

MLDS: Homework 1

ANDRAŽ DE LUISA

February 28, 2020

Given a data set with numeric input (independent) variables x_i in and binary target (dependent) y_i , we decide to use logistic regression: $(y_i|x_i, \beta) \sim \text{Bernoulli}(\theta_i)$, where $\theta_i = \frac{1}{1+e^{-\beta^T x_i}}$.

Exercise 1 Derive the log likelihood:

$$\begin{aligned} L(\beta, y, x) &= \prod_{i=1}^n \theta_i^{y_i} (1 - \theta_i)^{1-y_i} = \prod_{i=1}^n \frac{1}{1+e^{-\beta^T x_i}} y_i \cdot \left(1 - \frac{1}{1+e^{-\beta^T x_i}}\right)^{1-y_i} = \prod_{i=1}^n \frac{e^{-\beta^T x_i (1-y_i)}}{1+e^{-\beta^T x_i}} \\ l(\beta, y, x) &= \sum_{i=1}^n -\beta^T x_i (1 - y_i) - \log(1 + e^{-\beta^T x_i}) \end{aligned}$$

Exercise 3b Derive the gradient of the log-likelihood:

$$\begin{aligned} \frac{\delta}{\delta \beta} l(\beta, y, x) &= \sum_{i=1}^n -x_i^T (1 - y_i) + \frac{x_i e^{-\beta^T x_i}}{1+e^{-\beta^T x_i}} = \\ \sum_{i=1}^n x_i \left(y_i - 1 + \frac{e^{-\beta^T x_i}}{1+e^{-\beta^T x_i}} \right) &= \\ = \sum_{i=1}^n x_i \left(y_i - \frac{1}{1+e^{-\beta^T x_i}} \right) \end{aligned}$$

The number of function calls before the optimization was varying between 5000 and 25000 (depending on the degree of polynomial expansion), while after the optimization it stabilized around 20.

Exercise 5 Before the application of logistic regression I normalized the data in the 'housing' dataset, because the intervals in which data in different columns were lying were too much different between them. The 'painted' dataset was already normalized on the [0,1] interval.

In figures 1 and 2 the misclassification rates with respect to the growth of the degree of polynomial expansion of the dataset features are shown. Both plots show a similar pattern of overfitting the training data after some time, which is expected since the number of data instances in the datasets is substantially smaller than the number of features in an expanded

dataset of degree 4 or more. Still, the logistic regression in the 'housing' dataset is quite stable if we compare it to the 'painted' dataset, its double size does probably help here.

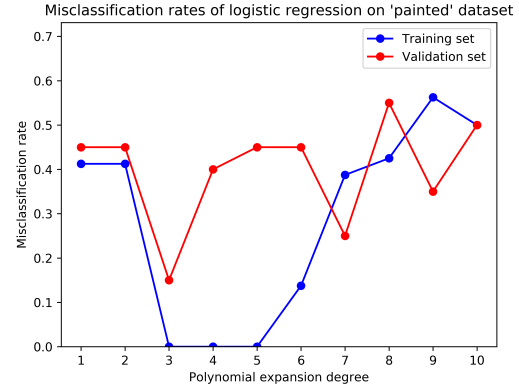


Figure 1: Changes of misclassification rates with respect to the polynomial expansion degree of the 'painted' dataset.

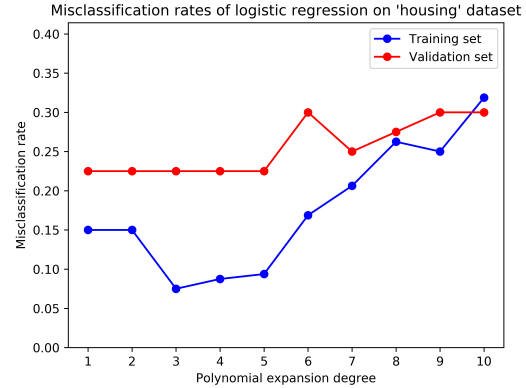


Figure 2: Changes of misclassification rates with respect to the polynomial expansion degree of the 'housing' dataset.