**Lab 09**

|  |  |
| --- | --- |
| Name: | Andrew Hsieh |
| Student ID: | B08611010 |
| Total Score: |  |

1. **Multiple Choice (35 points, 5 points each question)**

|  |  |  |  |
| --- | --- | --- | --- |
| # | Answer | Explanation (Optional) | Score |
| 1 | A |  |  |
| 2 | C |  |  |
| 3 | B |  |  |
| 4 | D | B: 如果改進optimization algorithm讓模型可以有更佳的error minimization的話，有可能會讓模型在training dataset上達成更低的loss，有可能會造成模型更加過擬合。  D: noise可以增加dataset的多樣性，如果在已經overfitting 的dataset中降低noise，將會造成dataset更加單一，過擬合更加嚴重。 |  |
| 5 | C |  |  |
| 6 | C |  |  |
| 7 | E |  |  |

1. **Simple Linear Regression (30 points, 10 points each question)**

|  |  |  |
| --- | --- | --- |
| # | Description | Score |
| 1 | 1) Done.  2) Yes, they are in correct sizes.  3) Done. |  |
| 2 | 1) Completed.  2) Yes, it seems very reasonable to me.  3) Done. |  |
| 3 | 1) Done.  2) Yes, they are basically the same.  3) Done. |  |

1. **Multiple and Polynomial Regression (35 points)**

|  |  |  |
| --- | --- | --- |
| # | Description | Score |
| 1 | 使用train\_test\_split直接將需要的資料分為training set與testing set。Train size : test size的比例直接依照預設值的3:1。預見的問題是沒辦法保證equal representation of each region，在分資料的時候使用random shuffle來盡量達成。 |  |
| 2 | 1) According to the R square score, I think the model is decent.  2) From the coefficients, we presume that:  ■ Age is important, while age squared is not.  ■ Log2(page\_views) is also a driving factor for market value, but not as powerful as age.  ■ The bigger the club a player is in, the higher his market value is.  3) A player should get older and go to bigger club to improve their market value. A player should go get 30.92436619 page reviews to increase their market value by 10.  4) The model performs equally well on the training and test data. I don’t think the model is overfitting because there are no significant gap between the R square score of training and testing dataset. |  |