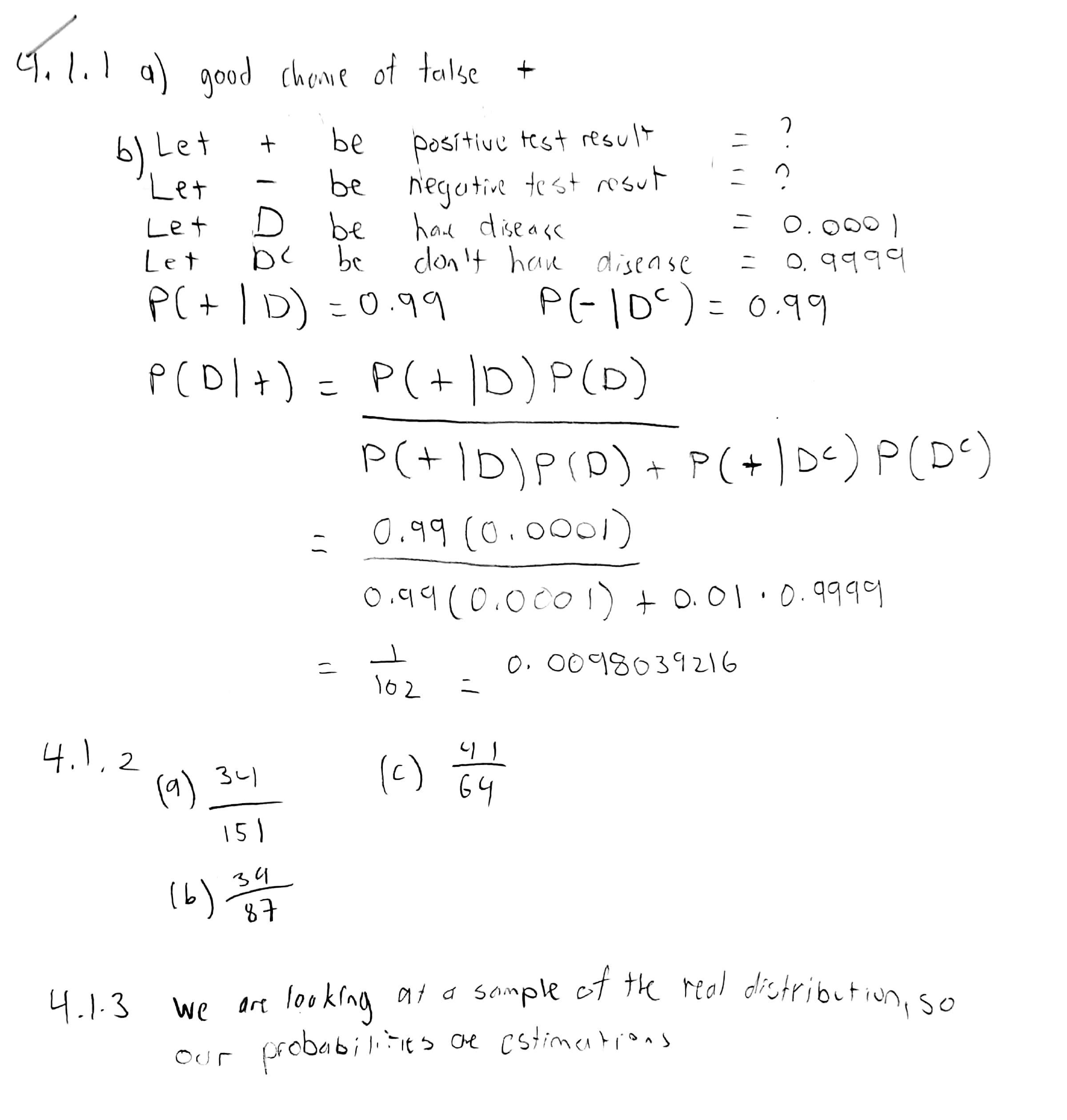
1. Our weight vector would converge to the linearly separable boundary. Without regularization, this would cause overfitting as the magnitude of the weights would go to infinity. This occurs because our data is linearly separable, causing the weight vector to converge to the boundary, but our lambda is 0, causing overfitting.
2. In the case of d > n, our weight vector will converge to a decision boundary that separates our data. In the case of d = n, our weight vector will converge to the exact solution. This is because our data is linearly independent. However, like (a), the magnitude of the weight vector will continue increasing to infinity, as lambda is 0.
3. Regularization makes sense as it will prevent the weight vector from overfitting.

**3. Multi-Class Classification using Least Squares**



**4. Probability and Maximum Likelihood Estimation**

**4.1 Probability Review**



**4.2 Maximum Likelihood Estimation**

