## Week 1 – Linear Classifiers and Logistic Regression

1. A linear classifier assigns the predicted class based on the sign of .
   1. True.
2. For a conditional probability distribution over , where takes on two values (+1, -1), .
   1. True.
3. Which function does logistic regression use to “squeeze” the real line to [0, 1]?
   1. Logistic.
4. If , what is true about ?
   1. .
5. Consider training a 1 vs. all multiclass classifier for the problem of digit recognition using logistic regression. There are 10 digits, thus there are 10 classes. How many logistic regression classifiers will we have to train?
   1. 10.

## Week 2 – Training Linear Classifiers

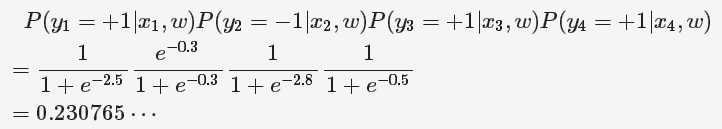
1. A linear classifier can only learn positive coefficients.
   1. False.
2. In order to train a logistic regression model, we find the weights that maximise the likelihood of the model.
   1. True.
3. The data likelihood is the product of the probability of the inputs given the weights andresponse .
   1. False.
4. Consider the situation where the inputs are 1-dimensional, the current estimates of the weights are and where is the intercept and is the weight for , and the data is:

|  |  |
| --- | --- |
| **x** | **y** |
| 2.5 | +1 |
| 0.3 | -1 |
| 2.8 | +1 |
| 0.5 | +1 |

Calculate the likelihood of this data.

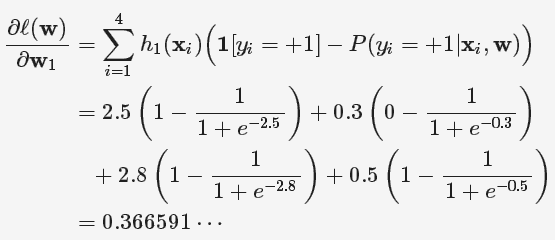
Given, and

|  |  |  |
| --- | --- | --- |
| **x** | **y** |  |
| 2.5 | +1 |  |
| 0.3 | -1 |  |
| 2.8 | +1 |  |
| 0.5 | +1 |  |



1. Calculate the derivate of the log-likelihood with respect to .

|  |  |  |  |
| --- | --- | --- | --- |
| **x** | **y** |  |  |
| 2.5 | +1 | 0.9241 | 0.1898 |
| 0.3 | -1 | 0.4256 |  |
| 2.8 | +1 | 0.9427 |  |
| 0.5 | +1 | 0.6225 |  |



1. What is true about gradient descent?
   1. It is an iterative algorithm.
   2. It finds the maximum by “hill climbing”.

## Week 2 – Overfitting and Regularisation in Logistic Regression